

Judgment under Uncertainty: Heuristics and biases

Edited by

*Daniel Kahneman
University of British Columbia*

*Paul Slovic
Decision Research
A Branch of Perceptronics, Inc.
Eugene, Oregon*

*Amos Tversky
Stanford University*

This work in preparing this book was supported by Office of Naval Research Grant N00014-79-C-0077 to Stanford University and by Office of Naval Research Contract N0014-80-C-0150 to Decision Research.



Cambridge
UNIVERSITY PRESS, 2001

*Даниель Канеман Пауль Словик
Амос Тверски*

Принятие решений в неопределенности

Правила и предубеждения

*Под редакцией
Заслуженного работника высшей школы
Российской Федерации, доктора психологических наук,
проф. Г. В. Суходольского*

Перевод с английского



Гуманитарный Центр
Харьков, 2005

УДК 159.922.27

ББК 88.53

Научный редактор: доктор психол. наук *Г. В. Суходольский* (СПбГУ),

Редактор: доцент, кандидат психол. наук *С. Г. Тарасов* (СПбГУ)

Перевод. Изд-во Институт прикладной психологии
“Гуманитарный Центр”

Д. Канеман, П. Словик, А. Тверски. Принятие решений
в неопределенности: Правила и предубеждения / Пер. с англ. –Х.:
Изд-во Институт прикладной психологии
“Гуманитарный Центр”, 2005, – 632 с.

ISBN 966-8324-14-5

Перед Вами фундаментальный труд по психологии принятия решений. Ссылки на отдельные работы этих авторов достаточно часто встречаются в академической литературе, но полный сборник этих статей на русском языке опубликован впервые. Выход этой книги, безусловно, важное событие для специалистов в области управления, стратегического планирования, принятия решений, поведения потребителей и т.д.

Книга представляет интерес для специалистов в области менеджмента, экономики, психологии, как в теории, так и в практике, которые имеют дело с такой сложной и интересной областью человеческой деятельности, как принятие решений.

You can see a fundamental work on psychology of decision-making. The references to separate works of these authors can be found often enough in the academic literature, but the complete volume of these articles has been published in Russian for the first time. The issue of this book is undoubtedly an important event for the experts in the field of management, strategic planning, decision-making, consumers' conduct etc.

The book is of interest to the experts in the field of management, economy, psychology, in theory as well as in practice, who deal with such complex and interesting sphere of human action as decision-making.

Все права защищены. Никакая часть данной книги не может быть воспроизведена в какой бы то ни было форме без письменного разрешения владельцев авторских прав. Права на публикацию книги на русском языке принадлежат издательству «Гуманитарный Центр».

ISBN 0-521-28414-7 (англ.)

ISBN 966-8324-14-5

© Cambridge University Press, 2001

© Институт прикладной психологии «Гуманитарный центр», перевод на русский язык, оформление, 2005

Содержание

Предисловие научного редактора.....	9
Предисловие авторов.....	11
Часть I: Введение.....	15
1. Принятие решений в условиях неопределенности: правила и предубеждения.....	17
Часть II: Репрезентативность.....	37
2. Вера в закон малых чисел.....	39
3. Субъективная вероятность: оценка репрезентативности.....	49
4. О психологии прогнозирования.....	66
5. Изучение репрезентативности.....	89
6. Оценки репрезентативности и на основе репрезентативности....	105
Часть III: Причинность и атрибуция.....	121
7. Общепринятое положение: информация не обязательно информативна.....	123
8. Причинные схемы при принятии решений в условиях неопределенности.....	140
9. Недостатки процесса атрибуции: о происхождении и исправлении ошибочных социальных оценок.....	153
10. Очевидное воздействие базового значения.....	179

Часть IV: Доступность.....	189
11. Доступность: эвристика оценки частоты и вероятности.....	191
12. Эгоцентрические предубеждения в доступности и атрибуции..	208
13. Предубеждения доступности в социальном восприятии и взаимодействии.....	220
14. Эвристика моделирования.....	232
Часть V: Ковариация и контроль.....	241
15. Субъективная оценка ковариации: суждения, основанные на данных против суждений, основанных на теориях.....	243
16. Иллюзия контроля.....	265
17. Результаты тестов – такие, какими Вы их себе представляете.....	274
18. Вероятностные рассуждения в клинической медицине: проблемы и возможности.....	286
19. Получение знаний из опыта и условно-оптимальных правил при принятии решения.....	308
Часть VI. Чрезмерная уверенность.....	327
20. Чрезмерная уверенность в суждениях, основанных на конкретных примерах.....	329
21. Сообщение о процессе обучения оцениванию вероятности.....	337
22. Калибровка вероятностей: положение дел к 1980 г.....	351
23. Обреченным необходимо изучать прошлое: эвристики и предубеждения в ретроспективе.....	383
Часть VII: Многоступенчатая оценка.....	403
24. Оценка составных вероятностей в последовательном выборе...	405
25. Консерватизм в процессе обработки информации человеком...	410
26. Гипотеза угадывания в многоступенчатом выводе.....	422
27. Заключение о личных характеристиках на основе информации, восстановленной из памяти.....	431

Часть VIII: Коррективные процедуры.....	445
28. Устойчивая привлекательность ошибочных линейных моделей при принятии решений.....	447
29. Жизнеспособность мифических чисел.....	466
30. Интуитивное прогнозирование: предубеждения и корректирующие процедуры.....	473
31. Освобождение от предубеждения.....	482
32. Улучшение индуктивного вывода.....	508
Часть IX: Восприятие риска.....	527
33. Факты против страха: понимание воспринимаемого риска.....	529
Часть X: Поскриптум.....	563
34. К вопросу о статистической интуиции.....	565
35. Варианты неопределенности.....	582
Литература.....	595

Предисловие научного редактора

В предлагаемой Вашему вниманию книге изложены результаты размышлений и экспериментальных исследований зарубежных ученых, малоизвестные русскоязычному читателю.

Речь идет об особенностях мышления и поведения людей при оценке и прогнозировании неопределенных событий и величин, таких как, в частности, шансы на выигрыш или заболевание, предпочтения на выборах, оценка профпригодности, экспертиза аварийности и многое другое.

Как убедительно показано в книге, принимая решения в неопределенных условиях, люди обычно ошибаются, иногда весьма значительно, даже если они изучали теорию вероятности и статистику. Эти ошибки подчинены определенным психологическим закономерностям, которые выявлены и хорошо экспериментально обоснованы исследователями.

Надо сказать, что не только закономерные ошибки человеческих решений в ситуации неопределенности, но и сама организация экспериментов, выявляющих эти закономерные ошибки весьма интересна и практически полезна.

Можно с уверенностью думать, что перевод этой книги будет интересен и полезен не только отечественным психологам, врачам, политикам, разного рода экспертам, но и многим другим людям, так или иначе связанным с оценкой и прогнозом по сути случайных социальных и личных событий.

*Научный редактор,
доктор психологических наук,
профессор СПбГУ
Г.В. Суходольский,
Санкт-Петербург, 2004 г.*

Предисловие авторов

Подход к процессу принятия решений, представленный в данной книге, основывается на трех направлениях исследований, получивших развитие в 50-х и 60-х годах XX столетия. Например, сравнение клинического и статистического прогнозирования, начатое Паулем Милом (Paul Meehl); изучение субъективной вероятности в парадигме Байеса (Bayes), представленное в психологии Вардом Эдвардсом (Ward Edwards); и исследование эвристики и стратегий рассуждения, представленной Гербертом Саймоном (Herbert Simon) и Джеромом Брунером (Jerome Bruner). Наш сборник также включает современную теорию, находящуюся на стыке процесса принятия решений с другой отраслью психологического исследования: изучением причинной атрибуции и бытовой психологической интерпретации, впервые представленной Фрицом Хайдером (Fritz Heider).

Классическая книга Мила, изданная в 1954, подтверждает тот факт, что простые линейные комбинации высказываний превосходят интуитивные суждения экспертов в предсказании существенных поведенческих критериев. Не теряющее актуальности и сегодня интеллектуальное наследие этой работы, и шумная полемика, последовавшая за ней, вероятно, не доказывали, что клиницисты плохо справились со своей работой, за которую, как отмечал Мил, им и не стоило бы браться. Скорее, это была демонстрация существенного несоответствия между объективными показателями успеха людей в задачах предсказания и их искренним убеждением о собственной продуктивности. Это заключение верно не только для клиницистов и клинических прогнозов: мнения людей о том, как они делают выводы и то, насколько хорошо они это делают, не могут быть приняты в качестве основания. Ведь исследователи, практикующие клинический подход часто использовали самих себя или своих друзей в качестве испытуемых, а интерпретация ошибок и отклонений были скорее когнитивными, чем психодинамическими: в качестве модели использовались скорее впечатления об ошибках, чем действительные ошибки.

С момента привлечения Эдвардсом и его коллегами идей Байеса в психологическое исследование, психологам впервые предложили целостную и четко сформулированную модель оптимального поведения в условиях неопределенности, с которой можно было сравнить принятие решений человеком. Соответствие принятия решений нормативным моделям стало одной из главных парадигм исследова-

ния в области суждения в условиях неопределенности. Это неизбежно затрагивало проблему предубеждений, к которым люди тяготеют при индуктивных выводах, и методов, которые могли бы использоваться для их коррекции. Эти проблемы рассматриваются в большинстве разделов этого издания. Однако во многих ранних работах использовалась нормативная модель для объяснения человеческого поведения, а также вводились дополнительные процессы для объяснения отклонений от оптимальных показателей. Напротив, целью исследования в области эвристик в принятии решений является объяснение как правильных, так и ошибочных суждений в терминах одних и тех же психологических процессов.

Появление такой новой парадигмы, как когнитивная психология, оказало серьезное влияние на исследование принятия решений. Когнитивная психология рассматривает внутренние процессы, ментальные ограничения, и то, как именно ограничения влияют на эти процессы. Ранними примерами концептуальной и эмпирической работы в этой области являлись исследование стратегий мышления Брунера и его коллег, а также трактовка эвристик рассуждений и ограниченной рациональности Саймона. И Брунер, и Саймон занимались стратегиями упрощения, которые уменьшают сложность задач принятия решений для того, чтобы сделать их приемлемыми для образа мышления людей. Мы включили в эту книгу большинство работ, исходя из подобных соображений.

В течение последних лет большой объем исследований был посвящен эвристикам суждений, а также исследованию их эффектов. Данное издание комплексно рассматривает этот подход. В нем собраны новые работы, написанные специально для этого сборника, и уже изданные статьи по вопросам суждений и предположений. Хотя граница между суждением и принятием решения не всегда ясна, мы сосредоточились здесь скорее на суждении, чем на выборе. Тема принятия решения достаточно важна для того, чтобы быть предметом отдельного издания.

Книга состоит из десяти частей. Первая часть содержит ранние исследования эвристики и стереотипов при интуитивном принятии решений. В части II специально рассмотрена эвристика репрезентативности, которая, в части III, расширена до проблем причинной атрибуции. Часть IV описывает эвистику доступности и ее роль в социальном суждении. Часть V рассматривает понимание и изучение ковариации, а также показывает наличие иллюзорных корреляций в принятии решений обычными людьми и специалистами. В части VI обсуждается проверка вероятностных оценок и обосновывается распространенное явление самоуверенности в прогнозировании и объяснении. Предубеждения, связанные с многоступенчатым выводом рассматриваются в части VII. Часть VIII рассматривает формальные и неформальные процедуры для коррекции и улучшения интуитивного принятия решений. Часть IX суммирует изучение последствий стереотипов при принятии решений в области риска. Заключительная часть содержит некоторые современные мысли по поводу нескольких концептуальных и методологических проблем по изучению эвристики и предубеждений.

Для удобства все ссылки собраны в отдельном списке в конце книги. Номера, напечатанные жирным шрифтом, относятся к материалу, включенному в книгу, обозначая главу, в которой данный материал появляется. Мы использовали скобки (...) чтобы указать на удаленный материал из предварительно изданных статей.

Наша работа по подготовке этой книги была поддержана Службой военно-морс-

ких исследований, Грант N00014-79-C-0077 Стэнфордскому Университету и Службой военно-морских исследований, Контракт N0014-80-C-0150 по исследованию принятия решений.

Мы хотим поблагодарить Пэгги Рокер, Нэнси Коллинс, Джерри Хенсона и Дона МакГрегора за их помощь в подготовке этой книги.

Даниель Канеман

Пол Словик

Амос Тверски

Часть I
Введение

1. Принятие решений в условиях неопределенности: правила и предубеждения*

Амос Тверски и Даниель Канеман

Многие решения основаны на убеждениях о вероятности неопределенных событий, таких как, например, результат выборов, вина ответчика в суде или будущий курс доллара. Эти убеждения обычно выражаются в заявлениях типа “я думаю, что...”, “вероятность такова...”, “маловероятно, что...” и т.д. Иногда убеждения относительно неопределенных событий выражены в численной форме как шансы или субъективные вероятности. Что определяет такие убеждения? Как люди оценивают вероятность неопределенного события или значения неопределенной величины? Этот раздел показывает, что люди полагаются на ограниченное число эвристических принципов, которые сводят сложные задачи оценки вероятностей и прогнозирования значений величин до более простых операций суждения. Вообще, эти эвристики весьма полезны, но иногда они ведут к серьезным и систематическим ошибкам.

Субъективная оценка вероятности похожа на субъективную оценку физических величин, таких как расстояние или размер. Все эти оценки основаны на данных ограниченной достоверности, которые обработаны согласно правилам эвристики. Например, предположительное расстояние до объекта частично определяется его четкостью. Чем резче объект, тем ближе он кажется. Это правило имеет некоторое обоснование, потому что на любой местности более удаленные объекты кажутся менее четкими, чем более близкие объекты. Однако постоянное следование этому правилу ведет к систематическим ошибкам в оценке расстояния. Характерно, что при плохой видимости расстояния часто переоцениваются, потому что контуры объектов размыты. С другой стороны, расстояния часто недооцениваются, когда видимость хорошая, потому что объекты кажутся более четкими. Таким образом, использование четкости в качестве показателя расстояния ведет к распространенным предубеждениям. Такие предубеждения также можно

*Эта глава впервые появилась в *Science*, 1974, 185, 1124-1131. Авторское право © 1974 принадлежит Американской Ассоциации достижений науки. Переиздано в соответствии с разрешением.

обнаружить в интуитивной оценке вероятности. В этой книге описываются три вида эвристик, которые используются для оценки вероятности и прогнозирования значений величин. Приводятся предубеждения, к которым ведут эти эвристики, а также обсуждается практическое и теоретическое значение этих наблюдений.

Репрезентативность

Большинство вопросов о вероятности принадлежат к одному из следующих типов: Какова вероятность того, что объект А принадлежит классу В? Какова вероятность, что причиной события А является процесс В? Какова вероятность, что процесс В приведет к событию А? Отвечая на такие вопросы, люди обычно полагаются на эвристику репрезентативности, в которой вероятность определяется степенью, в которой А репрезентативно по отношению к В, то есть степенью, в которой А похоже на В. Например, когда А в высокой мере репрезентативно В, вероятность того, что событие А происходит из В, оценивается как высокая. С другой стороны, если А не похоже на В, то вероятность оценивается как низкая.

Для иллюстрации суждения репрезентативности, рассмотрим описание человека его бывшим соседом: “Стив очень замкнутый и застенчивый, всегда готов мне помочь, но слишком мало интересуется другими людьми и реальностью вообще. Он очень кроткий и опрятный, любит порядок и систематизированность, а также склонен к детализации.” Как люди оценивают вероятность того, кто Стив по профессии (например, фермер, продавец, пилот самолета, библиотекарь или врач)? Каким образом люди располагают эти профессии от наиболее до наименее вероятной? В эвристике репрезентативности, вероятность того, что Стив – библиотекарь, например, определяется степенью, в которой он репрезентативен библиотекарю, или соответствует стереотипу библиотекаря. Действительно, исследование подобных проблем показало, что люди распределяют профессии точно таким же образом (Kahneman и Tversky, 1973, 4). Этот подход к оценке вероятности приводит к серьезным ошибкам, потому что на подобие или репрезентативность не оказывают влияние отдельные факторы, которые должны влиять на оценку вероятности.

Нечувствительность к априорной вероятности результата

Одним из факторов, которые не оказывают влияния на репрезентативность, но значительно влияют на вероятность — является предшествующая вероятность, или частота базовых значений результатов (исходов). В случае Стива, например, тот факт, что среди населения намного больше фермеров, чем библиотекарей, обязательно принимается в расчет при любой разумной оценке вероятности того, что Стив скорее является библиотекарем, чем фермером. Принятие во внимание частоты базовых значений, однако, в действительности не влияет на соответствие Стива стереотипу библиотека-

рей и фермеров. Если люди оценивают вероятность посредством репрезентативности, следовательно, предшествующими вероятностями они будут пренебрегать. Эта гипотеза была проверена в эксперименте, в котором изменялись предшествующие вероятности (Kahneman и Tversky, 1973, 4). Испытуемым показывали краткие описания нескольких людей, выбранных наугад из группы 100 специалистов – инженеров и адвокатов. Тестируемых просили оценить, для каждого описания, вероятность того, что оно принадлежит скорее инженеру, чем адвокату. В одном экспериментальном случае, испытуемым сообщалось, что группа, описания из которой были даны, состоит из 70 инженеров и 30 адвокатов. В другом случае испытуемым сообщалось, что группа состоит из 30 инженеров и 70 адвокатов. Шансы того, что каждое отдельное описание принадлежит скорее инженеру, чем адвокату, должна быть выше в первом случае, где большинство инженеров, чем во втором, где большинство адвокатов. Это можно показать, применяя правило Байеса заключающееся в том, что пропорция этих шансов должна быть $(0.7/0.3)^2$, или 5.44, для каждого описания. Грубо нарушая правило Байеса, испытуемые в обоих случаях, продемонстрировали, в сущности, одинаковые оценки вероятности. Очевидно, участники эксперимента оценили вероятность того, что конкретное описание принадлежит скорее инженеру, чем адвокату как степень, в которой это описание было репрезентативно этим двум стереотипам, мало учитывая, если учитывая вообще, предшествующие вероятности этих категорий.

Испытуемые правильно использовали предшествующие вероятности, когда они не обладали иной информацией. В отсутствии краткого описания личности, они оценивали вероятность того, что неизвестный человек является инженером, как 0.7 и 0.3, соответственно, в обоих случаях, при обоих условиях частоты базовых значений. Однако предшествующие вероятности полностью игнорировались, когда было представлено описание, даже если оно было полностью неинформативно. Реакции на нижеизложенное описание иллюстрируют это явление:

Дик – 30-летний мужчина. Женат, еще не имеет детей. Очень способный и мотивированный сотрудник, подает большие надежды. Пользуется признанием коллег.

Это описание было задумано таким образом, чтобы не предоставить информации о том, является ли Дик инженером или адвокатом. Следовательно, вероятность того, что Дик является инженером, должна равняться пропорции инженеров в группе, как если бы не было дано описания вовсе. Испытуемые, однако, оценили вероятность того, что Дик является инженером, как 5 независимо от того, какая дана пропорция инженеров в группе (7 к 3 или 3 к 7). Очевидно, что люди реагируют по-разному в ситуациях, когда описание отсутствует и когда дано бесполезное описание. В случае, когда описания отсутствуют, предшествующие вероятности используются должным образом; и предшествующие вероятности игнорируются, когда дается бесполезное описание (Kahneman и Tversky, 1973, 4).

Нечувствительность к размеру выборки

Чтобы оценивать вероятность получения конкретного результата в выборке, отобранной из указанной совокупности, люди обычно применяют эвристику репрезентативности. То есть они оценивают вероятность результата в выборке, например, то, что средняя высота в случайной выборке из десяти человек будет 6 футов (180 сантиметров), в степени в какой этот результат подобен соответствующему параметру (то есть средней высоте людей среди всего населения). Подобие статистики в выборке типичному параметру у всего населения не зависит от размера выборки. Следовательно, если вероятность рассчитывается с помощью репрезентативности, то статистическая вероятность в выборке будет по существу независима от размера выборки. Действительно, когда тестируемые оценивали распределение средней высоты для выборок различных размеров, они производили идентичные распределения. Например, вероятность получения средней высоты более чем 6 футов (180 см) была оценена подобной для выборок 1. 000, 100 и 10 человек (Kahneman и Tversky, 1972b, 3). Кроме того, испытуемые не сумели оценить роль размера выборки даже, когда это было подчеркнуто в формулировке проблемы. Приведем пример, подтверждающий это.

Некоторый город обслуживается двумя больницами. В большей по размеру больнице рождаются приблизительно 45 младенцев каждый день, а в меньшей больнице, приблизительно 15 младенцев каждый день. Как Вы знаете, приблизительно 50% от всех младенцев – мальчики. Однако точный процент меняется со дня на день. Иногда он может быть выше, чем 50%, иногда ниже.

В течение одного года, каждая больница делала учет тех дней, когда больше чем 60% рожденных младенцев были мальчиками. Какая больница, по вашему мнению, сделала учет большего количества таких дней?

Большая по размеру больница (21)

Меньшая больница (21)

Примерно поровну (то есть в пределах разницы в 5%) (53)

Числа в круглых скобках указывают количество отвечавших студентов последних курсов.

Большинство тестируемых оценивало вероятность того, что будет более 60% мальчиков в равной степени и в маленькой, и в большой больнице, возможно, потому что эти события описаны одинаковой статистикой и, таким образом, одинаково репрезентативны по отношению ко всему населению. Напротив, согласно теории выборок, ожидаемое число дней, в которые больше чем 60% рожденных младенцев являются мальчиками, намного выше в маленькой больнице, чем в большой, потому что для большой выборки менее вероятно, отклонение от 50%. Это фундаментальное понятие статистики, очевидно, не является частью интуиции людей.

Подобная нечувствительность к размеру выборки была зафиксирована в оценках *апостериорной* (*posteriori*) вероятности, то есть вероятности, что выборка была отобрана скорее из одной совокупности, чем из другой. Рас-

смотрим следующий пример:

Вообразите корзину, наполненную шарами, из которых $\frac{2}{3}$ одного цвета и $\frac{1}{3}$ другого. Один человек вынимает из корзины 5 шаров и обнаруживает, что 4 из них красные, а 1 – белый. Другой человек вынимает 20 шаров и обнаруживает, что 12 из них красные, а 8 – белые. Который из этих двух людей должен с большей уверенностью сказать, что в корзине скорее $\frac{2}{3}$ красных шаров и $\frac{1}{3}$ белых шаров, чем наоборот? Каковы шансы у каждого из этих людей?

В этом примере, правильным ответом является оценка последующих шансов как 8 к 1 для выборки 4:1 и 16 к 1 для выборки 12:8, при условии, что предшествующие вероятности равны. Однако большинство людей думает, что первая выборка обеспечивает намного более серьезное подтверждение для гипотезы, что корзина наполнена в основном красными шарами, потому что процентное отношение красных шаров в первой выборке больше, чем во второй. Это снова показывает, что интуитивные оценки превалируют за счет пропорции выборки, а не ее размера, который играет решающую роль в определении реальных последующих шансов. (Kahneman и Tversky, 1972b). Кроме того, интуитивные оценки последующих шансов (posterior odds) гораздо менее радикальны, чем правильные значения. В проблемах этого типа неоднократно наблюдалась недооценка влияния очевидного (W. Edwards, 1968, 25; Slovic и Lichtenstein, 1971). Этот феномен назван «консерватизмом».

Ошибочные концепции шанса

Люди полагают, что последовательность событий, организованная как случайный процесс представляет существенную характеристику этого процесса даже, когда последовательность короткая. Например, относительно выпадения монеты «орлом» или «решкой», люди считают, что последовательность O-P-O-P-P-O, более вероятна, чем последовательность O-O-O-P-P-P, которая не кажется случайной, а также более вероятна, чем последовательность O-O-O-O-P-O, которая не отражает равнозначность сторон монеты (Kahneman и Tversky, 1972b, 3). Таким образом, люди ожидают, что существенные характеристики процесса будут представлены, не только *глобально*, т.е. в полной последовательности, но также и *локально* – в каждой из ее частей. Однако локально репрезентативная последовательность систематически отклоняется от ожидания шансов, на которые рассчитывали: в ней слишком много чередований и слишком мало повторений. Другое последствие убеждения по поводу репрезентативности — хорошо известная ошибка игрока в казино. Видя, что красные слишком долго выпадают на колесе рулетки, например большинство людей, ошибочно полагает, что, скорее всего, теперь должно выпасть черное, потому что выпадение черного завершит более репрезентативную последовательность, чем выпадение еще одного красного. Шанс обычно рассматривается как саморегулирующийся

процесс, в котором отклонение в одном направлении приводит к отклонению в противоположном направлении с целью восстановления равновесия. На самом деле отклонения не исправляются, а просто «растворяются» по мере протекания случайного процесса.

Неправильные представления о шансе характерны не только для неопытных тестируемых. Изучение интуиции в статистических предположениях у опытных психологов-теоретиков (Tversky и Kahneman, 1971, 2) показало устойчивое верование в то, что можно назвать законом малых чисел, согласно которому даже маленькие выборки являются высоко репрезентативными по отношению к совокупностям, из которых они отобраны. Результаты этих исследователей отразили ожидание того, что гипотеза, достоверная относительно всей совокупности, будет представлена как статистически значимый результат в выборке, причем размер выборки не имеет значения. Как следствие, специалисты слишком верят в результаты, полученные на маленьких выборках, и слишком переоценивают повторяемость этих результатов. При проведении исследования, это предубеждение ведет к отбору выборок неадекватного размера и к преувеличенной интерпретации результатов.

Нечувствительность к надежности прогноза

Люди иногда вынуждены делать числовые предсказания, такие как будущий курс акции, спрос на товар или результат футбольной игры. Такие предсказания основываются на репрезентативности. Например, предположим, некто получил описание компании, и его просят предсказать ее будущую прибыль. Если описание компании очень благоприятно, то по этому описанию наиболее репрезентативной будет казаться очень высокая прибыль; если описание посредственно – то наиболее репрезентативным будет казаться заурядное развитие событий. То, насколько описание является благоприятным, не зависит от достоверности этого описания или степени, в которой оно позволяет проводить точное прогнозирование.

Следовательно, если люди делают прогноз, исходя исключительно из благоприятности описания, их предсказания будут нечувствительны к надежности описания и к ожидаемой точности предсказания.

Этот способ делать суждения нарушает нормативную статистическую теорию, в которой экстремум и диапазон предсказаний зависит от предсказуемости. Когда предсказуемость равна нулю, во всех случаях должно быть сделано одно и то же предсказание. Например, если описания компаний не содержат информации относительно прибыли, тогда для всех компаний должна быть спрогнозирована одна и та же величина (в размере среднего значения прибыли). Если прогнозируемость идеальна, конечно, предсказываемые величины будут соответствовать фактическим величинам, и диапазон прогнозов будет равняться диапазону результатов. Вообще, чем выше предсказуемость, тем шире диапазон предсказанных величин.

Некоторые исследования численного прогнозирования показали, что

интуитивные предсказания нарушают это правило, и что испытуемые мало учитывают, если учитывают вообще, соображения прогнозируемости (Kahneman и Tversky, 1973, 4). В одном из этих исследований, испытуемым было предоставлено несколько абзацев текста, каждый из которых описывал работу преподавателя ВУЗа в течение отдельно взятого практического занятия. Некоторых тестируемых попросили оценить качество урока, описанного в тексте с помощью процентной шкалы, относительно указанной совокупности. Других тестируемых просили спрогнозировать, также используя процентную шкалу, положение каждого преподавателя ВУЗа через 5 лет после этого практического занятия. Суждения, сделанные при обоих условиях были идентичными. То есть предсказание отдаленного во времени критерия (успех преподавателя через 5 лет) было идентично оценке информации, на основе которой это предсказание было сделано (качество практического занятия). Студенты, которые предположили это, несомненно, знали о том, насколько ограничена прогнозируемость компетентности преподавателя, основанная на единственном пробном уроке, проведенного 5 годами ранее; однако их прогнозы были столь же крайними, сколь и их оценки.

Иллюзия валидности

Как мы уже говорили, люди часто делают прогнозы, выбирая исход (например, профессию), который является наиболее репрезентативным по отношению к входным данным (например, описание человека). То, насколько они уверены в своем прогнозе, зависит, прежде всего, от степени репрезентативности (то есть качества соответствия выбора входным данным) безотносительно факторов, которые ограничивают точность их прогноза. Таким образом, люди вполне уверены в прогнозе, что человек является библиотекарем, когда дано описание его личности, которое соответствует стереотипу библиотекаря, даже если оно скудно, ненадежно или устарело. Необоснованная уверенность, которая является следствием удачного совпадения предсказываемого результата и входных данных, может называться *иллюзией валидности*. Эта иллюзия сохраняется даже тогда, когда испытуемый знает факторы, которые ограничивают точность его прогнозов. Вполне распространенным является высказывание, что психологи, которые проводят выборочные интервью, часто обладают значительной уверенностью по поводу своих прогнозов, даже если они знакомы с обширной литературой, которая показывает что в выборочных интервью высока вероятность ошибок. Продолжительная уверенность в правильности результатов клинического выборочного интервью, несмотря на повторные свидетельства его неадекватности, является достаточным свидетельством силы этого эффекта.

Внутренняя согласованность образца входных данных является основным показателем степени уверенности в прогнозе, основанном на этих входных данных. Например, люди выражают больше уверенности в прогнозе средне-

го балла успеваемости студента, чей табель за первый год обучения состоит полностью из В (4 балла), чем в прогнозе среднего балла студента, в чьем табеле за первый год обучения много оценок, как А (5 баллов), так и С (3 балла). Высоко согласованные образцы наиболее часто наблюдаются, когда входные переменные очень избыточны или взаимосвязаны. Следовательно, люди склонны быть уверенными в прогнозах, основанных на избыточных входных переменных. Однако элементарное правило в статистике корреляции, утверждает, что, если у нас есть входные переменные определенной валидности, прогноз, основанный на нескольких таких входных данных, может достигать более высокой точности, когда переменные независимы друг от друга, чем если они являются избыточными или взаимосвязанными. Таким образом, избыточность входных данных уменьшает точность, даже если она увеличивает уверенность, таким образом, люди часто уверены в прогнозах, которые, скорее всего, будут ошибочными (Kahneman и Tversky, 1973, 4).

Неправильные представления о регрессии

Предположим, что большая группа детей была протестирована с помощью двух подобных версий теста на способности. Если некто отберет десять детей из числа тех, кто справился лучше всех с одной из этих двух версий, он обычно будет разочарован выполнением ими второй версии теста. И напротив, если некто отберет десять детей из числа тех, кто хуже всех справился с первой версией теста, то в среднем он обнаружит, что с другой версией они справились несколько лучше. Обобщая, рассмотрим две переменные X и Y , которые имеют одинаковое распределение. Если выбрать людей, чьи средние X оценки отклоняются от среднего X на k единиц, тогда среднее от их Y шкалы будет обычно отклоняться от среднего Y меньше чем на k единиц. Эти наблюдения иллюстрируют общее явление известное как регресс к среднему, которое было открыто Гальтоном более чем 100 лет назад.

В обычной жизни все мы сталкиваемся с большим количеством случаев регресса к среднему, сравнивая, например, рост отцов и сыновей, уровень интеллекта мужей и жен, или результаты сдачи экзаменов, следующих один за другим. Тем не менее, у людей отсутствуют предположения по этому поводу. Во-первых, они не ожидают регрессии во многих контекстах, где она должна произойти. Во-вторых, когда они признают возникновение регрессии, они часто изобретают неверные объяснения причин. (Kahneman и Tversky, 1973, 4). Мы полагаем, что явление регресса остается неуловимым, потому что оно несовместимо с мнением о том, что прогнозируемый результат должен быть максимально репрезентативен по отношению к входным данным, и, следовательно, значение переменной результата должно быть настолько же крайним, как и значение входной переменной.

Неспособность признать смысл регрессии может иметь пагубные последствия, как было проиллюстрировано в следующих наблюдениях (Kahneman

и Tversky, 1973, 4). При обсуждении учебных полетов, опытные инструкторы отметили, что похвала за исключительно мягкое приземление обычно при следующей попытке сопровождается более неудачным приземлением, в то время как резкая критика после жесткого приземления обычно сопровождается улучшением результатов при следующей попытке. Инструкторы сделали вывод, что словесные поощрения вредны для обучения, в то время как выговоры приносят пользу, вопреки принятой психологической доктрине. Это заключение несостоятельно из-за присутствия регресса к среднему. Как и в других случаях, когда экзамены следуют один за другим, улучшение обычно следует за плохим выполнением работы, а ухудшение – за отлично проделанной работой, даже если преподаватель или инструктор никак не реагирует на достижения учащегося при первой попытке. Поскольку инструкторы похвалили своих учеников после хороших приземлений и поругали их после плохих, они пришли к ошибочному и потенциально вредному заключению, что наказание является более эффективным, чем награда.

Таким образом, неспособность понимать эффект регрессии ведет, к тому, что эффективность наказания оценивается слишком высоко, а эффективность награды недооценивается. В социальном взаимодействии, также, как и в обучении, награда обычно применяется, когда работа выполнена хорошо, и наказание, когда работа выполнена плохо. Следуя только закону регрессии, поведение, вероятней всего, улучшится после наказания и, скорее всего, ухудшится после награды. Следовательно, выходит так, что, по чистой случайности, людей поощряют за то, что они наказывают других, и наказывают за их поощрение. Люди, в целом, не знают об этом обстоятельстве. Фактически, неуловимая роль регрессии в определении очевидных последствий награды и наказания, кажется, ускользнула от внимания ученых, работающих в этой области.

Доступность

Бывают ситуации, в которых люди оценивают частоту класса или вероятность событий на основе легкости, с которой они вспоминают примеры случаев или события. Например, можно оценивать вероятность риска сердечного приступа среди людей средних лет, вспоминая такие случаи среди своих знакомых. Похожим образом некто может оценивать вероятность того, что некоторое бизнес-предприятие потерпит неудачу, представляя себе различные трудности, с которыми оно могло бы столкнуться. Эта эвристика оценки называется *доступностью*. Доступность очень полезна для оценки частоты или вероятности событий, потому что события, принадлежащие большим классам, обычно вспоминаются и быстрее, чем случаи менее частых классов. Однако на доступность воздействуют и другие факторы, кроме частоты и вероятности. Следовательно, уверенность относительно доступности ведет к вполне прогнозируемым предубеждениям, некоторые из которых проиллюстрированы ниже.

Предубеждения, обусловленные степенью восстанавливаемости событий в памяти

Когда размер класса оценивается на основе доступности его элементов, класс, элементы которого легко восстанавливаются в памяти, будет казаться более многочисленным, чем класс такого же размера, но элементы которого, менее доступны и хуже вспоминаются. При простой демонстрации этого эффекта, испытуемым зачитали список известных людей обоих полов, и затем попросили оценить, было ли в списке больше мужских имен, чем женских. Различные списки были предоставлены разным группам тестируемых. В некоторых из списков мужчины были более известны, чем женщины, а в других, женщины были более известны, чем мужчины. В каждом из списков, испытуемые ошибочно считали, что класс (в данном случае пол), в котором были более известные люди, был более многочисленным (Tversky и Kahneman, 1973,11).

В дополнение к узнаваемости, имеются другие факторы, такие как яркость, которая влияет на восстанавливаемость событий в памяти. Например, если человек наблюдал воочию пожар в здании, то он будет считать возникновение таких несчастных случаев, наверное, более субъективно вероятным, чем, если бы, он прочитал об этом пожаре в местной газете. Кроме того, недавние происшествия, вероятно, будут вспоминаться несколько проще, чем более ранние. Часто бывает, что субъективная оценка вероятности возникновения дорожных происшествий временно повышается, когда человек видит около дороги перевернутый автомобиль.

Предубеждения, обусловленные эффективностью направления поиска

Предположим, из английского текста наугад выбрано слово (из трех букв или больше). Что является более вероятным, то, что слово начинается с буквы г или что г является третьей буквой? Люди подходят к решению этой проблемы, вспоминая слова, которые начинаются с г (road – дорога), и слова, которые имеют г в третьей позиции (car – автомобиль), и оценивают относительную частоту, основываясь на легкости, с которой слова этих двух типов приходят на ум. Поскольку гораздо легче искать слова по первой букве, чем по третьей, большинство людей считают, что больше слов, которые начинаются с этой согласной, чем слов, в которых тот же самый согласный появляется в третьей позиции. Они делают такой вывод даже для таких согласных, как г или к, которые чаще появляются в третьей позиции, чем в первой (Tversky и Kahneman, 1973,11).

Различные задачи требуют различных направлений поиска. Например, предположим, Вас попросили оценить частоту, с которой слова с абстрактным значением (мысль, любовь) и конкретным значением (дверь, вода) появляются в письменном английском языке. Естественный способ ответить

на этот вопрос – найти контексты, в которых эти слова могли бы появляться. Кажется, легче вспомнить контексты, в которых может быть упомянуто абстрактное значение (любовь в женских романах), чем вспомнить контексты, в которых упоминается слово с конкретным значением (например, дверь). Если частота слов определяется на основании доступности контекстов, в которых они появляются, слова с абстрактным значением, будут оценены как относительно более многочисленные, чем слова с конкретным значением. Этот стереотип наблюдался в недавнем исследовании (Galbraith и Underwood, 1973), которое показало, что частота возникновения слов с абстрактным значением была намного выше частоты слов с конкретным значением, в то время как их объективная частота равна. Оценивалось также, что абстрактные слова появлялись в намного большем разнообразии контекстов, чем слова с конкретным значением.

Предубеждения, обусловленные способностью к представлению образов

Иногда нужно оценить частоту класса, элементы которого не хранятся в памяти, а могут быть созданы согласно определенному правилу. В таких ситуациях, обычно воспроизводятся некоторые элементы, а частота или вероятность оценивается той легкостью, с которой могут быть построены соответствующие элементы. Однако легкость воспроизведения соответствующих элементов не всегда отражает их фактическую частоту, и этот способ оценки приводит к предубеждениям. Для иллюстрации этого рассмотрим группу из 10 человек, которые формируют комитеты, состоящие из k членов, причем $2 \leq k \leq 8$. Сколько различных комитетов, состоящих из k членов может быть сформировано? Правильный ответ на эту проблему дается биномиальным коэффициентом (k^{10}), который достигает максимума, равного 252 для $k = 5$. Ясно, что число комитетов, состоящих из k членов, равняется числу комитетов, состоящих из $(10 - k)$ членов, потому что для любого комитета, состоящего из k членов, существует единственно возможная группа, состоящая из $(10 - k)$ человек, не являющихся членами комитета.

Одним из способов ответить без вычисления – это мысленно создать комитеты, состоящие из k членов, и оценить их количество, используя легкость, с которой они приходят на ум. Комитеты, состоящие из небольшого количества членов, например, 2, более доступны, чем комитеты, состоящие из большого количества членов, например, 8. Самая простая схема создания комитетов – разделение группы на непересекающиеся множества. Сразу видно, что легче создать пять непересекающихся комитетов по 2 члена в каждом, в то время как невозможно сгенерировать и двух непересекающихся комитетов по 8 членов. Следовательно, если частота оценивается за счет способности представить это, или доступностью к мысленному воспроизведению, будет казаться, что маленьких комитетов больше, чем больших, в отличие от правильной параболической функции. Действительно, когда

тестируемых-неспециалистов просили оценить число различных комитетов разных размеров, их оценки представляли собой монотонно-убывающую функцию от размера комитета (Tversky и Kahneman, 1973, 11). Например, средняя оценка числа комитетов, состоящих из 2 членов, была 70, в то время как оценка для комитетов, состоящих из 8 членов — 20 (правильный ответ — 45 в обоих случаях).

Способность представлять образы играет важную роль в оценке вероятностей возникновения реальных жизненных ситуаций. Риск, с которым связана опасная экспедиция, например, оценивается, посредством мысленного воспроизведения непредвиденных обстоятельств, для преодоления которых у экспедиции нет достаточного оборудования. Если многие из таких трудностей ярко изображаются, экспедиция может показаться чрезвычайно опасной, хотя легкость, с которой воображаются бедствия, вовсе не обязательно отражает их фактическую вероятность. Наоборот, если возможную опасность трудно вообразить, или она просто не приходит на ум, риск, связанный с каким-либо событием, может быть чрезвычайно недооценен.

Иллюзорная взаимосвязь

Чепмен и Чепмен (Chapman and Chapman, 1969) описали интересное предубеждение в оценке частоты, с которой два события произойдут одновременно. Они предоставили испытуемым-неспециалистам информацию относительно нескольких гипотетических пациентов с психическими расстройствами. Данные по каждому пациенту включали клинический диагноз и рисунки пациента. Позже испытуемые оценили частоту, с которой каждый диагноз (такой как паранойя или мания преследования) сопровождался различными особенностями рисунка (специфической формой глаз). Испытуемые заметно переоценили частоту совместного появления двух естественных событий, таких как мания преследования и специфическая форма глаз. Это явление получило название иллюзорная корреляция. В ошибочных оценках представленных данных, испытуемые «заново открыли» многое из уже известных, но необоснованных, клинических знаний относительно интерпретации рисуночного теста. Иллюзорный эффект корреляции был чрезвычайно стойкий по отношению к противоречащим данным. Он сохранился даже, когда взаимосвязь между признаком и диагнозом была фактически негативной, что не позволило испытуемым определить действительные отношения между ними.

Доступность является естественным объяснением эффекта иллюзорной корреляции. Оценка того, насколько часто два явления взаимосвязаны и происходят одновременно, может быть основано на силе ассоциативной связи между ними. Когда ассоциация сильна, можно, скорее всего, прийти к выводу, что события часто происходили одновременно. Следовательно, если ассоциация между событиями крепкая, то, по оценке людей, они будут часто происходить одновременно. Согласно этой точке зрения, иллюзорная кор-

реляция между диагнозом мании преследования и специфической формой глаз на рисунке, к примеру, появляется из-за того, что мания преследования скорее ассоциируется с глазами, чем с любой другой частью тела.

Продолжительный жизненный опыт научил нас, что, в общем, элементы больших классов вспоминаются лучше и быстрее, чем элементы менее частотных классов; что более вероятные события легче вообразить, чем маловероятные; и что ассоциативные связи между событиями укрепляются, когда события часто происходят одновременно. В результате, человек получает в свое распоряжение процедуру (эвристику доступности) для оценки размера класса, вероятность события, или частота, с которой могут одновременно происходить события, оцениваются той легкостью, с которой могут быть выполнены соответствующие ментальные процессы вспоминания, воспроизведения или ассоциации. Однако, как показали предшествующие примеры, эти процедуры оценивания систематически приводят к ошибкам.

Корректировка и «привязка» (anchoring)

Во многих ситуациях, люди делают оценки, отталкиваясь от начальной величины, которая специально подобрана таким образом, чтобы получить окончательный ответ. Начальная величина или отправная точка, может быть получена посредством формулировки проблемы, или она может быть частично результатом вычисления. В любом случае, такой «прикидки» обычно недостаточно (Slovic и Lichtenstein, 1971). То есть, различные отправные точки приводят к различным оценкам, которые смещены к этим отправным точкам. Мы называем этот феномен «*привязкой*» (anchoring).

Недостаточная «корректировка»

Для демонстрации эффекта «привязки», тестируемых просили оценить различные величины, выраженные в процентах (например, процент Африканских стран в Организации Объединенных Наций). Каждой величине с помощью случайного выбора в присутствии тестируемых, был присвоен номер от 0 до 100. Тестируемых вначале попросили указать, больше или меньше этот номер, чем значение самой величины, и затем оценить значение этой величины, двигаясь в большую или меньшую сторону относительно его номера. Различным группам тестируемых предлагались различные номера для каждой величины, и эти произвольные номера имели значительное влияние на оценки тестируемых. Например, средние оценки процента Африканских стран в Организации Объединенных Наций были 25 и 45 для групп, которые получили 10 и 65 в качестве отправных точек соответственно. Денежные вознаграждения за точность не уменьшали эффект «привязки».

«Привязка» происходит не только, когда испытуемому задана отправная точка, но также когда испытуемый основывает свою оценку на результате некоторого неполного вычисления. Изучение интуитивной числовой оцен-

ки иллюстрирует этот эффект. Две группы студентов средней школы оценивали, в течение 5 секунд, значение числового выражения, которое было написано на доске. Одна группа оценивала значение выражения

$$8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1,$$

в то время как другая группа оценивала значение выражения

$$1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8.$$

Для быстрого ответа на такие вопросы, люди могут произвести несколько шагов вычисления и оценить значение выражения с помощью экстраполяции или «корректировки». Поскольку «корректировки» обычно недостаточно, эта процедура должна вести к недооценке значения. Более того, так как результат первых немногих шагов умножения (выполненный слева направо) выше в убывающей последовательности, чем в возрастающей, первое упомянутое выражение должно быть оценено больше последнего. Оба предсказания были подтверждены. Средняя оценка для возрастающей последовательности была 512, в то время как средняя оценка для убывающей последовательности была 2. 250. Правильный ответ – 40. 320 для обеих последовательностей.

Предубеждения в оценке конъюнктивных и дизъюнктивных событий

В недавнем исследовании Бар-Хиллел (Bar-Hillel, 1973) тестируемым давали возможность сделать ставку на одно из двух событий. Использовались три типа событий: (i) простое событие, как, например, вытягивание красного шара из мешка, содержащего 50% красных и 50% белых шаров; (ii) связанное событие, как, например, вытянуть красный шар семь раз подряд из мешка (с возвращением шаров), содержащего 90% красных шаров и 10% белых; и (iii) несвязанное событие, как, например, вытянуть красный шар, по крайней мере, 1 раз в семи последовательных попытках (с возвращением шаров) из мешка, содержащего 10% красных шаров и 90% белых. В этой проблеме, существенное большинство тестируемых предпочло сделать ставку на связанное событие (вероятность которого - 0.48), а не на простое (вероятность которого - 0.50). Испытуемые также предпочитали, ставить скорее на простое событие, чем на дизъюнктивное, которое имеет вероятность 0.52. Таким образом, большинство тестируемых сделало ставку на менее вероятное событие при обоих сравнениях. Эти решения тестируемых иллюстрируют общий вывод: изучение решений в азартных играх и оценки вероятности указывают, что люди имеют тенденцию оценивать слишком высоко вероятность конъюнктивных событий (Cohen, Chesnick и Haran, 1972, 24) и склонны недооценивать вероятность дизъюнктивных событий. Эти стерео-

типы полностью объясняются эффектом «привязки». Установленная вероятность элементарного события (успех в любой стадии) обеспечивает естественную отправную точку для оценки вероятностей как конъюнктивных, так и дизъюнктивных событий. Так как «корректировки» от отправной точки обычно недостаточно, заключительные оценки остаются слишком приближенными к вероятностям элементарных событий в обоих случаях. Обратите внимание, что полная вероятность конъюнктивных событий ниже, чем вероятность каждого элементарного события, в то время как полная вероятность несвязанного события выше, чем вероятность каждого элементарного события. Следствием «привязки» является то, что полная вероятность будет завышена для конъюнктивных событий и занижена – для дизъюнктивных.

Предубеждения в оценке сложных событий особенно существенны в контексте планирования. Успешное завершение бизнес-предприятия, например, разработка нового продукта, обычно носит комплексный характер: чтобы предприятие преуспевало, каждое событие из ряда должно произойти. Даже, если каждое из этих событий весьма вероятно, полная вероятность успеха может быть довольно низкой, если количество событий большое. Общая тенденция оценивать слишком высоко вероятность конъюнктивных событий ведет к необоснованному оптимизму в оценке вероятности, что план будет удачным, или что проект будет закончен вовремя. Наоборот, с дизъюнктивными структурами событий обычно сталкиваются при оценке риска. Сложная система, такая как ядерный реактор или тело человека, повредится, если любой из его необходимых компонентов выйдет из строя. Даже, когда вероятность сбоя в каждом компоненте небольшая, вероятность отказа всей системы может быть высока, если в нее вовлечено много компонентов. Из-за предубеждения «привязки», люди имеют тенденцию недооценивать вероятность отказа в сложных системах. Таким образом, предубеждение привязки может иногда зависеть от структуры события. Структура события или явления похожая на цепочку звеньев ведет к переоценке вероятности этого события, структура события похожая на воронку, состоящая из дизъюнктивных звеньев, ведет к недооценке вероятности события.

«Привязка» при оценке распределения субъективной вероятности

При анализе принятия решений, от экспертов часто требуется выразить свое мнение относительно какой-либо величины, например, среднего значения индекса Доу-Джонса (Dow-Jones) в отдельно взятый день, в виде распределения вероятности. Такое распределение обычно строится путем выбора значений для величины, которые соответствуют его процентной шкале распределения вероятности. Например, эксперта можно попросить выбрать число, X_{90} так, чтобы субъективная вероятность того, что это число будет выше, чем значение среднего числа Доу-Джонса, была 0.90. То есть, он должен выбрать значение X_{90} , так чтобы в 9 случаях к 1 среднее значение ин-

декса Доу-Джонса не превышало это число. Распределение субъективной вероятности значения среднего числа Доу-Джонса может быть построено из нескольких таких оценок, выраженных с помощью различных процентных шкал.

Путем накопления таких субъективных распределений вероятности для различных величин, можно проверить правильность оценок эксперта. Эксперт считается калиброванным (см. гл. 22) должным образом в определенном наборе проблем, если только 2 процента правильных значений оцененных величин будут ниже заданных значений X_2 . Например, правильные значения должны быть ниже X_{01} для 1% величин и быть выше X_{99} , для 1% величин. Таким образом, истинные значения должны строго попадать в интервал между X_{01} и X_{99} в 98% задач.

Несколько исследователей (Alpert и Raiffa, 1969, 21; Stael von Holstein, 1971b; Winkler, 1967) проанализировали нарушения в оценке вероятности для многих количественных величин для большого числа экспертов. Эти распределения указали на обширные и систематические отклонения от надлежащих оценок. В большинстве исследований, фактические значения оцененных величин или меньше X_{01} или больше, чем X_{99} приблизительно для 30% задач. То есть, испытуемые устанавливают слишком узкие строгие интервалы, которые отражают их уверенность, чем их знания относительно оцененных величин. Это предубеждение характерно как для подготовленных, так и для простых тестируемых, и этот эффект не устраняется путем введения правил оценки, которые обеспечивают стимулы для внешней оценки. Этот эффект, по крайней мере, частично, относится к «привязке».

Чтобы выбрать X_{90} как значение среднего числа Доу-Джонса, например, естественно начать с размышления о лучшей оценке индекса Доу-Джонса и «скорректировать» верхние значения. Если эта «корректировка» – как и большинство других – является недостаточной, то X_{90} не будет достаточно экстремальным. Подобный эффект фиксирования произойдет в выборе X_{10} , который предположительно получен путем корректировки чьей-либо лучшей оценки в сторону понижения. Следовательно, достоверный интервал между X_{10} и X_{90} будет слишком узкий, и оцененное распределение вероятности будет слишком жестким. В подтверждение этой интерпретации можно показать, что субъективные вероятности систематически меняются посредством процедуры, в которой чья-либо лучшая оценка не служит «привязкой».

Распределения субъективной вероятности для данной величины (среднее число Доу-Джонса) могут быть получены двумя различными способами: (i), попросить испытуемого выбрать значение числа Доу-Джонса, которое соответствует распределению вероятности выраженной с помощью процентной шкалы и (ii), попросить испытуемого оценить вероятности того, что истинное значение числа Доу-Джонса превысит некоторые указанные величины. Эти две процедуры формально эквивалентны и должны приводить в результате к идентичным распределениям. Однако они предлагают различные способы корректировки от различных «привязок». В процедуре (i),

естественная отправная точка – лучшая оценка качества. В процедуре (ii), с другой стороны, тестируемый может «привязаться» к величине, установленной в вопросе. В противоположность этому, он может «привязаться» к равным шансам, или к шансам 50 на 50, которые являются естественной отправной точкой при оценке вероятности. В любом случае, процедура (ii) должна завершаться менее крайними оценками, чем процедура (i).

Чтобы противопоставлять эти две процедуры, группе тестируемых был предоставлен набор 24 количественных измерений (как, например, расстояние по воздуху от Нью-Дели до Пекина), которые оценивали или X_{10} или X_{90} для каждой задачи. Другая группа тестируемых получила средние оценки первой группы для каждой из этих 24 величин. Их попросили оценить шансы того, что каждая из данных величин превысила истинное значение соответствующей величины. В отсутствии какого-либо предубеждения вторая группа должна восстановить вероятность, указанную первой группой, то есть 9:1. Однако если равные шансы или заданная величина служат «привязкой», вероятность, указанная второй группой должна быть менее экстремальной, то есть ближе к 1:1. В действительности, средняя вероятность, указанная этой группой, во всех задачах, была 3:1. Когда суждения из этих двух групп были проверены, обнаружилось, что испытуемые в первой группе были слишком экстремальны в оценках в соответствии с более ранними исследованиями. События, вероятность которых, они определили как 0.10, фактически произошли в 24% случаев. Напротив, тестируемые во второй группе были слишком консервативны. События, вероятность которых, они определили как 0.34, фактически произошли в 26% случаев. Эти результаты иллюстрируют то, как степень правильности оценки зависит от процедуры оценки.

Обсуждение

В этой части книги рассматривались когнитивные стереотипы, которые возникают как результат уверенности в эвристиках оценивания. Эти стереотипы не характерны для эффектов мотивации, таких как принятие желаемого за действительное или искажения суждений из-за одобрения и порицания. Действительно, как уже сообщалось ранее, некоторые серьезные ошибки в оценивании происходили, несмотря на тот факт, что тестируемых поощряли за точность и вознаграждали за правильные ответы (Kahneman и Tversky, 1972b, 3; Tversky и Kahneman, 1973, 11).

Уверенность в эвристиках и распространенность стереотипов свойственна не только обывателям. Опытные исследователи также склонны к тем же самым предубеждениям – когда они думают интуитивно. Например, тенденция предсказывать результат, который наиболее репрезентативен по отношению к данным, без достаточного внимания к априорной вероятности наступления такого результата, наблюдались в интуитивных суждениях людей, которые обладали обширными познаниями в статистике (Kahneman

и Tversky, 1973, 4; Tversky и Kahneman, 1971, 2). Хотя те, кто имеют познания в статистике и избегают элементарных ошибок, как, например, ошибки игрока в казино, в интуитивных суждениях для более запутанных и менее понятных задач делают подобные ошибки.

Не удивительно, что полезные разновидности эвристики, такие как репрезентативность и доступность сохраняются, даже при том, что они иногда ведут к ошибкам в прогнозах или оценках. Что возможно и является удивительным, так это неспособность людей вывести из длительного жизненного опыта такие фундаментальные статистические правила как регресс к среднему или эффект размера выборки при анализе изменчивости внутри выборки. Хотя все мы в течение жизни встречаемся с многочисленными ситуациями, к которым эти правила могут быть применимы, очень немногие самостоятельно открывают принципы отбора выборки и регресса на своем опыте. Статистические принципы не познаются на основе каждодневного опыта, потому что соответствующие примеры не закодированы нужным образом. Например, люди не обнаруживают, что средняя длина слова в строках следующих друг за другом в тексте, отличается больше чем на следующих друг за другом страницах, потому что они просто не обращают внимания на среднюю длину слова в отдельных строках или страницах. Таким образом, люди не изучают отношение между размером выборки и изменчивостью внутри выборки, хотя данных для такого вывода предостаточно.

Недостаток соответствующей кодировки также объясняет, почему люди обычно не обнаруживают стереотипы в своих суждениях о вероятности. Человек мог бы узнать, правильны ли его оценки, подсчитывая число событий, которые действительно происходят из тех, которые он считает равновероятными. Однако для людей не естественно группировать события по признаку их вероятности. При отсутствии такой группировки человек не может обнаружить, например, что только 50% предсказаний, вероятность которых он оценил как 0.9 или выше фактически сбылись.

Эмпирический анализ когнитивных стереотипов имеет значение для теоретической и прикладной роли оценки вероятностей. Современная теория принятия решений (de Finetti, 1968; Savage, 1954) рассматривает субъективную вероятность как количественное мнение идеализированного человека. Определенно, субъективная вероятность данного события определяется набором шансов относительно этого события, из которых человеку предлагается выбрать. Может быть получено внутренне последовательное или целостное измерение субъективной вероятности, если выборы человека среди предложенных шансов подчиняются некоторым принципам, то есть аксиомам теории. Полученная вероятность субъективна в том смысле, что различные люди могут иметь различные оценки вероятности одного и того же события. Главный вклад этого подхода состоит в том, что он обеспечивает строгую субъективную интерпретацию вероятности, которая является применимой к уникальным событиям и является частью общей теории рационального принятия решений.

Возможно, стоит отметить, что, в то время как субъективные вероятности могут иногда выводиться из выбора шансов, они обычно не формируются этим способом. Человек делает ставку скорее на команду А, чем на команду В, потому что верит, что команда А, вероятнее всего, победит; он не выводит свое мнение как результат предпочтений тех или иных шансов. Таким образом, в действительности, субъективные вероятности определяют предпочтения в шансах, но не выводятся из них, в отличие от аксиоматической теории рационального принятия решений (Savage, 1954).

Субъективный характер вероятности привел многих ученых к мнению, что целостность, или внутренняя последовательность – единственный имеющий силу критерий, в соответствии с которым должны быть оценены вероятности. С точки зрения формальной теории субъективной вероятности, любой набор внутренне последовательных вероятностных оценок столь же хорош как любой другой. Этот критерий не вполне удовлетворителен, потому что внутренне последовательный набор субъективных вероятностей может быть несовместим с другими мнениями, которых придерживается человек. Рассмотрите человека, чьи субъективные вероятности для всех возможных результатов подбрасывания монеты отражают ошибку игрока в казино. То есть его оценка вероятности появления «решки» при каждом конкретном подбрасывании увеличивается с числом последовательно выпавших «орлов», которые предшествовали этому подбрасыванию. Суждения такого человека могут быть внутренне последовательными и поэтому приемлемыми как адекватные субъективные вероятности согласно критерию формальной теории. Эти вероятности, однако, являются несовместимыми с общепринятым мнением, что у монеты “нет памяти” и поэтому она не способна производить последовательные зависимости. Чтобы оцененные вероятности, считались адекватными, или рациональными, внутренней последовательности недостаточно. Суждения должны быть совместимы со всеми прочими взглядами этого человека. К сожалению, не может быть простой формальной процедуры для оценки совместимости набора вероятностных оценок с полной системой взглядов субъекта. Рациональный эксперт будет, однако, бороться за совместимость, даже при том, что внутреннюю последовательность более легко достичь и оценить. В частности он будет пытаться делать свои вероятностные суждения совместимыми с его знаниями относительно предмета, законов вероятности и своей собственной эвристики оценки и предубеждений.

Резюме

Эта статья описывает три типа эвристики, которые используются при оценках в условиях неопределенности: (i) репрезентативность, которая обычно используется, когда людей просят оценить вероятность того, что объект А или случай принадлежит классу или процессу В; (ii) доступность событий или сценариев, которая часто используется, когда людей просят оценить

частоту класса или правдоподобия отдельно взятого варианта развития событий; и (iii) корректировка или «привязка», которая обычно используется при количественном прогнозировании, когда доступна соответствующая величина. Эти эвристики высоко экономичны и обычно эффективны, но они приводят к систематическим ошибкам в прогнозе. Лучшее понимание этих эвристик и отклонений, к которым они приводят, могло внести вклад в оценку и принятие решений в условиях неопределенности.

Часть II
Репрезентативность

2. Вера в закон малых чисел*

Амос Тверски и Даниель Канеман

“Предположим, что Вы провели эксперимент с 20 испытуемыми, и получили значимый результат, который подтверждает Вашу теорию ($z = 2.23$, $p < 0.05$, «двустороннее»). Вы теперь имеете основание для проведения эксперимента с дополнительной группой из 10 испытуемых. Как Вы думаете, какова вероятность того, что результаты будут значимыми, если будет проводиться испытание с односторонним критерием, причем отдельно для этой группы?”

Если Вы считаете, что вероятность где-то около 0.85, Вы можете быть удовлетворены тем, что принадлежите к большинству. Действительно, это был среднестатистический ответ двух малых групп, которые любезно согласились ответить на анкету, распространенную на встрече Группы математической психологии в Американской Психологической Ассоциации.

С другой стороны, если Вы считаете, что вероятность около 0.48, Вы принадлежите к меньшинству. Только 9 из наших 84 респондентов дали ответы между 0.40 и 0.60. Однако оказывается, что 0.48 — намного более обоснованная оценка, чем 0.85¹.

¹ Требуемую оценку можно интерпретировать несколькими способами. Один возможный подход состоит в том, чтобы следовать общей практике исследования, где оценка, полученная в одном исследовании, принимается для того, чтобы определить правдоподобную альтернативу отсутствующей гипотезе. Требуемую вероятность можно тогда интерпретировать как достоверность второго теста (то есть, вероятность получения значимого результата во второй выборке) в противоположность альтернативной гипотезе, определенной результатом первой выборки. В особом случае теста среднего значения с известной изменчивостью, можно было бы вычислить достоверность теста по отношению к гипотезе, что среднее для всей совокупности равняется среднему для выборки. Так как размер второй выборки — половина первой выборки, подсчитанная вероятность получения $z \geq 1.645$ — только 0.473. Теоретически более достоверный подход состоит в том, чтобы интерпретировать требуемую вероятность в пределах структуры Байеса, вычислить ее относительно соответственно выбранного предшествующего распределения. Допуская предыдущую равномерность распределения, требуемая вероятность в последующем — 0.478. Ясно, что, если предшествующее

Очевидно, большинство психологов преувеличенно верят в вероятность успешного повтора полученных результатов. Вопросы, которых касается эта часть книги – это источники такой уверенности, и их последствия для проведения научного исследования. Наш тезис состоит в том, что люди обладают сильными предубеждениями относительно случайной выборки; что эти предубеждения ошибочны фундаментально; что эти предубеждения характерны как для простых испытуемых, так и для подготовленных ученых; и что ее применение в ходе научного исследования имеет неудачные последствия.

Мы представляем на обсуждение тезис о том, что люди рассматривают выборку отобранную случайным образом из совокупности как высоко репрезентативную, то есть подобную всей совокупности во всех существенных характеристиках. Следовательно, они ожидают, что любые две выборки, взятые из ограниченной совокупности, будут более подобны друг другу и совокупности, чем предполагает теория выборок, по крайней мере, для малых выборок.

Тенденция расценивать выборку как репрезентативную наблюдается в самых разнообразных ситуациях. Когда тестируемых просят создать случайную последовательность гипотетических подбрасываний монеты, например, они создают последовательности, где пропорция «орла» на любом коротком отрезке гораздо ближе к 0.50, чем предсказала бы теория вероятности (Tune, 1964). Таким образом, каждый отрезок полученной последовательности высоко репрезентативен по отношению к «справедливости» монеты. Подобные эффекты наблюдаются, когда тестируемые последовательно предсказывают события в созданном случайным образом ряде событий, как в экспериментах по изучению вероятности (Estes, 1964) или в других последовательных играх с шансами. Тестируемые действуют так, как будто, *каждый* отрезок случайной последовательности должен отражать правильную пропорцию: если последовательность отклонилась от пропорции во всей совокупности, ожидается корректирующее отклонение в другом направлении. Это получило название ошибки игрока казино.

Суть ошибки игрока казино – неправильное представление о справедливости закона случайности. Игрок чувствует, что равнозначность сторон монеты дает ему право ожидать, что любое отклонение в одном направлении будет скоро компенсировано соответствующим отклонением в другую сторону. Даже самая сбалансированная монета, однако, обладая ограничениями морали и памяти, не может при подбрасывании выдавать столь же равновероятные результаты, как того ожидает игрок в казино. Эта ошибка свойственна не только игрокам. Рассмотрим следующий пример:

распределение подтверждает нулевую гипотезу, как это часто бывает, следующая вероятность будет даже меньшей.

Данная глава первоначально появилась в Психологическом Бюллетене, 1971, 2, 105 - 110. Авторское право © 1971 Американской Психологической Ассоциацией. Переиздано в соответствии с разрешением.

Средний IQ среди восьмиклассников в городе, как известно, является 100. Вы выбрали случайную выборку из 50 детей для изучения достижений в учебе. Первый протестированный ребенок имеет IQ 150. Каким, как Вы ожидаете, будет средний показатель интеллекта для всей выборки?

Правильный ответ – 101. Неожиданно большое количество людей полагают, что ожидаемый IQ для выборки все равно 100. Это может быть оправдано только мнением, что случайный процесс самокорректируется. Высказывания типа “ошибки компенсируют друг друга” отражают представление людей об активном процессе самокоррекции случайных процессов.

Некоторые распространенные процессы в природе подчиняются таким законам: отклонение от устойчивого равновесия порождает силу, которая восстанавливает равновесие. Законы вероятности, напротив, не работают подобным образом: отклонения не отменяются по мере перебора элементов выборки, они ослабляются.

До сих пор, мы пытались описать два взаимосвязанных вида предубеждений для определения шансов. Мы предложили гипотезу репрезентативности, согласно которой люди полагают, что выборки будут очень подобными друг другу и совокупностям, из которых они отобраны. Мы также предположили, что люди считают, что процессы в выборке — это самокорректирующийся процесс. Эти два мнения ведут к одним и тем же последствиям. Оба порождают ожидания относительно выборки, и изменчивость этих прогнозов меньше, чем реальная изменчивость, по крайней мере, для малых выборок.

Закон больших чисел гарантирует, что очень большие выборки будут действительно высоко репрезентативны по отношению к совокупности, из которой они взяты. Если, кроме того, действует тенденция саморегуляции, то маленькие выборки должны также быть высоко репрезентативными и подобными друг другу. Интуиция людей относительно случайных выборок, кажется, соответствует закону малых чисел, который утверждает, что закон больших чисел применяется также и к малым числам.

Рассмотрим гипотетического ученого, который живет в соответствии с законом малых чисел. Как его убеждения влияли бы на его научную работу? Предположим, что наш ученый изучает явления, величина которых маленькая относительно ее неконтролируемой изменчивости, то есть, соотношение сигнала и помехи в сообщениях, которые он получает от природы изучаемого явления, невелико. Наш ученый мог быть метеорологом, фармакологом, или возможно психологом.

Если ученый верит в закон малых чисел, то он будет переоценивать выводы, сделанные на маленьких выборках. Чтобы проиллюстрировать это, предположим, что он занят изучением того, какую из двух игрушек предпочли бы маленькие дети. Из первых пяти малышей, которых он изучил, четыре отдали предпочтение одной и той же игрушке. Немало психологов на этом этапе будут ощущать некоторую уверенность в том, что нулевая гипотеза о том, что никакой из игрушек не отдается предпочтение, ложна. К

счастью, такая убежденность не является достаточным условием для публикации статьи в журнале, хотя оно может подойти для книги. С помощью быстрых подсчетов, наш психолог обнаружит, что вероятность результата, который был бы столь же предельным, как полученный им, столь же высока как $3/8$ при нулевой гипотезе.

Безусловно, применение статистической проверки гипотезы для научного вывода связано с серьезными трудностями. Тем не менее, вычисление уровней значимости (или коэффициентов вероятности, как предпочел бы Байес) заставляет ученого оценивать полученный эффект скорее в терминах оценки надежности изменчивости выборки, чем в терминах своей субъективной оценки, в которой присутствуют отклонения. Статистические испытания, исходя из этого, защищают научное сообщество от чрезмерно поспешного отказа от нулевой гипотезы (ошибка первого рода) контролируя многих своих членов, которые предпочли бы жить в соответствии с законом малых чисел. С другой стороны, не имеется никаких эффективных гарантий против того, что не удастся утвердить вполне обоснованную исследовательскую гипотезу (ошибка второго рода).

Вообразите психолога, который изучает взаимосвязь между потребностью в достижениях и ученой степенью. Решая, какой размер будет у его выборки, он может рассуждать следующим образом: “Какую взаимосвязь я ожидаю? $r = 0.35$. Какой N мне нужен для того, чтоб мой результат был значимым? (Смотрит в таблицу.) $N = 33$. Прекрасно, это — моя выборка”.

Единственный недостаток в этом рассуждении — то, что наш психолог забыл о разнообразии среди элементов выборки, возможно, потому что он полагает, что любая выборка должна быть высоко репрезентативна по отношению к совокупности, из которой она взята. Однако, если его предположение относительно корреляции в совокупности верно, корреляция в выборке может быть около 0.35 . Следовательно, вероятность получения значимого результата (то есть, надежность теста) для $N = 33$ приблизительно равна 0.50 .

В детальном исследовании мощности статистических критериев, Дж. Коен (J. Cohen, 1962, 1969) привел правдоподобные определения больших, средних, малых результатов, обширного набора вычислительных средств для оценки мощности разнообразных статистических тестов. В нормальном тесте значение разности между двумя средними, например, разность 0.25σ считается маленькой, разность 0.50σ — средней, и разность в 1σ является большой, согласно предложенным определениям. Средняя разность между уровнем интеллекта служащих и уровнем интеллекта рабочих со средней квалификацией — это средний результат. В исследовании научной практики, Дж. Коен (1962) рассмотрел все статистические исследования, изданные в одном томе журнала «Психопатология и социальная психология», и вычислил вероятность обнаружения каждого из трех значений результата. Средняя мощность для обнаружения маленьких результатов была 0.18 , 0.48 — для средних результатов, и 0.83 — для больших результатов. Если психологи как обычно ожидают появление средних результатов и

выбирают размер выборки по такому же принципу, как и в вышеупомянутом примере, надежность их исследований должна действительно быть приблизительно 0.50.

Анализ Дж. Коена показывает, что статистическая мощьность многих психологических исследований очень мала. Это – пагубная практика: она разочаровывает ученых и понижает эффективность исследования. Исследователю, который проверяет действенную гипотезу, но не может получить значимых результатов, ничего не остается, кроме как рассматривать природу ненадежной или даже враждебной. Кроме того, как показал Оверолл (Overall, 1969), распространенность исследований, несовершенных в статистической мощьности не только иррациональна, но и фактически вредна: результатом ее является большое количество необоснованных отказов от использования нулевой гипотезы среди опубликованных результатов.

Поскольку изучение статистической мощьности имеет особую важность в смысле организации повторных исследований, мы исследовали отношение к повторяемости в нашем анкетном опросе.

Предположим, что один из Ваших докторантов закончил трудный и продолжительный по времени эксперимент, включающий в качестве испытуемых 40 животных. Он зафиксировал и проанализировал большое количество переменных. Пока результаты не окончательные, но одно предварительное сравнение оказалось высоко значимым $t = 2.70$, что было очень неожиданным и могло представлять большое теоретическое значение.

Учитывая важность результата и его неожиданное значение, и количество анализов, которые проделал докторант, порекомендовали бы Вы ему воспроизвести исследование перед публикацией работы? Если Вы рекомендуете ему повторить исследование, то с каким количеством животных Вы посоветуете ему поработать?

Среди опрошенных психологов, преобладало мнение относительно повторного исследования: его порекомендовало 66 из 75 респондентов, вероятно, потому что они подозревали, что единственно значимый результат был случайным. Среднестатистическая рекомендация докторанту была обработать 20 испытуемых в повторном исследовании. Поучительно рассмотреть вероятные последствия этого совета. Если среднее значение и дисперсия элементов во второй выборке фактически идентичны тем же показателям первой выборки, то результирующее значение t будет 1.88. Следуя рассуждениям в Примечании 1 шанс докторанта получить значимый результат при повторении – только немного выше половины (для $p = 0.05$, одностороннего теста). Так как мы ожидали, что повторная выборка из 20 испытуемых, будет казаться обоснованной для наших респондентов, мы добавляли следующий вопрос:

Предположите, что Ваш несчастный докторант фактически повторил начальное исследование, увеличив размер выборки животных на 20, и не получил значительного результата в этом направлении, $t = 1.24$. Что бы Вы порекомендовали ему теперь? Числа в круглых скобках указывают число респондентов, которые отметили этот ответ

(А) Он должен объединить результаты и издавать свои выводы как факт. (0)

- (В) Он должен сообщить о результатах как о предварительном открытии. (26)
 (С) Он должен поработать с другой группой животных (состоящей в среднем из 20). (21)
 (D) Он должен пробовать найти объяснение различиям между этими двумя группами. (30)

Обратите внимание, что независимо от уверенности в первоначальном открытии, его правдоподобность, конечно, возросла после повторного проведения опыта. Не только был получен похожий экспериментальный результат в двух выборках, но также и значимость результата в повторном опыте составила полных две трети величины результата в первоначальном исследовании. Ввиду размера выборки (20), который порекомендовали наши респонденты, повторное проведение эксперимента было приблизительно настолько успешным, насколько можно ожидать. Распределение ответов, однако, отражает существующий и поныне скептицизм относительно открытия докторанта, за которым последовал повторный эксперимент, как и рекомендовалось. Это плачевное состояние дел – типично последствие недостаточной статистической надежности.

В отличие от ответов В и С, которые могут иметь некоторые основания, наиболее популярный ответ, ответ D, является неприемлемым. Мы сомневаемся, что этот же ответ был бы получен, если бы респонденты понимали, что разница между двумя исследованиями даже приблизительно не является значимой. (Если дисперсия элементов этих двух выборок равны, то $t = 0.53$.) В отсутствии статистической проверки, наши респонденты следовали гипотезе репрезентативности: поскольку различие между этими двумя выборками была больше, чем, они ожидали, они рассмотрели это как основание для объяснений. Однако попытка «найти объяснение различию между этими двумя группами», является, по всей вероятности, попыткой объяснить появление помехи.

Наши респонденты, все вместе, оценили проведение повторного исследования довольно грубо. Это следует из гипотезы репрезентативности: если мы ожидаем, что все выборки очень похожи друг на друга, то почти все повторения правомерной гипотезы должны быть статистически значимы. Грубость критерия для успешного повторения опыта может быть продемонстрирована в ответе на следующий вопрос:

Исследователь сообщил результат, который Вы считаете неправдоподобным. Он провел исследование с 15 испытуемыми, и получил значимую величину, $t = 2.46$. Другой исследователь попытался повторить его опыт, но он получил не значимую величину t для того же числа субъектов. Инструкция была та же самая в обоих экспериментах. Вы просматриваете литературу. Какова наибольшая величина t во втором эксперименте, про которую Вы могли бы сказать, что ее невозможно воспроизвести?

Большинство наших респондентов оценило $t = 1.70$ как неудачу в повторении исследования. Если суммировать данные двух таких исследований ($t = 2.46$ и $t = 1.70$), значение t для объединенных данных приблизительно

3.00 (при равных дисперсиях). Таким образом, перед нами парадоксальное положение дел, при котором одни и те же данные, которые увеличили бы нашу уверенность по поводу результата, при рассмотрении как часть первоначального исследования, пошатнули бы нашу уверенность, если бы рассматривались как независимое исследование. Этот двойной стандарт особенно беспокоит с тех пор, как, по многим причинам, повторные проведения опытов обычно рассматриваются как независимые исследования, и гипотезы часто оцениваются с помощью перечня непоследовательных данных, подтверждающих их.

Вопреки широко распространенному мнению, может быть так, что выборка повторного опыта часто больше, чем изначальная. Решение воспроизвести когда-то полученный результат часто является следствием того, что исследователь очень доволен результатом и хочет, чтобы скептически настроенное научное сообщество его приняло. Так как сообщество необоснованно требует, чтобы результаты повторного эксперимента были значимы независимо от изначального эксперимента, или, по крайней мере, чтобы они были хотя бы приблизительно значимыми, существует необходимость обрабатывать большую выборку. В качестве иллюстрации приведем пример, что если неудачливый докторант, чья диссертация обсуждалась, принимает валидность своего первоначального результата ($t = 2.70$, $N = 40$), и если он хочет, чтобы риск того, что он получит t меньше 1.70, была только 0.1, он должен будет провести работу приблизительно с 50 животными в своем повторном исследовании. С несколько более слабым первоначальным результатом ($t = 2.20$, $N = 40$), размер повторной выборки, требуемый для получения такой же мощности теста, возрастает приблизительно до 75.

То, что результаты, обсужденные к настоящему времени, не ограничены только гипотезами относительно средней величины и дисперсии показывают ответы на следующий вопрос:

Вы выполнили исследование корреляции, оценивая 20 переменных для 100 испытуемых. Двадцать семь из 190 коэффициентов корреляции являются значимыми на уровне 0.05; и 9 из них значимы за пределами 0.01. Среднее абсолютного уровня значимых корреляций – 0.31, и образец результатов теоретически обоснован. Как вы считаете, сколько из 27 значимых корреляций, по Вашим ожиданиям, снова будут значимыми, при точном повторении исследования, при $N = 40$?

При $N = 40$, требуется корреляция приблизительно 0.31 уровня для того, чтобы она была существенной на уровне 0.05. Это средняя величина значимых корреляций в первоначальном исследовании. Таким образом, только около половины первоначально значимых корреляций (то есть, 13 или 14) остались бы значимыми при $N = 40$. Кроме того, конечно, корреляции в повторном исследовании должны отличаться от корреляций в оригинальном исследовании. Следовательно, благодаря регрессии, первоначально значимые коэффициенты, наиболее вероятно, уменьшатся. Таким образом, от 8 до 10 повторяемых корреляций из первоначальных 27 – это максимум из того,

что можно ожидать. Средняя оценка наших респондентов – 18. Это больше чем количество значимых корреляций повторного опыта, которые будут получены, если корреляции повторно вычислены для 40 испытуемых, выбранных наугад из первоначальных 100! Очевидно, люди ожидают больше, чем простое дублирование первоначальной статистики в выборке повторного исследования; они ожидают, повторения значимых результатов, не принимая в расчет размер выборки. Это ожидание требует нелепого продолжения гипотезы репрезентативности; даже закон малых чисел не способен произвести подобных результатов.

Надежда на то, что результаты возможно воспроизвести почти во всей полноте, объясняет обычное, хотя и довольно прискорбное явление. Исследователь, который вычисляет все корреляции между тремя индексами тревожности и тремя индексами зависимости, будет часто сообщать и довольно уверенно интерпретировать единственную полученную существенную корреляцию. Его уверенность в неточном результате является следствием его убеждения, что полученная матрица корреляций высоко репрезентативна и легко воспроизводима.

В этом разделе, мы увидели, что сторонник закона малых чисел ведет свою научную деятельность следующим образом:

1. Он подвергает риску свои исследовательские гипотезы на небольших выборках, не осознавая, что шансы в его пользу чрезвычайно низки. Он переоценивает мощность.

2. Он необоснованно уверен в ранних тенденциях (например, в данных, полученных на первых нескольких испытуемых) и в стабильности наблюдений (например, в количестве и идентичности значимых результатов). Он переоценивает значимость.

3. В оценке повторных исследований, своих собственных или чужих, он имеет необоснованно высокие ожидания относительно воспроизводимости значимых результатов. Он недооценивает величину доверительных интервалов.

4. Он редко объясняет отклонение от ожидаемых результатов выборки изменчивостью выборок, потому что он находит “объяснение” любому несоответствию. Таким образом, он имеет мало возможностей распознать изменчивость выборок в действии. Поэтому его вера в закон малых чисел, навсегда останется непоколебимой.

Наш анкетный опрос выявлял многочисленные примеры распространенности веры в закон малых чисел². Нашим типичным респондентом являлся сторонник этого закона, независимо от группы, к которой он принадлежит.

² В. Эдвардз (1968, 25) обосновал, что люди терпят неудачу в извлечении достаточного количества информации или определенности из вероятностных данных. Он назвал эту неудачу *консерватизмом*. Наши респонденты вполне подходят под описание консерваторов. Скорее, в соответствии с гипотезой репрезентативности, они имеют тенденцию извлекать из данных большее количество определенности, чем данные, фактически, содержат.

Не было фактически никаких различий между среднестатистическими ответами в аудитории на собрании, где обсуждались проблемы математической психологии и на общей сессии съезда Американской Психологической Ассоциации, хотя мы не заявляем о репрезентативности любой из этих двух выборок. Очевидно, познания в области формальной логики и теории вероятности не уменьшают ошибочную интуицию. Что же, в таком случае, можно сделать? Можно ли искоренить веру в закон малых чисел или, по крайней мере, контролировать ее?

Опыт исследований, вряд ли, сильно поможет, потому что дисперсия выборок также легко «объясняется». Корректирующие эксперименты – это эксперименты, которые не обеспечивают ни мотива, ни возможности для ложного объяснения. Таким образом, студент, изучающий статистику, может брать повторяемые выборки заданного размера из совокупности, и изучать влияние размера выборки на изменчивость выборки на личном опыте. Однако нет уверенности, что ожидания могут быть скорректированы таким способом, так как предубеждения, подобные таким, как ошибки игрока в казино, сохраняются несмотря на опыт.

Даже если нельзя отказаться от стереотипов, студенты могут научиться распознавать их существование и принимать необходимые меры предосторожности. Так как в обучении статистике немало предостережений, предупреждение относительно стереотипности статистической интуиции не может быть неуместным. Очевидная мера предосторожности это вычисление. Сторонник закона малых чисел имеет ошибочные убеждения относительно уровня достоверности, мощности и интервалов доверительности. Уровни значимости обычно вычисляются, и о них сообщается, а мощность и интервалы доверительности – нет. Возможно, они должны быть.

Явное вычисление мощности, относящееся к некоторой обоснованной гипотезе, например, малым, большим и средним результатам Дж. Коена (1962, 1969), несомненно, должно быть выполнено до того, как проведено исследование. Такие вычисления ведут к осознанию того, что нет никакого смысла проводить исследование, если, например, размер выборки не увеличить в 4 раза. Мы отказываемся от убеждения, что серьезный исследователь сознательно пойдет на риск 0.50 того, что его обоснованная гипотеза исследования так и не будет подтверждена. Кроме того, вычисление надежности существенны для интерпретации отрицательных результатов, то есть, если не удалось отказаться от нулевой гипотезы. Поскольку интуитивные оценки мощности читателями, вероятно, будут неправильными, то, что в данной книге приводятся подсчитанные значения величин, скорее всего, не будет пустой тратой времени читателей и бумаги.

В психологической литературе, изданной ранее, преобладало согласие при представлении отчетов, брать, например, выборку размером в среднем, например, $M \pm PE$, где PE – вероятная ошибка (то есть, доверительный интервал 50% относительно средней величины). Это соглашение было позже оставлено в пользу формулировки проверки гипотез. Доверительный интер-

вал, однако, обеспечивает полезный коэффициент для дисперсии выборки, и именно эту дисперсию мы имеем тенденцию недооценивать. Акцент на уровнях значимости имеет тенденцию уменьшать фундаментальное различие между размером результата и его статистической значимостью. Независимо от размера выборки, результат в одном исследовании является обоснованной оценкой результата при повторении. Напротив, оцененный уровень значимости в повторном эксперименте сильно зависит от размера выборки. Нереалистичные ожидания относительно воспроизводимости уровней значимости могут быть скорректированы, если будет прояснено различие между размером и значимостью, и если вычисленный размер наблюдаемых результатов описан как положено. С этой точки зрения, по крайней мере, принятие модели проверки гипотез не было долгожданным счастьем для психологии.

Ревностный приверженец закона малых чисел совершает множество грехов против логики статистического вывода с самыми честными намерениями. Гипотеза репрезентативности описывает когнитивный стереотип или стереотип чувственного восприятия, который действует независимо от мотивации. Таким образом, в то время как поспешный отказ от использования нулевой гипотезы является привлекательным, отклонение от взвешенной гипотезы, является ухудшающим, все же истинный сторонник закона малых чисел зависит и от того, и от другого. Его интуитивные ожидания подчиняются скорее постоянному неправильному восприятию мира, чем удобному принятию желаемого за действительное. В ответ на замечания редакторов, он может решить относиться с надлежащим недоверием к своей статистической интуиции и заменять впечатления вычислением всякий раз, когда это возможно.

3. Субъективная вероятность: оценка репрезентативности*

Даниель Канеман и Амос Тверски

Субъективная вероятность играет важную роль в нашей жизни. Решения, которые мы принимаем, выводы, к которым мы приходим, и объяснения, которые мы подыскиваем, обычно основаны на наших суждениях о вероятности неопределенных событий, таких как успех на новой работе, результаты выборов или состояние рынка. Действительно, большое количество литературы по экспериментальным исследованиям было посвящено вопросу, как люди воспринимают, обрабатывают и оценивают вероятность неопределенных событий, в контексте изучения вероятности, интуитивной статистики и принятия решения в условиях риска. И хотя никакой систематической теории в психологии неопределенности не появилось, благодаря этой литературе было сформулировано несколько обобщений. Возможно наиболее общий вывод, полученный из многочисленных исследований заключается в том, что люди не следуют принципам теории вероятности в оценке вероятности неопределенных событий. Вряд ли это заключение можно считать удивительным, потому что многие из законов случайности не являются ни интуитивно очевидными, ни удобными для применения. Менее очевидным, однако, является тот факт, что отклонения субъективной от объективной вероятности¹ представляются надежными, систематическими, и кажется, что их тяжело устранить. Очевидно, люди заменяют законы случайности эвристикой, оценки которой иногда бывают разумными, по очень часто — нет.

¹ Мы используем термин “ субъективная вероятность” для обозначения любой оценки вероятности события, которую дает испытуемый, или которая выводится из его поведения. Не предполагается, что эти оценки, должны удовлетворять каким-либо аксиомам или требованиям последовательности. Мы используем термин “ объективная вероятность “ для обозначения числовых значений, подсчитанных на основе установленных допущений, согласно законам вычисления вероятности. Разумеется, эта терминология не совпадает с каким-либо философским представлением вероятности.

* Эта глава — сокращенная версия статьи, которая появилась в *Cognitive Psychology*, 1972, 3, 430-454. Авторское право © Academic Press, Inc. 1972. Переиздано в соответствии с разрешением.

В настоящей книге мы подробно исследуем одну из таких эвристик, называемую репрезентативностью. Человек, который следует этой эвристике, оценивает вероятность неопределенных событий или выборок, степенью, в которой они: (1) подобны в своих существенных свойствах исходной совокупности; и (2) отражают существенные особенности процесса, с помощью которого они были созданы. Наш тезис заключается в том, что во многих случаях, событие А оценивается как более вероятное, чем событие В всякий раз, когда оно кажется более репрезентативным, чем В. Другими словами, упорядочивание событий по их субъективной вероятности совпадает с упорядочиванием их по репрезентативности.

Репрезентативность, как подобие при восприятии, легче оценить, чем охарактеризовать. В обоих случаях, отсутствует общее определение, хотя существуют случаи, в которых люди выражают согласие в том, какой из двух стимулов является более похожим на стандарт, или какое из двух событий более репрезентативно по отношению к данному процессу. В этой книге мы не измеряем репрезентативность с помощью шкалы, хотя это вполне выполнимо. Вместо этого, мы рассматриваем случаи, когда упорядочивание событий по их репрезентативности является очевидным, и показываем, что люди постоянно оценивают более репрезентативное событие как более вероятное, независимо от того, так это или нет. Хотя репрезентативность может играть важную роль во многих суждениях вероятности, например, в прогнозировании политической ситуации или клиническом диагнозе, существующий подход ограничен только повторяющимися ситуациями, где объективные вероятности легко можно вычислить.

Большинство данных, которые упоминаются в этой части книги, собраны в форме анкетного опроса приблизительно 1. 500 респондентов в Израиле. Испытуемыми были учащиеся 10, 11, и 12 классов средне-образовательных школ, готовящихся к поступлению в высшее учебное заведение (в возрасте 15-18 лет). Чтобы привлечь внимание и мотивировать испытуемых, были приложены специальные усилия. Анкетные опросы были оформлены в виде коротких тестов, проводились в естественной ситуации — в классной комнате, а имена респондентов были написаны на листах с ответами. Каждый респондент отвечал на небольшое число вопросов (обычно 2-4), на каждый из которых требовалось максимум 2 мин. Было объявлено, что вопросы изучают интуицию людей относительно определения шансов. Испытуемым давали стандартные устные инструкции, которые подробно объясняли соответствующий вопрос. Процедура проведения опыта была построена таким образом, чтобы не вызывать ассоциаций с какой-либо школой или возрастом. Большинство вопросов было предварительно опробовано на студентах старших курсов университета (в возрасте 20-25 лет) и результаты, полученные от этих двух совокупностей, были идентичными.

Детерминанты репрезентативности

В этом разделе мы обсудим характеристики выборок или событий, которые делают их репрезентативными, и демонстрируют их влияние на субъективную вероятность. Сначала, мы опишем некоторые особенности, которые определяют подобие выборки ее исходной совокупности, затем мы приступим к анализу детерминантов очевидной случайности.

Подобие выборки и совокупности

Понятие репрезентативности лучше объяснять на особых примерах. Рассмотрим следующий пример.

Были обследованы все семьи в городе, в которых было шестеро детей. В 72 семьях мальчики и девочки рождались в таком порядке Д М Д М М Д. Как вы думаете, в скольких рассмотренных семьях порядок рождения детей был М Д М М М М?

Две последовательности рождений приблизительно одинаково вероятны, но большинство людей, конечно, согласятся, что они не одинаково репрезентативны. Последовательность с пятью мальчиками и одной девочкой не сможет отразить пропорцию мальчиков и девочек в совокупности. Действительно, 75 из 92 испытуемых оценили, что эта последовательность менее вероятна, чем стандартная последовательность ($p < 0.01$ согласно тесту знаков). Средняя оценка была 30. Подобные результаты были сообщены Коэн и Хенсел (Cohen и Hansel, 1956), а также Алберони (Alberoni, 1962).

Можно задаваться вопросом, не игнорируют ли просто испытуемые информацию о порядке появления детей, и ответить на вопрос, оценив частоту появления семей с пятью мальчиками и одной девочкой относительно частоты семей с тремя мальчиками и тремя девочками. Однако, когда мы попросили, тех же самых испытуемых оценить частоту последовательности М М М Д Д Д, они сочли ее значительно менее вероятной, чем Д М М Д М Д ($p < 0.01$), предположительно потому, что первая упомянутая последовательность кажется менее случайной. Таким образом, порядок появления детей не игнорируется.

Описанная детерминанта репрезентативности заключается в сохранении в выборке соотношения меньшинства или большинства такого же как и в совокупности. Мы ожидаем, что выборка, которая сохраняет это соотношение, будет оценена как более вероятная, чем выборка, появление которой (объективно) столь же вероятно, но где это отношение нарушено. Этот эффект проиллюстрирован в следующей задаче:

В средней школе имеются две программы. Мальчиков большинство (65%) в программе А и меньшинство (45%) в программе В. В каждой из этих двух программ равное число классов.

Вы входите в класс наугад, и видите, что 55% учеников – мальчики. Как вы предполагаете, этот класс относится к программе А или программе В?

Так как большинство учеников в классе – мальчики, этот класс более репрезентативен по отношению к программе А, чем программе В. Соответственно, 67 из 89 испытуемых предположили, что класс принадлежит программе А ($p < 0.01$, по тесту знаков). На самом деле, чуть более вероятно, что класс принадлежит программе В (так как дисперсия выборки для $p = 0.45$ превышает дисперсию для $p = 0.65$).

Выборка, в которой представлены разнообразные возможные результаты, вообще, более репрезентативна, чем сравниваемая с ней выборка, в которой некоторые из возможных результатов отсутствуют. Например, данный биномиальный процесс с $p = 4/5$ значимое большинство испытуемых оценивают выборку с 10 успехами и 0 неудачами как менее вероятную, чем выборку с 6 успехами и 4 неудачами, хотя первая упомянутая выборка на самом деле более вероятна.

Влияние стереотипа репрезентативности не сводится только к простым неподготовленным испытуемым. Этот стереотип также встречается (Tversky и Kahneman, 1971, 2) в интуитивных суждениях опытных психологов. Статистическая значимость обычно рассматривается как подтверждение научной истины. Следовательно, реальный результат (в совокупности) как ожидается, будет представлен значимым результатом (в выборке), практически не принимая во внимание размер выборки. Вследствие этого, исследователи склонны переоценивать вероятность значимого результата всякий раз, когда они полагают, что нулевая гипотеза будет ложной.

Например, участникам заседания Группы Математической Психологии и Американской Психологической Ассоциации был предложен следующий вопрос:

Предположим, что Вы проводите эксперимент с 20 испытуемыми, и получили значимый результат, который подтверждает Вашу теорию ($z = 2.23$, $p < 0.05$, двусторонний критерий). Теперь у Вас есть основание провести эксперимент с дополнительной группой, состоящей из 10 испытуемых. Как Вы думаете, какова вероятность того, что результаты будут значимыми в тесте с односторонним критерием, отдельно для этой группы?

Реальная оценка желательной вероятности – несколько ниже 0.50. Средняя оценка респондентов была 0.85. Эта необоснованная уверенность в воспроизводимости значимости имеет серьезные последствия для проведения исследования: она приводит к нереалистичным ожиданиям относительно значимости и результатам планирования исследований, несовершенных в статистической мощности (см. J. Cohen, 1962).

Отражение случайности

Для того чтобы неопределенное событие было репрезентативным, не достаточно, чтобы оно было подобно своей исходной совокупности. Событие также должно отражать свойства неопределенного процесса, породившего его, то есть оно должно казаться случайным. Как и в случае с подобием выборки

и совокупности, специфические особенности, которые определяют очевидную случайность, различаются в зависимости от контекста. Несмотря на это, две общих особенности, беспорядочность и локальная репрезентативность, кажется, захватывают интуитивное представление о случайности. Обсудим эти свойства по порядку.

Главная характеристика очевидной случайности – отсутствие систематических образцов. Последовательность выпадений монеты², например, которая очевидно упорядочена – не репрезентативна. Таким образом, чередующиеся последовательности орлов и решек, как, например, ОРОРОРОР или РРООРРОО, не смогут отразить случайность процесса. Действительно, испытуемые считают такие последовательности относительно невероятными и избегают их в моделировании случайных последовательностей (Tune, 1964; Wagenaar, 1970).

Ожидается некоторая неупорядоченность, не только в появлении результатов, но также и в их распределении, как показано в следующей задаче:

В каждом туре игры, 20 шариков наугад распределяются среди пятерых детей:

Аланом, Беном, Карлом, Дэном и Эдом.

Рассмотрите следующие распределения шариков:

1. Алан — 4, Бен — 4, Карл — 5, Дэн — 4, Эд — 3.

2. Алан — 4, Бен — 4, Карл — 4, Дэн — 4, Эд — 4.

Если устроить несколько туров игры, какого типа результаты встретятся больше – первого или второго типа?

Однородное распределение шариков (II), объективно более вероятно, чем неоднородное распределение (I), но все же кажется слишком упорядоченным, чтобы быть результатом случайного процесса. Распределение первого типа, которое слегка отклоняется от равномерного распределения, является более репрезентативным по отношению к случайному распределению. Подавляющее большинство испытуемых (36 из 52, $p < 0.01$ по тесту знаков) сочли распределение первого типа более вероятным, чем распределение II типа. Присутствие некоторого беспорядка влияет на репрезентативность и, следовательно, на очевидную вероятность неопределенных событий.

Испытуемые дают ответы на вышеупомянутую задачу, как будто они игнорируют индивидуальный характер этих двух распределений и сравнивают, вместо этого, два соответствующих класса распределений, игнорируя каждое отдельное распределение шариков детям. Это не означает, что испытуемые не оценивают различие между классом и его представителями. Что они все-таки не оценивают – это надлежащее воздействие этого различия на оценку относительной частоты.

Люди рассматривают шанс как непредсказуемый, но по сути справедливый. Таким образом, они ожидают, что в абсолютно случайном распределении шариков каждый ребенок получит приблизительно (хотя и не точно)

² В этой книге мы рассматриваем только сбалансированные монеты.

одинаковое число шариков. Точно так же они ожидают, что даже короткие последовательности подбрасываний монеты, будут содержать относительно одинаковое количество «орлов» и «решек». Вообще, репрезентативная выборка – эта та, в которой существенные характеристики исходной совокупности представлены в целом не только в полной выборке, но также и локально в каждой из ее частей. Выборка, репрезентативная локально, однако, систематически отклоняется от ожиданий шансов: в ней слишком много изменений и слишком мало группировок.

Закон больших чисел гласит, что очень большие выборки высоко репрезентативны по отношению к совокупности, из которой они взяты. В другом случае (Tversky и Kahneman, 1971, 2), мы охарактеризовали ожидание локальной репрезентативности как приверженность закону малых чисел, согласно которому “закон больших чисел применим также и к малым числам”. Эта убеждение, как мы предполагаем, лежит в основе ошибок интуиции относительно случайности, которая представлена в широком разнообразии контекстов.

Исследования восприятия случайности (например, Tune, 1964; Wagenaar, 1970) показывают, что, когда людей просят смоделировать случайный процесс, такой как серии подбрасываний монеты, они создают последовательности, которые являются репрезентативными локально, с большим количеством коротких отрезков. Более того, люди имеют тенденцию рассматривать как маловероятные, или отклонять как неслучайные последовательности, которые имеют правильное распределение на длине отрезка, возможно, потому что длинные отрезки не репрезентативны локально.

Подобные результаты также были получены в сотнях исследований, изучающих вероятность и бинарное прогнозирование (Estes, 1964; M. R. Jones, 1971). Ошибка игрока в казино или эффекта отрицательной новизны является проявлением убежденности в локальной репрезентативности. Поскольку, если соотношение этих двух результатов сохраняются на коротких отрезках, то за длинной последовательностью одного результата для восстановления равновесия должен идти другой результат. В мире, где действует локальная репрезентативность, фактически, ошибка игрока в казино больше не является ошибкой.

В своем *Введении в Теорию Вероятности*, Феллер (1968, р. 160) описывает пример, который иллюстрирует ошибочную веру в локальную репрезентативность. Во время интенсивной бомбежки Лондона во Вторую Мировую войну, считалось вообще, что выбор целей бомбежки не может быть случайным, потому что некоторые районы города были поражены несколько раз, в то время как на многие другие бомбы не падали совсем. Таким образом, рисунок попаданий бомб нарушил закон локальной репрезентативности, и гипотеза случайности попаданий казалась недопустимой. Чтобы проверить эту гипотезу, всю территорию Юга Лондона разделили на маленькие области равной площади, и фактическое распределение попаданий в этом районе сравнили с ожидаемым (пуассоновским) распределением согласно

предположению о том, что бомбежка велась по случайному принципу. Вопреки общим ожиданиям, соответствие между распределениями было очень сильным. “Для нетренированного глаза, – замечает Феллер, – случайность кажется упорядоченностью или тенденцией к группировке”.

Большинство студентов удивляются, когда узнают, что в группе, состоящей только из 23 человек, вероятность того, что, по крайней мере, у двух из них день рождения в один и тот же самый месяц и день, превышает 0.5. Ясно, что при выборке в 23 человека, ожидаемое количество дней рождения в один день - меньше чем $1/15$. Таким образом, день с двумя днями рождения, а также наличие 343 “пустых” дней, – очень нерепрезентативны, и рассматриваемый случай, поэтому, кажется маловероятным. Обобщая, мы считаем, что характер многих результатов в теории вероятности, противоречащий интуиции, относится к нарушениям репрезентативности. (Для убедительного примера из теории случайных шагов, см. Feller, 1968, стр. 84-88.)

Репрезентативная выборка, в таком случае, является подобной совокупности в своих существенных характеристиках, и отражает случайность такой, как ее видят люди; то есть все ее части репрезентативны, и ни одна из них не является слишком упорядоченной. Только немногие из всех возможных выборок удовлетворяют всем этим ограничениям. Большинство выборок — нет, и поэтому они не кажутся случайными. Среди 20 возможных последовательностей шести подбрасываний монеты, например, мы рискуем столкнуться с тем, что только выборка ОРРОРО окажется действительно случайной. Подбросив монету 4 раза, мы можем не получить подобной последовательности.

Тенденция расценивать некоторые двоичные последовательности как более случайные, чем другие, имела драматические последствия для эксперимента на радио Зенит³, в котором аудитории было предложено предположить значения пяти двоичных символов, которые были “телепатированы” им группой радиовещания. Анализ более чем миллиона ответов (Goodfellow, 1938) показал, что число выборов для некоторых последовательностей значительно превышало их вероятность, а для других было существенно занижено, причем оно в значительной степени зависело от очевидной случайности входящих («телепатированных») последовательностей. Значение этих результатов для ESP исследований является очевидным.

Последовательности, кажутся случайными, когда словесное описание наиболее длинное. Вообразите, что Вы диктуете длинную последовательность двоичных символов, например, орлов и решек. Вы будете, несомненно, использовать сокращенные выражения типа «четыре решки» или «Орел-Решка три раза подряд». Последовательность с большим количеством повторяющихся подряд символов, позволяет использовать сокращения пер-

³ Мы благодарим R. P. Abelsona за предоставление нам своего исследования.

вого типа. Структура таких отрезков, похожих на случайные, минимизирует полезность таких сокращений, и, следовательно, не поддается экономичным описаниям. Очевидная случайность, поэтому, является разновидностью сложности структуры. Детерминанты сложности структуры, такие как кодируемость (Garner 1970; Glanzer & Dark, 1963; Vitz и Todd, 1969) также принимают вид очевидной случайности.

Распределение выборок

Мы предположили, что испытуемые оценивают вероятность события таким образом, чтобы более репрезентативным событиям приписывать более высокие вероятности, а одинаково репрезентативным событиям – равные вероятности. В этом параграфе мы исследуем значение этой гипотезы для изучения субъективных распределений выборок, то есть, вероятность того, что испытуемые возьмут выборки данного размера из указанной совокупности.

Когда выборка описана в терминах единичной статистики, например, пропорции или среднего, то степень, в которой она репрезентативна совокупности, определяется подобием этой статистики соответствующему параметру совокупности. Так как размер выборки не отражает никакой специфической особенности исходной совокупности, он не ассоциируется с репрезентативностью. Таким образом, событие, в котором обнаруживается больше чем 600 мальчиков в выборке из 1.000 младенцев, например, столь же репрезентативно, как обнаружение больше чем 60 мальчиков в выборке из 100 младенцев. Поэтому, два эти события, были бы оценены как равновероятные, хотя последнее, на самом деле, значительно более вероятно. Точно так же согласно данному анализу, субъективная вероятность того, что средний рост в выборке людей — между 6 футами 0 дюймами и 6 футов 2 дюймами будет независима от размера выборки.

Чтобы проверять эти прогнозы, девять различных групп испытуемых воспроизвели субъективные распределения выборок для трех размеров выборки ($N = 10, 100, 1.000$) и для каждой из следующих трех совокупностей.

Распределение полов. (Биномиальное, $p = 0.50$) Испытуемым сказали, что каждый день рождается приблизительно N младенцев в определенном регионе. Для $N = 1.000$, например, вопрос, звучал так:

Каков процент дней, в которые число мальчиков среди 1.000 младенцев будет следующим:

- До 50 мальчиков*
- от 50 до 150 мальчиков*
- от 150 до 250 мальчиков*
-*
- От 850 до 950 мальчиков*
- Больше чем 950 мальчиков*

Обратите внимание, что эти категории учитывают все возможности, так что сумма процентов в Ваших ответах должна составить в целом приблизительно 100%.

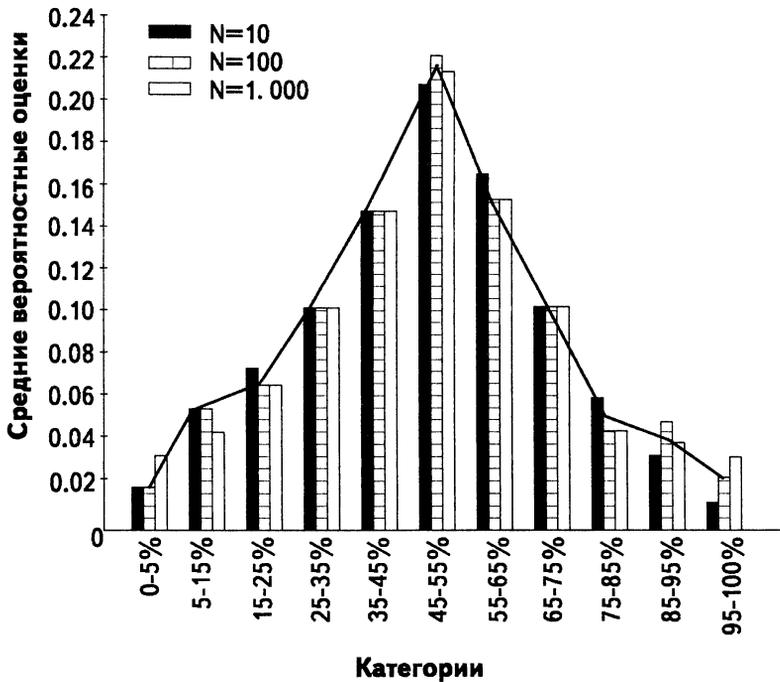


Рис. 1. Распределение полов

Для $N = 100$, эти 11 категорий были: до 5, 5-15, и т.д.

Для $N = 10$, каждая категория содержала единичный результат, например, 6 мальчиков.

Распределение типа сердцебиения. (Биномиальное, $p = 0.80$) В этом случае, испытуемым сказали, что каждый день в некоторой области рождаются приблизительно N младенцев, и что 80% всех новорожденных имеет сердцебиение типа α , а остальные 20% имеют сердцебиение типа β .

Для каждого размера выборки, испытуемые произвели распределения выборок для количества младенцев, рожденных каждый день, с сердцебиением типа α , используя 11 категорий как те, что приведены выше.

Распределение роста. Испытуемым сказали, что региональный военкомат делает учет среднего роста N количества людей, которых обследуют каждый день. Им также сообщили, что средний рост мужчин в совокупности — между 170-175 см (в Израиле рост измеряется в сантиметрах), и что частота появления различного уровня роста уменьшается по мере удаления от среднего. Для каждого размера выборки испытуемые воспроизвели распределение выборки людей среднего роста, в следующих семи категориях: до 160, 160-165, ..., больше 185.

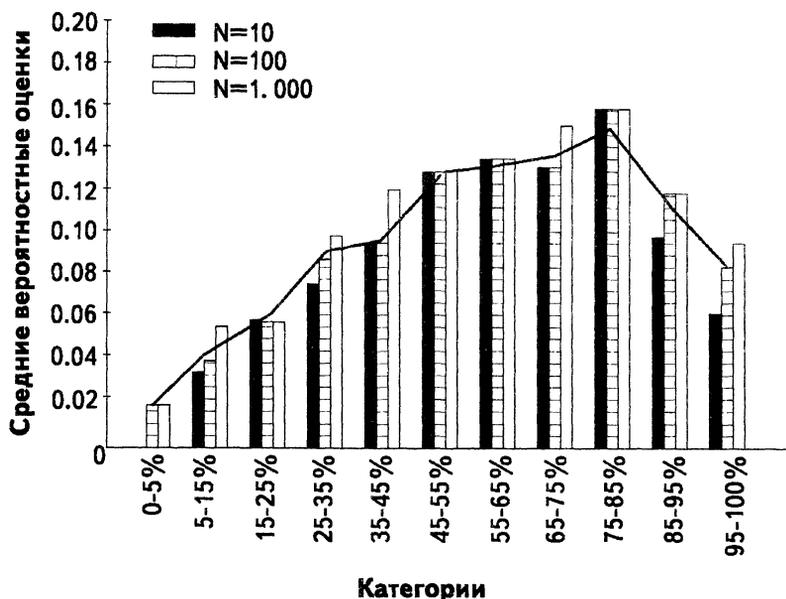


Рис. 2. Распределение типов сердцебиения

Средняя оценка для этих трех совокупностей, соответственно, показана на рисунках 1, 2 и 3 для всех трех значений N . (Размер группы варьировался от 45 до 84, со средним – 62.) Очевидно, что размер выборки вообще не оказывает никакого влияния, независимо от субъективных распределений выборок. Независимые группы тестируемых, которые решали задачи, отличающиеся друг от друга только размером выборки, производят совершенно одинаковые распределения. Этот результат остается тем же и для совокупностей, определенных абстрактно, например, биномиальные, также как для совокупностей, которые известны испытуемым из их ежедневного опыта, например, рост людей.

Так как субъективные распределения выборок независимы от N , сплошные линии на каждом графике, которые соединяют значения средних оценок, могут рассматриваться как “универсальные” распределения выборок для соответствующей совокупности. Для изображения того, насколько в действительности размер выборки влияет на распределение, который испытуемые полностью игнорируют, показаны правильные распределения выборок для $p = 0.50$ и $p = 0.80$, вместе с соответствующим “универсальным” распределением выборок, на Рисунках 4 и 5, соответственно.

Можно заметить, что “универсальные” кривые даже ровнее, чем правильные кривые для $N = 10$. Для $p = 0.50$, “универсальная” дисперсия (0.048) грубо равна нормальной дисперсии для выборки для $N = 5$ (0.05). Для $p = 0.80$, дисперсия “универсальной” кривой (0.068) находится между нормальной дисперсией выборки для $N = 2$ и $N = 3$.

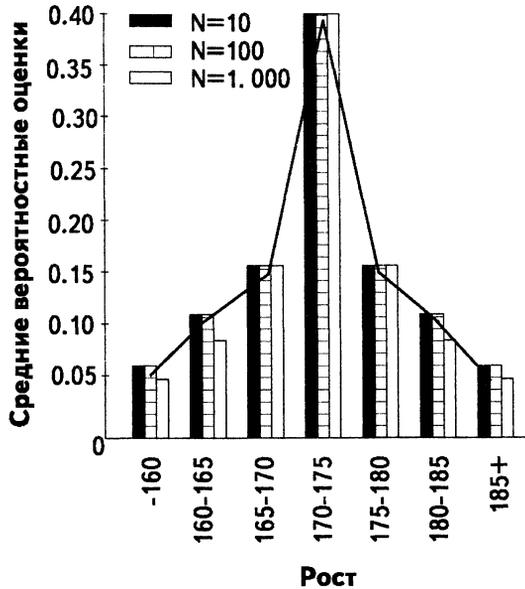


Рис. 3. Распределение роста

В биномиальных распределениях, средняя величина вообще совпадает с модой. Следовательно, когда $p \neq 0.50$, короткий «хвост» должен быть больше, чем длинный; см., например, правильное распределение для $N = 10$ на Рисунке 4. Рисунок 4 также показывает, что это свойство нарушено «универсальной» кривой для $p = 0.80$, чье среднее равно только 0.63. Таким образом, хотя мода в субъективном распределении выборок должным образом расположена в области наиболее репрезентативной величины, средняя величина смещена к длинному «хвосту». Тот же самый результат был получен и в других исследованиях, например, Cohen и Hansel (1956), Peterson, DuCharme и Edwards (1968). Таким образом, для $p = 0.80$ «универсальное» выборочное распределение пропорции – вообще не биномиальное!

Следующий эксперимент отличается от предыдущих субъективных биномиальных распределений (Peterson, DuCharme и Edwards, 1968; Wheeler & Beach, 1968). Сначала в более ранней работе описывались размеры выборок, намного меньшие, чем размеры выборок данного исследования. Во-вторых, что является более важным, число событий, для которых были распределены вероятности, не было одинаковым для различных размеров выборок: для выборки размера N , испытуемые оценивали $N + 1$ результатов. В данном исследовании, напротив, испытуемые оценивают то же самое число категорий для всех размеров выборки. Неизменность субъективного распределения выборки относительно N , которая показана на Рисунках 1, 2 и 3, не может оставаться той же самой, когда изменяется число категорий, или когда выборка достаточно маленькая, чтобы шансы можно было задать перечислением. Для больших выборок, задание всех шансов перечислением невозможно,

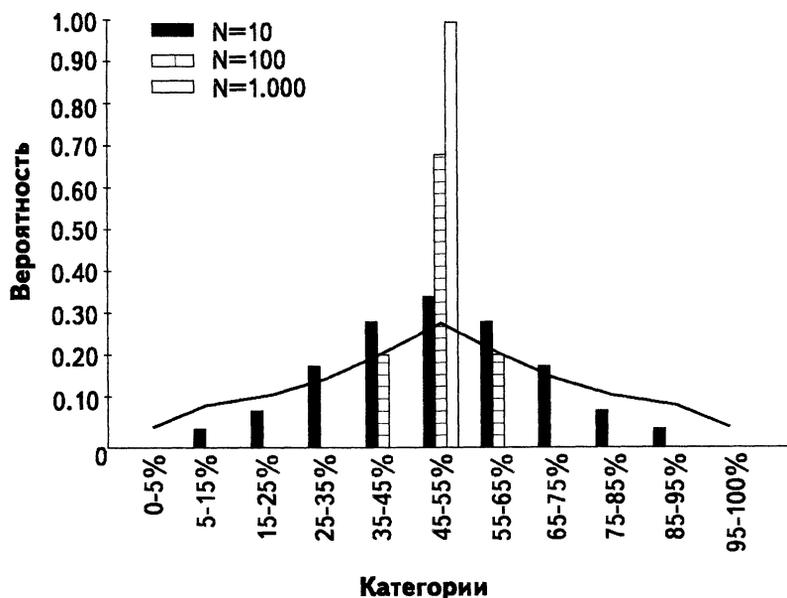


Рис. 4. Распределение выборок $p=0.50$

и вполне естественно обращение к прямой оценке репрезентативности, на которую влияет среднее выборки, или пропорция выборки.

Чтобы далее исследовать прогнозирование репрезентативности относительно размера выборки, был проведен дополнительный эксперимент. В качестве испытуемых были 97 студентов старших курсов Стэнфорда без какой-либо подготовки в теории вероятности или статистике, собранные в маленькие группы от 5 до 12 человек в каждой. Им были предоставлены, в установленном порядке, три задачи, каждая из которых определяла собой процесс формирования выборки с указанным средним и критической величиной, которая была выше этого среднего, и их попросили оценить, является ли отдельно взятый результат выборки более вероятным в большой или в маленькой выборке. Каждому испытуемому заплатили \$1 за участие в эксперименте и еще один дополнительный \$1, если его ответ на одну из задач (отобранный наугад после завершения задания) был правильным.

Чтобы контролировать предубеждения ответа, каждая задача была представлена в двух формах. Половина испытуемых оценивала, для всех трех задач, является ли более вероятным, что исход, который более экстремален, чем указанная критическая величина, встретится в маленькой или в большой выборке. Правильный ответ, конечно, – то, что экстремальный исход, более вероятно, встретится в маленькой выборке. Остальные испытуемые оценили, что более вероятно: если результат, который является менее экстремальным, чем указанная критическая величина, встретится в маленькой или в большой выборке. Правильный ответ здесь – то, что такой результат, более вероятно, встретится в большой выборке. Данные три задачи представлены ниже. Указанные числа – количество испытуемых, которые выб-

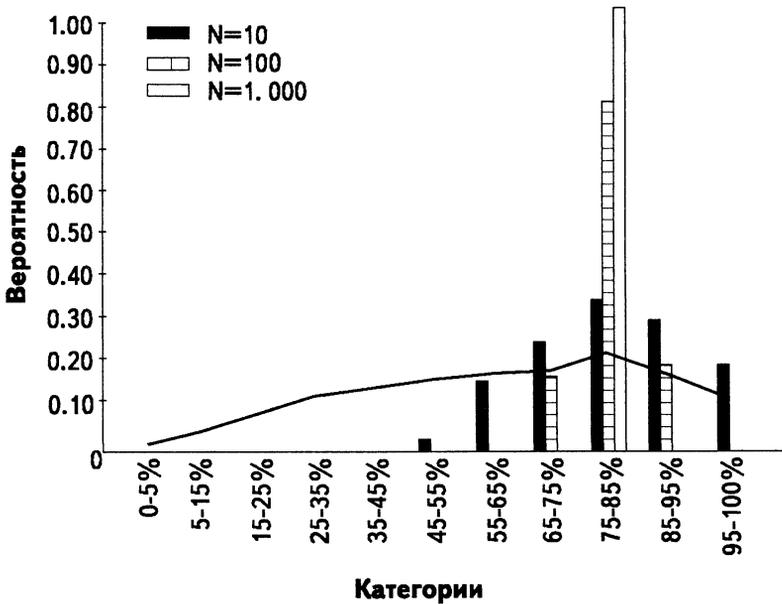


Рис. 5. Распределение выборки $p = 0.80$

рали каждую категорию ответа, для каждой из двух форм. Правильные ответы помечены звездочками.

1. Некий город обслуживается двумя больницами. В большей больнице каждый день рождается приблизительно 45 младенцев, а в меньшей больнице каждый день рождается приблизительно 15 младенцев. Как Вы знаете, приблизительно 50% всех младенцев – мальчики. Точный процент мальчиков-младенцев, однако, изменяется со дня на день. Иногда он может быть выше, чем 50%, иногда понижаться.

	Больше чем 60 %	Меньше чем 60 %
Большая больница	12	9*
Маленькая больница	10*	11
Примерно поровну (т.е. в пределах 5% разницы)	28	25

В течение 1 года каждая больница вела учет дней, в которые (больше / меньше) чем 60% рожденных младенцев были мальчики. Какая больница, как Вы думаете, сделала учет большего количества таких дней?

2. Исследователь, изучающий некоторые свойства языка, взял книгу в мягкой обложке и вычислил среднюю длину слова на каждой странице книги (то есть, количество букв на той странице, поделенное на количество слов). Другой исследователь взял пер-

вую строку на каждой странице и вычислил среднюю длину слова в строке. Средняя длина слова во всей книге – четыре. Однако, не в каждой строке или на странице точно такая же средняя длина слова. В некоторых средняя длина слова может быть больше или меньше.

	Больше 6	Меньше 6
Ученый, исследовавший страницы	8	10*
Ученый, исследовавший строки	21*	15
Примерно поровну (т.е. в пределах 5% разницы)	20	23

Первый исследователь сосчитал количество страниц, на которых средняя длина слова 6 или (больше / меньше), а второй исследователь сосчитал количество строк, в которых средняя длина слова 6 или (больше / меньше). Какой исследователь, как Вы думаете, сделал запись большего количества таких единиц (страницы для одного, строки для другого)?

3. Проводится медицинское обследование, с целью изучения некоторых факторов, влияющих на коронарные заболевания. Две группы собирают данные. Одна обследует по три человека в день, а другая – одного человека. Эти люди выбраны наугад из совокупности. Во время осмотра замеряется рост каждого человека. Средний рост взрослого мужчины 5 футов 10 дюймов. Людей, чей рост выше среднего, столько же, сколько и людей, чей рост ниже среднего.

	Больше 5 футов 11 дюймов	Меньше 5 футов 11 дюймов
Команда, проверявшая по 3 человека в день	7	14*
Команда, проверявшая по одному человеку в день	18*	17
Примерно поровну (т.е. в пределах 5% разницы)	23	17

Группа, проверяющая трех человек в день, оценивает их рост, и отмечает дни, когда средняя величина роста человека была (больше / меньше) чем 5 футов 11 дюймов, другая группа просто считает дни, когда человек, которого они проверили, был (выше\ниже) 5 футов 11 дюймов. Какая команда, как Вы думаете, насчитала большее количество таких дней?

Если бы у испытуемых было какое-то понимание роли размера выборки, они должны были бы без труда выбрать правильные ответы на эти простые вопросы. С другой стороны, если они считают, что одинаково репрезентативные результаты в равной степени вероятны, они не должны проявлять какого-либо систематического предпочтения правильному ответу. И это ясно видно из примеров. Модальный ответ – “поровну” почти во всех примерах;

кроме того, в любой из задач нет никакого существенного предпочтения правильному ответу.

Этот эксперимент подтверждает заключения первоначального исследования, несмотря на некоторые процедурные различия. Здесь, каждый субъект прямо оценивает вероятность результата, имея выборки двух размеров при условиях, разработанных, таким образом, чтобы стимулировать его давать правильные ответы. Эта процедура должна увеличить выразительность такой характеристики выборки, как размер. Кроме того, последняя задача сравнивает единичное наблюдение со средней величиной выборки из трех наблюдений. Как было показано, респонденты не смогли заметить даже тот очевидный факт, что средние величины должны быть менее изменчивыми, чем единичные наблюдения.

Понятие, что разнообразие элементов выборки уменьшается с увеличением количества элементов выборки, очевидно, не является частью интуитивных представлений человека. Действительно, неправильные представления о роли типового размера часто проявляются в ежедневной жизни. С одной стороны, люди часто принимают всерьез результат, выраженный в процентах, не заботясь о количестве наблюдений, которое может быть смехотворно маленьким. С другой стороны, люди часто остаются скептически перед лицом неоспоримого свидетельства из большой выборки. Как в случае известного политического деятеля, который горько жаловался, что стоимость проживания не основана на данных о целом населении, а только на большой выборке, и добавил, “И что еще хуже, на *случайной* выборке”.

Конечно, мы не хотим сказать, что человек не способен понять влияние размера выборки на разнообразие ее элементов. Людей можно научить верному правилу, возможно даже почти без труда. Дело в том, что люди не следуют этим правилам, когда предоставлены сами себе. Более того, изучение поведения психолога-исследователя (J. Cohen, 62, Tversky и Kahneman, 1971, 2) показывает, что сильная тенденция недооценивать влияние размера выборки не исчезает, несмотря на знание верного правила и обширного обучения в области статистики. Любого, кто хотел бы рассматривать человека как разумного интуитивного статистика, такие результаты обескураживают.

Нормативные модели и дескриптивные эвристики

Существует мнение (см., например, W. Edwards, 1968, 25), что человек, вообще говоря, следует правилу Байеса, но не в состоянии оценить полное воздействие очевидности, и поэтому – консервативен. Петерсон и Бич, например, сделали вывод, что нормативная модель обеспечивает хорошее первое приближение к поведению испытуемых, которые находятся “под влиянием определенных переменных и в определенных направлениях” (Peterson и Beach, 1967, с. 43). Это точка зрения не была разделена всеми. В более современном обзоре литературы, Словик и Лихтенштейн (Slovic и Lichtenstein, 1971) доказывали, что вышеупомянутая оценка поведения человека как

интуитивного статистика “слишком щедра,” в то время как Пиц, Даунинг и Рейнхольд пришли к выводу, на основе своих данных, что человеческое поведение в задачах Байеса “неоптимально в более фундаментальном смысле, чем подразумеваются в обсуждениях консерватизма” (Pitz, Downing и Reinhold, 1967, с. 392).

Польза, которую может принести нормативный подход Байеса анализу и моделированию субъективной вероятности зависит, прежде всего, не от точности субъективных оценок, а скорее от того, охватывает ли модель существенные детерминанты процесса суждения. Исследование, которое обсуждается в этой книге, показывает, что это не так. В частности мы видели, что размер выборки не оказывает влияния на субъективные распределения выборок, что последующие биномиальные оценки определены (по крайней мере, в общем случае) скорее пропорцией выборки, чем разнообразием элементов выборки, и что они не зависят от пропорции совокупности. В своей оценке очевидности человек – вероятно не консервативный последователь Байеса: он – не последователь Байеса вовсе.

Является спорным, что неспособность нормативной модели описать человеческое поведение свойственна только для неопытных испытуемых, которые столкнулись с незнакомыми случайными процессами, и что она могла бы обеспечивать адекватную оценку более знакомых случайных процессов, с которыми люди сталкиваются в повседневной жизни. Имеется мало подтверждений, однако, которые могли бы засвидетельствовать это мнение. Сначала, было показано (Tversky и Kahneman, 1971, 2), что те же самые систематические ошибки, предложенные соображениями репрезентативности, могут обнаружиться в интуитивных суждениях хорошо обученных ученых. Очевидно, знакомство с теорией вероятности не устраняет все ошибки интуиции относительно законов случая. Во-вторых, в нашей повседневной жизни мы сталкиваемся с многочисленными случайными процессами (например, рождение мальчика или девочки, красный свет светофора на перекрестке, набор карт без единой карты червовой масти), которые подчиняются биномиальному закону, например, в высокой степени приближения. Люди, однако, будут не в состоянии извлечь адекватную концепцию бинарного процесса из этого опыта. Очевидно, использование постоянных многочисленных проверок не приводит к оптимальному поведению.

В своей жизни люди задают себе и другим вопросы типа: Какова вероятность того, что этот 12-летний мальчик вырастет и станет ученым? Какова вероятность, что этот кандидат будет принят на службу? Какова вероятность, что эта компания уйдет из бизнеса? Эти проблемы отличаются от тех, которые обсуждались ранее в этой книге тем, что в силу их уникального характера, их нельзя решить ни в терминах частоты возникновения в прошлом, ни в терминах некоторого четкого процесса осуществления выборки.

В этой книге, мы довольно детально исследовали одну эвристику, согласно которой вероятность события определяется степенью, в которой оно репрезентативно по отношению к главным характеристикам процесса или

совокупности, из которой оно взято. Хотя наши примеры исследований были ограничены четкими процессами осуществления выборки (где объективная вероятность легко вычислима), мы считаем, что точно такая же эвристика играет важную роль в оценке неопределенности в уникальных ситуациях, где никакой “правильный” ответ не доступен. Вероятность того, что отдельно взятый 12-летний мальчик станет ученым, например, может быть определена степенью, в которой роль ученого репрезентативна по отношению к нашему представлению об этом мальчике. Точно так же, когда мы размышляем о вероятности того, что компания уйдет из бизнеса, или что политического деятеля назначат на должность, у нас в голове есть модель компании или политической ситуации, и мы оцениваем как наиболее вероятные те результаты, которые лучше всего представляют существенные особенности соответствующей модели....

4. О психологии прогнозирования*

Даниель Канеман и Амос Тверски

В этой части книги мы рассматриваем правила, которые определяют интуитивные прогнозы и оценки достоверности, и сопоставляем их с нормативными принципами статистического прогнозирования. Предлагаются два вида прогнозов: *прогнозирование категории* и *числовое прогнозирование*. При прогнозировании категории, предсказание осуществляется в номинальной форме, например, победитель на выборах, диагноз пациента или будущая профессия человека. При числовом прогнозировании, предсказание дается в числовой форме, например, котировка акций или средняя оценка успеваемости студента.

При прогнозировании и принятии решений в условиях неопределенности, людям не свойственно определять вероятность исхода или прибегать к статистической теории прогнозирования. Вместо этого, они полагаются на ограниченное число эвристик, что иногда приводит к верным суждениям, а иногда влечет за собой серьезные и систематические ошибки (Kahneman & Tversky, 1972b, 3; Tversky & Kahneman, 1971, 2; 1973, 11). В этой книге мы рассматриваем роль в интуитивных прогнозах одной из таких эвристик – репрезентативности.

При наличии определенных данных (например, краткого описания личности), соответствующие исходы (например, род занятий или уровень достижений) можно определить степенью, в которой они репрезентативны по отношению к этим данным. Мы утверждаем, что люди прогнозируют на основе репрезентативности, то есть выбирают или предсказывают последствия, анализируя степень, в которой исходы отражают существенные особенности исходных данных. Во многих ситуациях репрезентативные последствия действительно более вероятны, чем другие. Однако это не всегда так, потому что существует ряд факторов (например, априорные вероятности исходов и на-

* Эта глава впервые появилась в *Psychological Review*, 1973, 80, 237-251. Авторское право © 1973

дежность первичных данных), которые влияют на вероятность исходов, а не на их репрезентативность. Поскольку эти факторы люди не принимают во внимание, то их интуитивные предсказания систематически и существенно нарушают статистические правила прогнозирования.

Чтобы подтвердить эту гипотезу, мы покажем, что предсказание исходов на основе кажущегося подобия совпадает с их предсказанием на основе репрезентативности, и что априорная вероятность и ожидаемая точность прогноза, по существу, не влияют на интуитивные предсказания.

В первой части мы исследуем прогнозирование категории и демонстрируем, что оно подчиняется независимой оценке репрезентативности и по существу не зависит от предшествующих вероятностей исходов. В следующей части мы исследуем числовое прогнозирование и показываем, что оно недостаточно регрессивно и на него существенно не влияют соображения достоверности. Следующие три части описывают, соответственно, методологические проблемы в изучении прогнозирования, источники необоснованной уверенности по поводу прогнозов и некоторые ошибочные прогнозы относительно эффектов регресса.

Прогнозирование категории

Базовое значение, подобие и вероятность

Следующий экспериментальный пример иллюстрирует прогнозирование репрезентативности и ошибки, связанные с этим способом интуитивного прогнозирования. Группе из 69 испытуемых¹ (группа базового значения) задали следующий вопрос: “Возьмем всех аспирантов первого курса в США на сегодняшний день. Какой процент студентов, по Вашему мнению, задействован в каждой из девяти областей специализации? Области специализации представлены в Таблице 1. Первая колонка этой таблицы представляет средние оценки базового значения для различных областей.

Второй группе из 65 испытуемых (группа подобия) предложили следующее краткое описание личности:

Том В. обладает высоким интеллектом, но ему не хватает творческого потенциала. Он любит порядок и ясность и нуждается в системе, где каждая мелочь находится на своем месте. Манера письма — однообразная и механическая, иногда оживляется несколько банальными каламбурами и вспышками воображения подобно произведениям научной фантастики. Имеет сильную тягу к устойчивому материальному положению. Проявляет мало сочувствия и симпатии к другим людям, не любит взаимодействовать с другими. Будучи эгоцентричным, тем не менее, имеет глубокое чувство морали.

¹ При отсутствии дополнительных уточнений, под испытуемыми следует понимать оплачиваемых добровольцев, привлеченных через студенческую газету в Университете штата Орегон. Данные были собраны в классах для групповых занятий.

Испытуемых попросили проанализировать девять областей науки с точки зрения того, “насколько подобен Том В. типичному аспиранту каждой из областей науки?” Во второй колонке Таблицы 1 представлены средние значения подобия, соответствующие различным областям науки.

Наконец, группе прогнозирования, состоящей из 114 аспирантов-психологов трех крупных университетов Соединенных Штатов, было представлено краткое описание личности Тома В. со следующей дополнительной информацией.

Предыдущее краткое описание личности Тома В. было составлено психологом на основании проективного теста, когда Том был старшеклассником средней школы. В настоящее время Том В. – аспирант. Расположите следующие девять областей специализации аспирантов в порядке вероятности того, что Том В. является теперь аспирантом в одной из этих областей.

Таблица 1. Оценки базовых значений для девяти областей специализации и краткое описание данных подобия и прогнозирования для Тома В.

<i>Специализация аспиранта</i>	<i>Среднее базовое значение (в %)</i>	<i>Средняя величина подобия</i>	<i>Средняя величина вероятности</i>
Управление бизнесом	15	3.9	4.3
Программирование/ информатика	7	2.1	2.5
Технические науки	9	2.9	2.6
Гуманитарные науки и педагогика	20	7.2	7.6
Юриспруденция	9	5.9	5.2
Библиотечарское дело	3	4.2	4.7
Медицина	8	5.9	5.8
Физика и естественные науки	12	4.5	4.3
Социальные науки	17	8.2	8.0

Третья колонка Таблицы 1 отражает средние величины исходов, полученных испытуемыми в группе прогнозирования.

Позже было рассчитано произведение моментов между колонками Таблицы 1 (корреляция Пирсона). Корреляция между вероятностью и подобием составила – 0.97, в то время как корреляция между вероятностью и ба-

зовым значением² – 0.65. Очевидно, что суждения вероятности по существу совпадают с суждениями подобия и существенно отличаются от суждений базового значения. Полученный результат является прямым подтверждением гипотезы, что люди делают прогнозы, опираясь на репрезентативность или подобие.

Вероятностные суждения аспирантов-психологов грубо нарушают нормативные правила прогнозирования. Более 95% респондентов предположили, что Том В., вероятнее всего, изучает информатику, чем гуманитарные науки или педагогику, хотя они твердо знали, что в последней области аспирантов намного больше. Согласно оценкам базового значения, отраженным в Таблице 1, исходная разница между гуманитарными науками и информатикой — приблизительно 1 к 3 (реальная разница намного больше).

Согласно правилу Байеса, можно преодолеть первоначальное противоречие, позволяющее Тому В. заниматься информатикой, а не гуманитарными дисциплинами и педагогикой, если описание его личности является одновременно и точным, и диагностическим. Однако, аспиранты, принявшие участие в нашем исследовании, не считали, что эти условия были выполнены. После выполнения задачи прогнозирования, респондентов попросили подсчитать процент правильных ответов (то есть, исправить свой первоначальный выбор среди девяти областей специализации), который можно было бы достичь, при наличии у них нескольких типов информации. Средние значения правильных ответов составили 23% для прогнозов, основанных на проективных тестах, для сравнения с 53%, для прогнозов, основанных на сообщениях старшеклассников средней школы касательно их интересов и планов. Очевидно, проведение проективных испытаний не пользовалось большим доверием. Однако аспиранты полагались на описание, полученное из таких испытаний, и не принимали во внимание базовые значения.

В общем, для статистического прогнозирования важны три типа информации: (а) первичная или фоновая информация (например, базовые значения областей специализации выпускников ВУЗов); (б) дополнительная информация для отдельного взятого случая (например, описание Тома В.); (в) ожидаемая точность прогноза (например, априорная вероятность правильных ответов). Фундаментальное правило статистического прогнозирования гласит, что ожидаемая точность влияет на удельный вес, приписываемый дополнительной и первичной информации. При уменьшении ожидаемой точности, предсказания должны становиться более регрессивными, то есть близкими к прогнозам, основанным на первичной информации. В случае Тома В., ожидаемая точность была низкой, и испытуемым следовало опираться на априорную вероятность. Вместо этого, они делали прогнозы, на

² При вычислении этой корреляции, оценки были инвертированы таким образом, чтобы высоко оцененной вероятности соответствовало высокое значение.

основе репрезентативности, то есть они прогнозировали исходы по их подобию дополнительной информации, не принимая во внимание априорную вероятность.

Испытуемые в группе прогнозирования целиком полагались на краткое описание личности, и, очевидно, не принимали во внимание следующие соображения. Во-первых, учитывая известную неадекватность проективного тестирования личности, вполне вероятно, что Том В. действительно не страдал навязчивыми стремлениями и не был таким равнодушным, как предполагает его описание. Во-вторых, даже если описание соответствовало действительности, когда Том В. был учащимся средней школы, оно могло стать ложным при поступлении в аспирантуру. Наконец, даже если описание все-таки достоверно, очевидно, что под него подходит больше студентов, изучающих гуманитарные науки и педагогику, чем студентов, изучающих информатику, просто потому что в первой области студентов намного больше, чем в последней.

Манипуляция ожидаемой точностью

Гипотезой проведенного нами дополнительного исследования являлось то, что вопреки статистической модели, манипуляция ожидаемой точностью не затрагивает структуру прогнозов. Экспериментальный материал состоял из пяти кратких описаний личности мальчиков-девятиклассников, предположительно составленных консультантом-психологом на основе интервью как части лонгитюдного исследования. Описания были построены также, как и в исследовании с Томом В. Для каждого описания, испытуемые первой группы ($N = 69$) рассмотрели девять областей специализации аспирантов (см. Таблицу 1) относительно соответствия отдельно взятого мальчика “образу типичного аспиранта-первокурсника в данной области”. Опираясь на суждения подобия, они оценили частоту базового значения девяти областей специализации аспирантов. Результаты представлены в Таблице 1. Оставшимся испытуемым сообщили, что эти описания были наугад отобраны из числа участников первоначального исследования, которые теперь являются аспирантами-первокурсниками. Группе с высокой точностью ($N = 55$), указали, что “на основе данных описаний, испытуемые студенты делают правильные прогнозы приблизительно в 55% случаев”. Группе с низкой точностью ($N = 50$) сообщили, что предсказания студентов правильны приблизительно в 27% случаев. Для каждого описания, испытуемые оценили эти девять областей согласно “вероятности того, что данный человек является на данный момент аспирантом в этой области”. Они также определили вероятность правильности первого выбора для каждого описания.

Манипуляция ожидаемой точностью имела существенное влияние на суждения о вероятности. Средние оценки были 0.70 и 0.56, для группы с высокой и с низкой точностью соответственно ($t = 3.72$, $p < 0.001$). Однако, распределение девяти исходов, полученных после инструкции о низкой точности, были

незначительно ближе к распределению базового значения, чем распределение, полученное после инструкции о высокой точности. Для каждого испытуемого было вычислено произведение моментов между средней величиной, которую он предписал каждому из этих девяти исходов (по этим пяти описаниям) и базовым значением. С помощью этой корреляции можно измерить степень, в которой прогнозы испытуемого соответствуют распределению базового значения. Среднее значение индивидуальных корреляций равнялось 0.13 для испытуемых в группе с высокой точностью и 0.16 для испытуемых в группе с низкой точностью. Эта разница не является значимой ($t = 0.42$, $df = 103$). Данный пример суждения нарушает нормативную теорию прогнозирования, согласно которой любое уменьшение ожидаемой точности должно сопровождаться изменением предсказаний в сторону базового значения.

Так как манипуляция ожидаемой точностью не оказала влияния на прогнозы, две группы прогнозов были объединены. Последующие исследования были проведены тем же образом, что и в исследовании с Томом В. Для каждого описания, было вычислено две корреляции: (а) между средней величиной вероятности и средней величиной подобия и (б) между средним значением вероятности и средней величиной базового значения. Корреляции отражены в Таблице 2, с учетом исхода, оцененного как наиболее вероятный для каждого описания. Высокие значения корреляции были получены между прогнозом и подобием. И наоборот, не обнаружено систематической связи между базовым значением и прогнозом: значения корреляции широко изменялись в зависимости от того, являются ли наиболее репрезентативные исходы для каждого описания частыми или редкими.

Таблица 2. Корреляция средней величины вероятности со средним значением подобия и базовым значением.

	<i>Первое модальное предсказание</i>				
	<i>Юрис- пруден- ция</i>	<i>Программи- рование</i>	<i>Меди- цина</i>	<i>Библио- текарское дело</i>	<i>Управ- ление бизнесом</i>
Со средним значением подобия	0.93	0.96	0.92	0.88	0.88
С базовым значением	0.33	-0.35	0.27	-0.03	0.62

Как и в предыдущем случае, обсуждением базового значения здесь также пренебрегли. В теории статистики, игнорировать базовое значение можно

только в том случае, когда есть уверенность в безошибочности прогноза. Во всех других случаях, должен быть найден соответствующий компромисс между распределением, ожидаемым в соответствии с описанием, и распределением базового значения. Маловероятно, что поверхностное описание четырнадцатилетнего ребенка, основанное на единственном интервью, может служить оправданием степени безошибочности, которую ожидали увидеть испытуемые в прогнозах.

После пяти описаний личности, перед испытуемыми была поставлена дополнительная задача:

О Доне известно только то, что он участвовал в первоначальном исследовании и сейчас является аспирантом первого курса. Укажите вашу оценку, а также сообщите, насколько Вы уверены в правильности оценки этого случая.

Для Дона корреляция между средней величиной вероятности и оцененным базовым значением равнялась 0.74. Таким образом, знание базового значения, которое не учитывалось при наличии описания, использовалось, когда никакой дополнительной информации не было.

Доказательства, основывающиеся на априорной вероятности либо на информации об индивидууме

Следующее исследование представляет собой более тщательную проверку гипотезы, что интуитивные предсказания зависят от репрезентативности и относительно независимы от априорной вероятности. В этом исследовании априорные вероятности были сформулированы так, чтобы сделать их более выразительными и сопоставимыми со способом ответа. Испытуемым зачитали следующий рассказ:

Группа психологов проинтервьюировала и провела личностный тест для 30 инженеров и 70 адвокатов, причем все они добились успеха в своих областях. На основе этой информации, были написаны краткие описания личности 30 инженеров и 70 адвокатов. В ваших анкетах Вы найдете пять описаний, выбранные наугад из 100 имеющихся в наличии описаний. Для каждого описания, пожалуйста, укажите вероятность (от 0 до 100) того, что описанный человек является инженером.

Та же задача была выполнена группой экспертов, которые давали высоко точные оценки вероятностей для различных описаний. Вам будет выплачена премия, в зависимости от того насколько близкими Ваши оценки окажутся к оценкам группы экспертов.

Инструкции были даны группе из 85 испытуемых (группа с малым количеством инженеров, или L группа). Испытуемые в другой группе (группа с большим количеством инженеров, H группа; N = 86) получили идентичные инструкции, за исключением априорной вероятности: им указали, что из 100 изученных человек, 70 являются инженерами и 30 адвокатами. Испытуемым обеих групп были выданы одни и те же описания. Вот одно из них:

Джек – 45-летний мужчина. Женат, имеет четверых детей. В общем, консервативен, осторожен и честолюбив. Не проявляет интереса к политическим и социальным проблемам, тратит большую часть свободного времени на свои многочисленные хобби, такие как столярное дело, плавание и решение математических загадок.

Вероятность того, что Джек является одним из 30 инженеров в выборке из 100 человек = _____ %.

После пяти описаний, испытуемые столкнулись с пустым описанием:

Предположим, что у Вас нет никакой информации относительно человека, выбранного наугад из совокупности.

Вероятность того, что этот человек является одним из 30 инженеров в выборке из 100 человек = _____ %

И в группе с большим количеством инженеров, и в группе с малым количеством инженеров половину испытуемых просили оценить, для каждого описания, вероятность того, что описанный человек – инженер (как в примере выше), в то время как другие испытуемые оценивали для каждого описания вероятность того, что описанный человек – адвокат. Эта манипуляция оказалась неэффективной. Средние вероятности, приписываемые исходам *инженер* и *адвокат* в двух различных формах, в сумме составляли приблизительно 100% для каждого описания. Следовательно, данные для обеих форм были объединены, и результаты представлены с точки зрения исхода *инженер*.

Структура данного эксперимента позволяет вычислить подходящий образец суждений. Вычисления основываются на формуле Байеса, в форме вероятностей. Пусть O обозначает вероятность того, что отдельно взятое описание принадлежит скорее инженеру, чем адвокату. Согласно правилу Байеса, $O = Q \times R$, где Q обозначает априорную вероятность того, что наугад выбранное описание принадлежит скорее инженеру, чем адвокату; и R – отношение вероятности для отдельно взятого описания, то есть отношение вероятности того, что человек, наугад выбранный из совокупности инженеров, будет описан как инженер, к вероятности того, что человек, отобранный наугад из совокупности адвокатов, будет описан как адвокат.

Для группы с большим количеством инженеров, которой было сказано, что выборка состоит из 70 инженеров и 30 адвокатов, априорное отношение Q_H равно 70/30. Для группы с малым количеством инженеров Q_L равняется 30/70. Таким образом, для каждого описания, отношение апостериорной вероятности для этих двух групп равно

$$\frac{O_H}{O_L} = \frac{O_H \cdot R}{O_L \cdot R} = \frac{O_H}{O_L} = \frac{7/3}{3/7} = 5.44$$

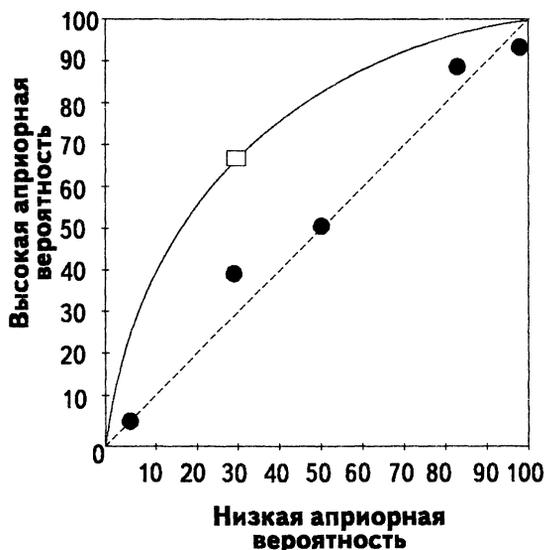


Рис. 1. Оцененная средняя вероятность (для инженеров) для пяти описаний и для “пустого” описания (квадратный символ) при высоких и низких априорных вероятностях. (Изогнутая линия показывает правильное распределение согласно правилу Байеса.)

Так как соотношение вероятностей в этой формуле сокращается, то для всех описаний должно быть получено одинаковое значение O_H/O_L . Следовательно, в данном случае, правильный эффект манипуляции априорной вероятности можно вычислить без знания соотношения вероятностей.

На Рисунке 1 можно увидеть средние оценки вероятности для каждого описания, как при высокой, так и при низкой априорной вероятности. Для каждого описания, средняя оценка вероятности, при высоком априорном соотношении вероятностей ($Q_H = 70/30$) соответствует на графике средней оценке, при низкой априорной вероятности ($Q_L = 30/70$). Согласно нормативному уравнению, полученному в предыдущем параграфе, все точки должны лежать на кривой Байеса. В действительности, только пустой квадрат, который соответствует “пустым” описаниям, лежит на этой линии: при отсутствии описания, испытуемые решили, что оценка вероятности будет 70% при высокой априорной вероятности и 30% при низкой априорной вероятности. В остальных же пяти случаях, точки лежат недалеко от линии идентификации.

Эффект априорной вероятности, хотя и небольшой, является статистически значимым. Для каждого испытуемого средняя оценка вероятности, была вычислена для всех случаев, кроме случая с пустым описанием. Среднее значение величин составило 50% для группы с низкой априорной вероятностью и 55% для группы с высокой априорной вероятностью ($t = 3.23$, $df = 169$, $p < 0.01$). Однако из Рисунка 1 следует, что каждая точка лежит ближе к линии идентификации, чем к кривой Байеса. Следовательно, мож-

но сделать вывод о том, что манипуляция априорным распределением оказала минимальное влияние на субъективную вероятность. Как и в предыдущем эксперименте, испытуемые применили свои знания о априорной вероятности только тогда, когда им не дали никакого подтверждения. Согласно гипотезе репрезентативности, априорная вероятность не принималась во внимание, когда информация об индивидууме была доступна.

Сила этого эффекта демонстрируется ответами на следующее описание:

Дик – 30-летний мужчина. Женат, еще не имеет детей. Очень способный и мотивированный сотрудник, подает большие надежды. Пользуется признанием коллег.

Это описание было построено таким образом, чтобы быть полностью неинформативным в отношении профессии Дика. Испытуемые обеих групп пришли к согласию: средние оценки составили 50% (см. Рисунок 1). Разница между ответами на это описание и на “пустое” описание разъясняет ситуацию. Очевидно, люди реагируют по-разному, когда не получают никакого описания, и когда дано бесполезное описание. В первом случае, априорная вероятность принимается во внимание; во втором, априорная вероятность игнорируется.³

Существуют ситуации, когда априорная вероятность, играет более существенную роль. Во всех примерах, рассмотренных до настоящего момента, четко выраженные стереотипы были связаны с альтернативными исходами, и, мы предполагаем, что суждения управлялись, степенью, в которой описания оказывались репрезентативными по отношению к этим стереотипам. В других задачах исходы естественно рассматривать как части измерения. Предположим, например, что Вас попросили оценить вероятность того, что каждый из нескольких студентов получит стипендию. В этой задаче, не существует хорошо обозначенных стереотипов студентов, которые получают или не получают стипендию. Скорее, естественно рассматривать исход (то есть, получение стипендии) как некий отрезок на прямой академических достижений или способностей. Априорная вероятность, то есть процент стипендий в определенной группе, могла использоваться, чтобы устранить пределы. Следовательно, они, вряд ли игнорируются. Кроме того, мы могли бы ожидать, что экстремальные априорные вероятности будут оказывать некоторое воздействие даже при наличии четких стереотипов исходов. Точная схема условий, при которых первичная информация используется или отвергается, требует дальнейшего исследования.

Один из основных принципов статистического прогнозирования гласит, что априорная вероятность, которая суммирует наши знания относительно проблемы до того, как мы получили определенное описание, остается уместной даже после того, как такое описание получено. Правило Байеса переводит этот качественный принцип в мультипликативное соотношение между

³ но см с.185

априорной вероятностью и отношением вероятности. Нашим испытуемым не удалось объединить априорную вероятность и дополнительную информацию. Когда им было предоставлено описание, каким бы неинформативным или недостоверным оно ни было, Тома В. или Дика (инженер/адвокат), испытуемые, очевидно, чувствовали, что распределение профессий в его группе больше не соответствовало действительности. Неспособность оценить роль априорной вероятности, если дано определенное описание – это, возможно, одно из наиболее существенных отклонений интуиции от нормативной теории прогнозирования.

Числовое прогнозирование

Фундаментальное правило нормативной теории прогнозирования гласит, что изменчивость предсказаний на основании ряда случаев, должна отражать точность прогноза. Когда точность прогноза совершенна, человек предсказывает то значение критерия, которое фактически будет получено. Когда неопределенность максимальна, во всех случаях прогнозируется фиксированное значение величины. (При прогнозировании категории, предсказывается наиболее частая категория. При числовом прогнозировании предсказывается среднее значение, мода, медиана или некоторая другая величина в зависимости от функции потерь.) Таким образом, изменчивость прогнозов равна изменчивости критерия, когда точность прогноза совершенна, и изменчивость прогноза равна нулю, когда точность прогноза нулевая. При средней точности прогноза, изменчивость прогноза принимает промежуточное значение, то есть прогнозы регрессивны относительно критерия. Таким образом, чем больше неопределенность, тем меньше изменчивость прогнозов. Прогнозы, сделанные на основании репрезентативности, не следуют этому правилу. В предыдущей главе мы продемонстрировали, что люди не делали регресс в сторону более частых категорий при уменьшении ожидаемой точности предсказаний. Эта глава демонстрирует аналогичную неспособность в контексте числового прогнозирования.

Прогнозирование исходов либо оценка исходных данных

Предположим, Вам сообщают, что консультант-психолог описал студента первого курса как умного, уверенного в себе, начитанного, трудолюбивого и любознательного. Рассмотрим два типа вопросов, которые можно было бы задать по этому описанию:

(А) *Оценка*: Какое мнение складывается у Вас относительно способностей к учебе после этого описания? Какой процент описаний первокурсников, по Вашему мнению, произвел бы на Вас большее впечатление? (Б) *Прогнозирование*: Как Вы думаете, какие средние оценки получит этот студент? Какой процент первокурсников получит более высокую среднюю оценку?

Между этими двумя вопросами имеется важное различие. В первом случае Вы оцениваете исходные данные; а во втором, Вы прогнозируете исход. Так как во втором вопросе существует большая неопределенность, чем в первом, ваше прогнозирование должно быть более регрессивным, чем ваша оценка. То есть процент, который Вы даете в качестве прогноза, должен быть ближе к 50%, чем тот процент, который Вы даете в качестве оценки. Чтобы подчеркнуть различие между этими вопросами, предположим, что описание является неточным. Это никак не должно повлиять на вашу оценку: оценка описаний относительно того, какое они производят впечатление, не зависит от их точности. С другой стороны, в прогнозировании вы должны быть регрессивны настолько, чтобы, предположить, что описание неточно или ваше прогнозирование недостоверно.

Как бы там ни было, гипотеза репрезентативности гласит, что прогнозирование и оценка должны совпадать. Оценивая данное описание, люди выбирают число, которое является наиболее репрезентативным по отношению к описанию. Если люди предсказывают на основе репрезентативности, то в качестве прогноза они также выберут самое репрезентативное число. Следовательно, оценка и прогнозирование – это, по сути, одно и то же. Чтобы проверить эту гипотезу, было проведено несколько исследований. В каждом из этих исследований испытуемым давали описательную информацию касательно некоего набора случаев. Группа *оценки* определяла качество каждого описания относительно данной совокупности, а группа *прогнозирования* предсказывала будущее развитие событий. Суждения этих двух групп сравнивались, чтобы проверить, являются ли прогнозы более регрессивными, чем оценки.

В двух исследованиях испытуемым предложили описания студентов-первокурсников, составленные на основе интервью при поступлении. В первом исследовании каждое описание состояло из пяти прилагательных, описывающих интеллектуальные способности и черты характера, как в приведенном примере. Во втором исследовании описания представляли собой тексты длиной примерно в абзац, в которых описывались детали биографии студентов, и то, как они адаптировались в колледже на настоящий момент. В обоих исследованиях группу оценки попросили рассмотреть каждое из описаний, определив “процент студентов во всем классе, чьи описания свидетельствуют о больших способностях”. Группе прогнозирования дали те же самые описания и попросили сделать прогноз относительно того, какую среднюю годовую оценку получит этот студент и его класс по окончанию учебного года в процентилях.

Результаты обоих исследований отражены на Рисунке 2, который показывает для каждого описания среднее значение прогноза среднегодовой успеваемости в процентилях по отношению к средней оценке. Единственное систематическое различие между прогнозами и оценками наблюдалось в исследовании с прилагательными, когда предсказания были значительно

выше, чем соответствующие оценки. Стандартное отклонение предсказаний или оценок было подсчитано, опираясь на данные каждого испытуемого. Сравнение этих величин не дало значительной разницы изменчивости между группами оценки и прогнозирования в пределах значений исследуемых величин. В исследовании с прилагательными средняя величина стандартного отклонения составила 25.7 для группы оценки ($N = 38$) и 24.0 для группы прогнозирования ($N = 36$, $t = 1.25$, $df = 72$, не значимо). При исследовании описаний средняя величина стандартного отклонения составила 22.2 для группы оценки ($N=37$) и 21.4 для группы прогнозирования ($N = 63$, $t = 0.75$, $df = 98$, не значимо). В двух исследованиях и группа оценки, и группа прогнозирования производили одинаково экстремальные суждения, хотя первая группа предсказывала отдаленный объективный критерий на основе поверхностной информации, полученной из интервью; в то время как вторая группа оценивала впечатление, которое на нее произвело каждое описание. В статистической теории прогнозирования наблюдаемое равенство между оценками и прогнозами было бы возможно, если точность прогноза совершенна, а в целом, это условие не соблюдалось в исследовании.

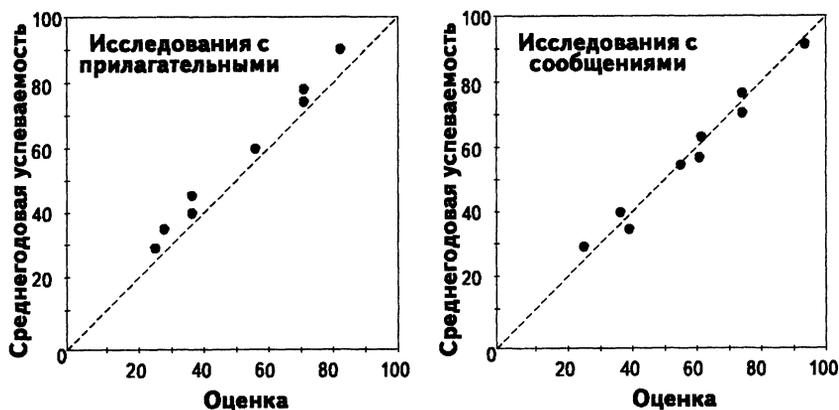


Рисунок 2. Предсказанная процентильная среднегодовая успеваемость как функция процентильной оценки для исследований с использованием прилагательных и сообщений.

Еще одно свидетельство равенства оценки и прогноза было получено Бейт (Beyth) в ее докторской диссертации. Она дала трем группам испытуемых семь абзацев текста, каждый из которых описывал поведение студента-практиканта на одном из уроков во время практики. В качестве испытуемых были привлечены изучающие статистику студенты университета Hebrew. Им сказали, что описания были получены из картотеки 100 преподавателей на-

чальной школы, которые, пятью годами ранее, окончили курс обучения. Испытуемых в группе оценки попросили оценить качество урока, описанного в тексте, длинной в абзац, в процентилях относительно данной совокупности. Испытуемых в группе прогнозирования попросили сделать прогноз (в процентилях) настоящего положения каждого преподавателя, то есть его полную компетентность через 5 лет после составления описания. И группа оценки, и группа прогнозирования выполнили обе задачи. Как и в исследованиях, описанных выше, разница между оценкой и прогнозом была несущественной. Этот результат оставался тем же и при сравнении между испытуемыми в группе, и при сравнении испытуемым собственных результатов. Хотя испытуемые знали, какая пропасть отделяет один пробный урок и время получения пятилетнего опыта работы, это знание не привело к тому, что их прогнозы стали более регрессивными, чем их оценки.

Прогнозирование или трансляция

Предыдущие исследования показали, что прогнозирования переменной не регрессивны при сравнении с оценками исходных данных в терминах этой переменной. В следующем исследовании мы покажем, что существуют ситуации, в которых прогнозирования значения переменной (успеваемости) не более регрессивны, чем простая трансляция этой переменной из одного масштаба в другой. В качестве результирующей переменной была выбрана среднегодовая оценка, потому что ее корреляции и свойства ее распределения хорошо известны испытуемым.

В эксперименте участвовали три группы испытуемых. Испытуемые во всех группах прогнозировали среднегодовую оценку 10 гипотетических студентов на основании только процентильной оценки, полученной каждым из этих студентов. Один и тот же набор процентильных оценок был представлен всем группам, но эти три группы получили различные интерпретации исходной переменной следующим образом:

1. *Процентиль среднегодовой оценки.* Испытуемым в Группе 1 ($N = 32$) сказали, что “для каждого из нескольких студентов вам предложат процентильную оценку, представляющую его успеваемость в течение первого года обучения, и Вас просят угадать его среднегодовую оценку в этом году”. Испытуемым объяснили, что “65 баллов, например, означает, что среднегодовая оценка, полученная этим студентом лучше, чем у 65% студентов его группы, и т.д.”.

2. *Умственная концентрация.* Испытуемым Группы 2 ($N = 37$) сообщили, что “тест на умственную концентрацию измеряет способность концентрироваться и извлекать всю информацию из сложных сообщений. Было обнаружено, что студенты с высокими среднегодовыми оценками склонны успешно проходить тесты на умственную концентрацию и наоборот. Однако выяснилось, что результаты теста на умственную концентрацию зависят от настроения и психического

состояния человека во время прохождения теста. Таким образом, один и тот же человек может получить различные результаты при многократном тестировании, в зависимости от того, хорошо ли он выспался предыдущей ночью или от самочувствия в этот день.

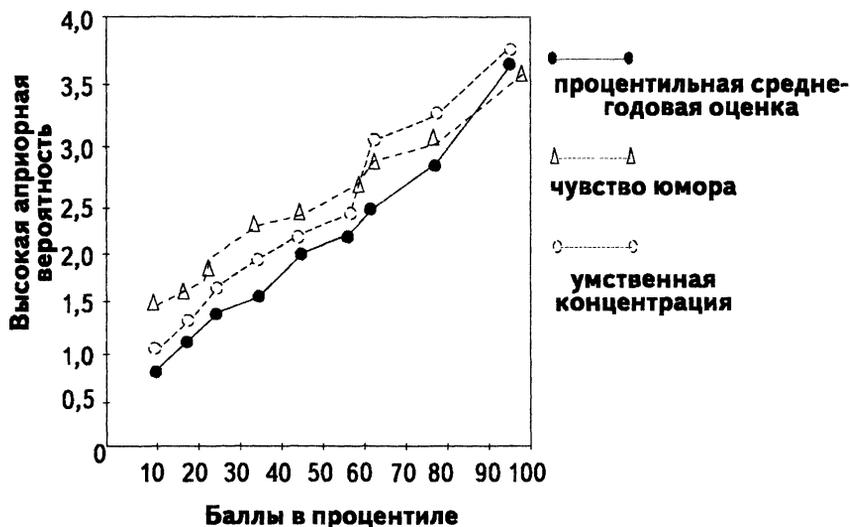


Рис. 3. Предсказания среднегодовой оценки в зависимости от процентильных оценок для трех переменных

3. Чувство юмора. Испытуемым в Группе 3 ($N = 35$) сообщили, что “тестирование чувства юмора измеряет способность людей придумывать остроумные надписи для комиксов и оценивать проявления юмора в различных формах. Было обнаружено, что студенты, показывающие высокие результаты при прохождении этого теста, как правило, получают более высокие среднегодовые оценки, чем студенты, которые показывают в этом тесте плохие результаты. Однако, невозможно предсказать среднегодовые оценки с высокой точностью, основываясь на чувстве юмора.”

В настоящем исследовании, все испытуемые прогнозировали среднегодовые оценки на основе одних и тех же процентильных оценок. Группа 1 просто перевела значения среднегодовой успеваемости в процентиле на шкалу среднегодовых оценок. С другой стороны, Группы 2 и 3 предсказывали среднегодовую оценку на базе более отдаленных исходных данных. Исходя из соображений нормативной теории, прогнозы этих групп должны быть более регрессивными, то есть менее изменчивыми, чем суждения Группы 1. Однако гипотеза репрезентативности предполагает другие результаты.

Группа 2 делала прогнозы на основе потенциально верного, но ненадежного теста на умственную концентрацию, который был представлен как мера успева-

емости. Мы выдвинули гипотезу, что предсказания этой группы были бы менее регрессивными при их сравнении с предсказаниями Группы 1. В общем, мы предполагаем, что балл успеваемости (например, среднегодовая оценка), который лучше всего представляет значение в процентиле, измеряя способности (например, умственную концентрацию), это тот, который соответствует тому же самому процентилю на шкале успеваемости.

Таблица 3. Средняя величина значения индивидуальных прогнозов для трех групп и результаты запланированного сравнения между группами 1 и 2, и между группами 2 и 3

Статистика	Группа				
	1. Среднегодовая оценка в процентиле	1 и 2	2. Умственная концентрация	2 и 3	3. Чувство юмора
Средняя спрогнозированная среднегодовая оценка	2.27	Не значимо	2.35	0.05	2.46
Стандартное отклонение прогнозов	0.91	Не значимо	0.87	0.01	0.69
Отклонение регрессии	0.030	Не значимо	0.029	0.01	0.022
Корреляция r	0.97	Не значимо	0.95	Не значимо	0.94

Так как ненадежность данных не влияет на репрезентативность, мы предполагаем, что прогнозы среднегодовой оценки на базе ненадежного теста на умственную концентрацию будут по существу идентичными предсказаниям среднегодовой оценки на базе среднегодовой оценки в процентиле. С другой стороны, ожидалось, что предсказания Группы 3 будут регрессивными, потому что чувство юмора обычно не рассматривается как мера успеваемости.

Средние прогнозы, приходящиеся на 10 баллов в процентиле по трем группам, показаны на Рисунке 3. Исходя из Рисунка 3, можно сказать, что прогнозы Группы 2 не более регрессивны, чем прогнозы Группы 1, в то время как прогнозы Группы 3 оказываются более регрессивными.

В пределах данных каждого отдельного испытуемого были вычислены 4 индекса: средняя величина его прогноза, стандартное отклонение его прогноза, отклонение регрессии спрогнозированной среднегодовой оценки на

исходном множестве, и корреляция между ними. Средние величины трех значений для этих трех групп показаны в Таблице 3.

Из таблицы видно, что испытуемые во всех трех группах привели правильные данные, что проявляется в высоких корреляциях между исходными данными и прогнозами (средние корреляции были получены за счет преобразования индивидуальных значений в z -значения Фишера). Результаты запланированных сравнений между Группами 1 и 2 и между Группами 2 и 3 подтверждают принцип, наблюдаемый на Рисунке 3. Нет существенных различий между прогнозами, сделанными на базе среднегодовой оценки в процентиле, и прогнозами, сделанными на основе теста на умственную концентрацию. Таким образом, люди не производили регресс при предсказании меры успеваемости на базе данных об уровне способностей, даже и ненадежных.

Предсказания, основанные на данных о чувстве юмора, с другой стороны, являются регрессивными, хотя и недостаточно. Отношение между среднегодовой оценкой и чувством юмора, выведенное из сравнения линий регрессии приблизительно 0.70. К тому же, прогнозы, основанные на данных о чувстве юмора значительно выше, чем прогнозы, основанные на данных об умственной концентрации. Кроме того, у прогнозов на основании данных об умственной концентрации существует тенденция превышать прогнозы, основанные на среднегодовой оценке в процентиле. Мы наблюдали это явление во многих исследованиях. Прогнозируя успеваемость человека на основе несовершенной информации, испытуемые проявляют снисходительность (Guilford, 1954). Вслед за уменьшением надежности, они все больше опираются на спрогнозированный уровень поведения.

Как ожидается, прогнозы будут по существу нерегрессивными, когда исходные переменные, а также переменные исхода рассматриваются как проявления той же самой черты. Пример таких прогнозов можно найти и в реальных условиях, например, Коллегия по отбору офицеров в израильской армии. Опытные офицеры, входящие в состав экзаменационной комиссии, обычно оценивают кандидатов по семибальной шкале по окончании нескольких дней экзаменов и наблюдений. В исследовательских целях для каждого прошедшего испытания кандидата им нужно было предсказать заключительный ранг, который он получит в офицерской школе. Из более чем 200 случаев обнаружилось, что предсказанное распределение рангов было практически идентичным фактическому распределению, с одним очевидным исключением: прогнозы неудач были менее часты, чем фактические неудачи. Как правило, частоты прогнозов в двух самых высоких категориях точно соответствовали фактическим частотам. Все испытуемые прекрасно знали, что исследование указывает на то, что надежность их прогнозов была только умеренной (от 0.20 до 0.40). Однако их предсказания были нерегрессивными.

Методологические соображения

Гипотеза репрезентативности гласит, что прогнозы не отличаются от оценок подобия, хотя нормативная теория статистики подразумевает то, что прогнозы должны быть менее экстремальными, чем оценки. Поэтому тестирование гипотезы репрезентативности требует исследования, в котором прогнозы сравниваются с другим типом суждений. Варианты двух сравнительных анализов использовались в исследованиях, о которых говорится в этой части книги.

В одном исследовании, обозначенном А-ХУ, различные группы испытуемых оценивали две переменные (Х и У) на основе одной и той же исходной информации (А). В случае Тома В., например, двум различным группам давали одну и ту же исходную информацию (А), то есть описание личности. Одна группа оценивала исходы с точки зрения подобия (Х), в то время как другая оценивала исходы с точки зрения вероятности (У). Точно так же в нескольких исследованиях, посвященных прогнозированию числа, различным группам давали одну и ту же информацию (А), например, список пяти прилагательных, описывающих студента. Одна группа проводила оценивание (Х), а другая — прогнозирование (У).

В другом исследовании, обозначенном АВ-Х, две различных группы испытуемых оценивали одну и ту же переменную результата (Х) на основе различной исходной информации (А и В). В исследовании “инженер/адвокат”, например, две различных группы проводили одну и ту же оценку (Х) вероятности того, что отдельно взятый человек является инженером. Им предоставили краткое описание его личности и различную информацию (А и В) относительно частот базового значения инженеров и адвокатов. В контексте прогнозирования числа, различные группы прогнозировали среднегодовую оценку (Х) на основе различных переменных — процентиля среднегодовой оценки (А) и умственной концентрации (В).

Гипотеза репрезентативности была подтверждена при сравнительном анализе, который показал, что в отличие от нормативной модели, прогнозы не более регрессивны, чем оценки или суждения подобия. Также возможно выяснить, являются ли интуитивные прогнозы регрессивными по сравнению с фактическими исходами, или с исходными данными, когда входные данные и исходы измеряются по одной и той же шкале. Даже, в том случае, когда предсказания не более регрессивны, чем трансляции, мы ожидаем, что они будут слегка регрессивными по сравнению с исходами из-за известной ошибки тяготеть к центральной тенденции (Jonson, 1972; Woodworth, 1938). В большом разнообразии оценочных задач, включая простую трансляцию исходных данных с одной шкалы на другую, испытуемые имеют тенденцию избегать крайних ответов и уменьшать изменчивость своих суждений (Stevens и Greenbaum, 1966). Из-за этого предубеждения, суждения будут регрессивны по сравнению с входными данными или с исходами. Ис-

следования, используемые в этой книге, нейтрализуют это влияние, сравнивая две оценки, подверженные одному и тому же предубеждению.

Настоящая серия исследований была посвящена ситуациям, в которых люди делают прогнозы на основе информации, которая доступна им до проведения исследования, в форме стереотипов (например, инженера) и ожиданий относительно отношений между переменными. Обратная связь исхода не обеспечивалась, и количество оценок, которые должен был осуществить каждый тестируемый, было мало. Напротив, самые ранние исследования прогнозирования касались изучения функциональных или статистических отношений между переменными, с которыми испытуемые не имели никакого предшествующего знакомства. Эти исследования обычно стимулируют большое количество попыток и различных форм обратной связи исхода. (Часть этой литературы была рассмотрена в работах Словика и Лихтенштейна, 1971.) В исследовании повторных прогнозов с обратной связью, испытуемые вообще предсказывают, выбирая результаты так, чтобы полная последовательность или образец предсказаний были высоко репрезентативными по отношению к распределению исходов. Например, испытуемые в исследованиях, изучающих вероятность, производят последовательности прогнозов, которые едва соответствуют статистическим характеристикам последовательности исходов. Точно так же испытуемые в задачах числового прогнозирования приблизительно воспроизводят график рассеивания, то есть объединенное распределение входных данных и исходов (см., например Grey, 1968). Для того чтобы это сделать, испытуемые прибегают к смешанной стратегии: для любых исходных данных они производят распределение различных предсказаний. Эти прогнозы отражают то, что любой набор исходных данных вызывает различные исходы при различных испытаниях. Очевидно, правила прогнозирования различны в этих двух парадигмах, хотя репрезентативность имеет отношение к обеим. В парадигме обратной связи испытуемые производят последовательности ответов, представляющие целостные модели связи между исходными данными и исходами. В ситуациях, исследуемых в этой книге, испытуемые выбирают прогноз, который лучше всего представляет их впечатления относительно каждого отдельного случая. Эти два подхода ведут к различным нарушениям нормативного правила: представление неопределенности с помощью смешанной стратегии в парадигме обратной связи и уменьшение неуверенности за счет прогнозирования с помощью оценок в данной парадигме.

Достоверность и иллюзия валидности

Как демонстрируется в предшествующих разделах книги, люди прогнозируют, выбирая исход, который является наиболее репрезентативным по отношению к исходным данным. Мы предполагаем, что степень достоверности, которую человек предполагает в предсказании, отражает степень, в которой выбранный исход является более репрезентативным по отношению

к исходным данным, чем другие исходы. Главный показатель репрезентативности в контексте предсказания чисел с мультиатрибутивными исходными данными (например, оценки профилей) — это упорядоченность или взаимосвязанность исходных данных. Чем более упорядочены исходные данные, тем более репрезентативной будет казаться предсказанная величина и тем большей будет достоверность этого прогноза. Например, люди прогнозируют среднее значение В с большей достоверностью на основе оценок В в двух разных предварительных испытаниях, чем на основе А и С. Действительно, было обнаружено, что внутренняя изменчивость или несогласованность исходных данных уменьшает достоверность прогнозов (Slovic, 1966).

Невозможно преодолеть заблуждение, для которого упорядоченные профили позволяют большую прогнозируемость, чем неупорядоченные. Стоит отметить, однако, что это убеждение несовместимо с обычно применяемой многовариантной моделью прогнозирования (то есть, нормальной линейной моделью), в которой ожидаемая точность прогноза не зависит от изменчивости в пределах профиля.

С упорядоченными профилями испытуемые обычно сталкиваются, когда прогнозируют на основании высоко коррелированного множества. Неупорядоченные профили, наоборот, являются более частыми при низких корреляциях. Поскольку достоверность растет с увеличением упорядоченности, достоверность будет высокой, когда исходные переменные высоко коррелированы. Как бы там ни было, если даны исходные переменные заданной валидности, многократная корреляция с критерием обратно пропорционально связана с корреляциями между исходными данными. Таким образом, возникает парадоксальная ситуация, при которой высокие взаимосвязи между исходными данными увеличивают надежность и уменьшают достоверность.

Чтобы продемонстрировать этот эффект, мы попросили испытуемых сделать прогноз среднегодовой оценки на основе двух пар тестов на способности. Испытуемым сказали, что одна пара тестов (творческое мышление и способность к символизации) была высоко коррелированной, в то время как другая пара тестов (гибкость ума и систематичность рассуждения) не была коррелированной. Множество, с которым они столкнулись, оправдало эти ожидания. (Для половины испытуемых ярлыки коррелированные и некоррелированные для пары тестов были изменены на противоположные). Испытуемым сообщили, что “все тесты были одинаково действенными в прогнозировании успеваемости в колледже”. Конечно, в этой ситуации, более высокая точность прогнозов может быть достигнута с некоррелированной парой тестов, чем с коррелированной парой. Но, как и ожидалось, испытуемые были более уверены в прогнозах на базе коррелированных испытаний, для полного диапазона множества прогнозирования ($t = 4.80$, $df = 129$, $p < 0.001$). То есть, они были более уверены в контексте с более низкой прогнозирующей валидностью.

Другой вывод, который может быть получен из исследований прогнозирования, включая и наше – это то, что достоверность представляет собой функцию спрогнозированного уровня действия, график которой по форме напоминает букву J (см. Johnson, 1972). Испытуемые прогнозируют высокую успеваемость с большой степенью достоверности, и они проявляют большую степень достоверности, прогнозируя полную неудачу, чем среднюю успеваемость. Как мы видели ранее, интуитивные предсказания часто недостаточно регрессивны. Поэтому несоответствия между прогнозами и исходами, являются самыми большими в экстремумах. J-образная функция достоверности подразумевает, что испытуемые наиболее уверены в предсказаниях, которые, скорее всего, ложны.

Предшествующий анализ показывает, что факторы, которые увеличивают достоверность, например, упорядоченность и экстремальность, часто отрицательно коррелированы с точностью прогнозов. Таким образом, люди склонны быть слишком уверенными в ошибочных суждениях, явление, которое можно назвать *иллюзией валидности*. Подобно прочим ошибкам восприятия и суждения, иллюзия валидности часто сохраняется даже тогда, когда признан ее иллюзорный характер. Интервьюируя кандидата, например, многие из нас испытывают большую уверенность по поводу прогноза его будущей деятельности, несмотря на знание о том, что интервью часто бывают ошибочными.

Представления относительно регресса

Последствия регресса окружают нас повсюду. В жизни у самых выдающихся отцов бывают посредственные сыновья, у замечательных жен заурядные мужья, неадаптированные имеют тенденцию приспособливаться, а от счастливых, в конечном счете, удача отворачивается. Несмотря на эти факторы, люди не приобретают надлежащее понятие регресса. Во-первых, они не ожидают проявления регресса во многих ситуациях, где он должен произойти. Во-вторых, что подтвердит любой преподаватель статистики, чрезвычайно трудно приобрести надлежащее понятие регресса. В-третьих, когда люди наблюдают регресс, они обычно изобретают для этого явления ложные динамические объяснения.

Что же делает понятие регресса противоречащим интуиции, который трудно приобрести и применить? Мы утверждаем, что главный источник трудностей – то, что эффекты регресса обычно нарушают интуицию, которая говорит нам, что спрогнозированный исход должен быть максимально репрезентативен по отношению к исходной информации.⁴

⁴ Ожидание того, что каждый значимый акт поведения высоко репрезентативен по отношению к исполнителю, может объяснить, почему как обыватели, так и психологи постоянно удивляются незначительным корреляциям среди, как кажется, взаимозаменяемых измерений честности, рискованности, агрессии и зависимости (Mischel, 1968).

Чтобы проиллюстрировать постоянство нерегрессивной интуиции, не смотря на то, что она подчиняется статистике, мы представили следующую проблему нашей выборке аспирантов, специализирующихся в психологии:

Проблема тестирования. Человек, выбранный наугад, обладает коэффициентом интеллекта, равным 140. Предположим, что коэффициент интеллекта представляет собой сумму “истинного” количества очков и случайную ошибку измерения.

Пожалуйста, назовите верхние и нижние пределы достоверности 95% для истинного коэффициента интеллекта этого человека. То есть назовите такой верхний предел, при котором Вы уверены на 95%, что истинный коэффициент интеллекта, фактически, ниже, чем эта цифра, и такой нижний предел, что Вы уверены на 95%, что истинный коэффициент интеллекта фактически выше.

В этой задаче, испытуемых попросили считать наблюдаемый коэффициент интеллекта суммой “истинного” показателя интеллекта и компонента ошибки. Так как наблюдаемый уровень интеллекта значительно выше среднего, более вероятно, что компонент ошибки положителен и что этот человек на последующих тестах получит более низкий результат. Большинство испытуемых (73 из 108), однако, установили интервалы достоверности симметрично относительно 140, не показав никакого ожидания регресса. Из оставшихся 35 испытуемых, 24 задали интервалы регрессивной достоверности, а 11 задали контррегрессивные интервалы. Таким образом, большинство испытуемых проигнорировало влияние ненадежности исходных данных и прогнозировало так, как будто величина 140 была истинным показателем интеллекта. В этой части книги неоднократно обращается внимание на тенденцию предсказывать безошибочность исходной информации.

Возникновение регресса также иногда признается, либо потому что мы обнаруживаем эффект регресса в наших собственных наблюдениях, либо потому что нам явно говорят, что регресс произошел. Когда эффект регресса обнаружен, он обычно рассматривается как систематическое изменение, которое требует независимого объяснения. Действительно, в социальных науках предлагалось много ложных объяснений эффектов регресса⁵. Использовались динамические принципы, чтобы объяснить, почему бизнес, одно время очень успешный, имеет тенденцию к ухудшению впоследствии, и почему обучение интерпретации выражений лица полезно для тех учащихся, которые показали плохие результаты на предварительных экзаменах, и вредно для тех, кто показал самые хорошие результаты. Некоторые из этих объяснений не были бы предложены, если бы их авторы поняли, что если даны две переменные равной изменчивости, следующие два утверждения логически эквивалентны: (а) Y регрессивен относительно X; (б) корреляция

⁵ Для освещения обсуждений ошибок в исследованиях, связанных с регрессом, см., например, Campbell (1969) и Wallis и Roberts (1956).

между Y и X – меньше единицы. Поэтому объяснение регресса является равносильным объяснению того, почему корреляция меньше единицы.

В качестве заключительной иллюстрации того, насколько трудно распознать и должным образом интерпретировать регресс, рассмотрим следующий вопрос, который был задан нашей выборке аспирантов. Описанная задача фактически возникла в опыте одного из авторов.

Задача обучения. Инструкторы в летной школе использовали политику последовательного положительного поощрения, рекомендуемую психологами. Они устно поощряли каждое успешное выполнение маневра в полете. После некоторого времени применения этого подхода обучения, инструкторы заявили, чтобы вопреки психологической доктрине, высокая похвала за хорошее выполнение сложных маневров, обычно приводит к ухудшению их выполнения при следующей попытке. Что должен ответить психолог?

Регресс неизбежен в летных маневрах, потому что его выполнение не является абсолютно надежным, и прогресс при их последовательном выполнении идет медленно. Следовательно, пилоты, которые исключительно успешно проявили себя на одном испытании, вероятно, проявят себя хуже на следующем, независимо от реакции инструкторов на их изначальный успех. Опытные инструкторы летной школы фактически обнаружили регресс, но приписали это вредному влиянию поощрения. Эта история из жизни иллюстрирует печальный аспект человеческого положения. Мы обычно хвалим других, когда они хорошо себя ведут, и наказываем их, когда они ведут себя плохо. Только благодаря регрессу, они будут показывать хорошие результаты после наказания и плохие после похвалы. Следовательно, перед нами жизненная схема, согласно которой нас часто хвалят за то, что мы наказываем других, и наоборот.

Ни один из аспирантов, отвечавших на этот вопрос, не предположил, что регресс мог быть причиной проблемы. Вместо этого, они предложили, что устное поощрение могло быть неэффективным для пилотов, или что оно могло привести к самонадеянности. Некоторые аспиранты даже сомневались относительно валидности впечатлений инструкторов и обсуждали возможные источники предубеждения в их восприятии ситуации. Эти респонденты несомненно были полностью ознакомлены с понятием статистического регресса. Однако они не сумели распознать случай регресса, когда он не был изложен в знакомых терминах роста отцов и сыновей. Очевидно, одно лишь обучение статистике не меняет фундаментальную интуицию в вопросах неопределенности.

5. Изучение репрезентативности

Майя Бар-Хиллер

Даниель Канеман и Амос Тверски предположили, что при оценке вероятности неопределенных событий люди часто обращаются к эвристике или к эмпирическим правилам, которые почти не коррелируют (если коррелируют вообще) с переменными, которые фактически определяют вероятность события. Одна из таких эвристик — *репрезентативность*, определяемая как субъективная оценка степени, в которой рассматриваемое событие “является в существенных свойствах подобным своей исходной совокупности” или “отражает существенные особенности процесса, который его породил” (Kahneman & Tversky, 1972b, с. 431, 3). Хотя в некоторых случаях более вероятные события оказываются также более репрезентативными (и наоборот), уверенность относительно репрезентативности случая как показателя его вероятности может привести к двум видам систематических ошибок в суждении. Во-первых, это может придать чрезмерный вес переменным, которые влияют на репрезентативность события, а не на его вероятность. Во-вторых, это может уменьшать важность переменных, крайне необходимых для определения вероятности события, но не связанных с его репрезентативностью.

Концепция репрезентативности иногда критиковалась как слишком неопределенная и неуловимая, возможно, потому что в ней не хватает общего определения. Однако мы не утверждаем, что невозможно оценить репрезентативность независимо от оценки вероятности, — заключение, которое часто подразумевалось критиками. Например, в исследовании “Тома В.”, Канеман и Тверски (1973, 4) определяли репрезентативность как подобие некоторого человека, Тома В., “типичному аспиранту в ... (нескольких) областях специализации” (1973, с. 238) и оценивали ее независимо от вероятности того, что Том В. действительно занимался этими областями науки. В других исследованиях, независимая оценка на основе репрезентативности не применялась только потому, что читатели могли без труда провести ее самостоятельно с помощью мысленного эксперимента.

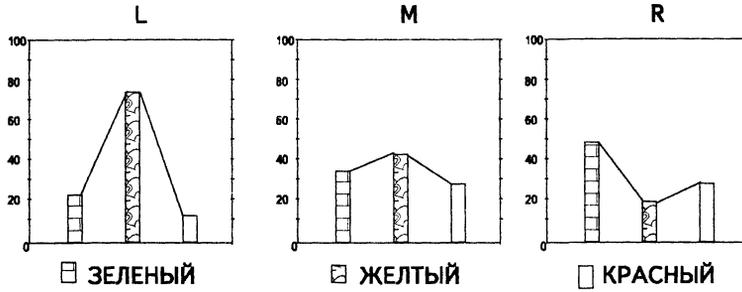


Рис. 1. Характерный стимул, использованный при изучении подобию и вероятности
(Источник: Бар-Хиллер, 1974).

Подобие и вероятность

Вскоре после того, как было введено понятие репрезентативности, я провела небольшое исследование, в котором репрезентативность оценивалась, как в соответствии с априорными критериями (что будет разъяснено позже в этой главе), так и экспериментально, в соответствии с оценками испытуемых. Затем я сравнила эти оценки с независимо собранными суждениями о вероятности (Bar-Hillel, 1974). Преимуществом изучаемых задач являлась их уникальность, четко определенная вероятность, а также простота визуального представления, позволяющая легко делать суждения о видимом подобии.

В качестве стимулов в этом исследовании были выбраны тройки гистограмм, обозначенные L, M и R для левой, средней и правой гистограммы, соответственно. Пример приведен на Рисунке 1. Каждая диаграмма состоит из трех полос, которые окрашены, слева направо, в зеленый, желтый и красный цвет. В длину каждая полоса поделена на 5 частей, а длины этих трех полос в пределах каждой гистограммы, в сумме составляют 100.

Все тройки гистограмм (L, M, R), используемые в этом эксперименте обладали следующими свойствами:

1. Каждая полоса в средней гистограмме является средней по длине между соответствующими полосами в гистограммах слева и справа (однако если, средняя длина левой и правой полос не достигала числа, кратного пяти, средняя полоса округлялась к ближайшему числу, кратному пяти.) Например, на Рисунке 1, длина зеленой полосы в M-гистограмме = 35, что находится между 20 и 50 (соответствующие длины полос в L и R гистограммах). То же можно сказать про желтые и красные полосы. (Обратите внимание, однако, что длины этих полос округлены: от 42.5 до 40 и от 22.5 до 25.)

2. Порядок длин полос в М-гистограмме совпадает с порядком или в L- или в R-гистограмме, но не с двумя, так как L- и R-гистограмма всегда упорядочены по-разному. Например, на Рисунке 1 порядок M и L один и тот же, причем красная полоса – самая короткая, а желтая – самая высокая в обеих диаграммах, что отличается от порядка в R диаграмме, где желтая полоса самая короткая. Это подчеркивается тонкой линией, соединяющей вершины всех полос в гистограмме, хотя ось X была номинальной переменной.

3. Предположим, что M-гистограмма интерпретируется как описание триномиальной совокупности, а L и R-гистограммы интерпретируются как выборки из 20 субъектов, которые выбраны из этой популяции. Тогда выборка, чей порядок совпадает с порядком M-гистограммы, является менее вероятной выборкой. Например, на Рисунке 1, M описывает триномиальную совокупность с параметрами 0.35, 0.40, и 0.025. Вероятность того, что взята выборка R (то есть, 10 зеленых, 3 желтых и 7 красных бусинок), а не выборка L (то есть, 4 зеленых, 14 желтых и 2 красные бусинки) – приблизительно 8 к 5, т. к.

$$\frac{P(L | M)}{P(R | M)} = \frac{(20! / 4!14!2!)}{(20! / 10!3!7!)} \times \frac{(0.35)^4 (0.40)^4 (0.25)^2}{(0.35)^{10} (0.40)^3 (0.25)^7} = 0.6127$$

Таким образом, R – более вероятная выборка.

4. Предположим, наоборот, что L и R интерпретируются как совокупности, а M как выборка. Появление выборки M наименее вероятно из совокупности, с которой у нее совпадает порядок. Например, на Рисунке 1, M описывает выборку из 7 зеленых, 8 желтых и 5 красных бусинок. Вероятность того, что M скорее взята из совокупности R (то есть, параметры 0.50, 0.15, и 0.35), чем из совокупности L (то есть, параметры 0.20, 0.70, и 0.10) – приблизительно 7 к 5, т.к.

$$\frac{P(M | L)}{P(M | R)} = \frac{(20! / 7!8!5!)}{(20! / 7!8!5!)} \times \frac{(0.20)^7 (0.70)^8 (0.10)^5}{(0.50)^7 (0.15)^8 (0.35)^5} = 0.7017$$

Таким образом, R – более вероятно совокупность.

Двадцать восемь таких троек гистограмм были показаны трем различным группам испытуемых, которые руководствовались тремя различными наборами инструкций. Одной группе было сказано, что каждая гистограмма описывает триномиальное распределение, где L и R описывают две совокупности, а M описывает выборку, взятую из одной из этих совокупностей. 25 испытуемых в этой группе оценивали вероятность того, взята ли выборка, отображенная в M, из совокупности L или из совокупности R. Другой группе сказали, что M описывает триномиальную совокупность, в то время как L и R – две триномиальные выборки. 26 испытуемых в этой группе оценивали вероятность того, какая из этих двух выборок взята из данной сово-

купности. Этим группам сначала предоставили большой стеклянный сосуд, заполненный зелеными, желтыми и красными бусинками. Им показали, как содержимое сосуда (то есть, совокупность) может быть описано соответствующей гистограммой. В их присутствии была наугад взята выборка из 20 бусинок, а также было продемонстрировано ее представление в виде гистограммы. Испытуемых попросили считать, что L и R являются совокупностями, а M представляет выборку 20 бусинок, или же L и R — это выборки, а M представляет совокупность, в зависимости от группы, к которой принадлежали испытуемые.

		A				B	
		L	R			L	R
Подобие	R	2	15		R	4	13
	L	10	1		L	11	0
		Вероятность выборки				Вероятность совокупности	
		$\phi=0.79$				$\phi=0.75$	

Рис. 2. Совместное распределение субъективных оценок L и R по подобию в сравнении с вероятностью (A) L и R – выборки из совокупности M или (B) – M выборка из совокупности L или R.

(Источник: Бар-Хиллер, 1974.)

Третьей и последней группе, состоящей из 25 испытуемых, не было представлено интерпретации гистограмм, и их просто просили оценить, какая из этих двух гистограмм, L или R, более походила на среднюю гистограмму M.

Стимул рассматривался как L или R согласно гистограмме, выбранной большинством испытуемых отдельно для каждого из этих трех критериев ((1) более вероятно, что производит выборку M; (2) более вероятно, что по-явится из совокупности M; (3) более подобен M).

Двумя основными результатами этого исследования были:

1. Эмпирические оценки подобию, представленные испытуемыми, абсолютно совпали с априорным критерием подобию, выраженным в свойстве 2, обозначенном выше. Другими словами, для всех 28 стимулов, более чем 80% испытуемых решили, что M более подобно гистограмме, в которой высоты цветных полос были упорядочены таким же образом.

2. Упорядочение L и R по признаку вероятности, в качестве выборок, либо в качестве совокупностей, коррелировались с их упорядочением по подобию. Из 17 троек гистограмм, в которых было оценено, что R более подобно M, только двум была присуща малая вероятность того, что они являются выборками, и четырем, что они являются совокупностями. Из этих 11 выборок, в которых было оценено, что L более подобно M, только одна была оценена ниже, чем R, в качестве выборки. Таким образом, упорядочение по подобию не совпадало с упорядочением по вероятности в менее чем 13% сти-

мулов. Совместное распределение суждений подобия с оценками вероятностей дана на Рисунке 2; где в (А) L и R — выборки, а в (В) — совокупности.

Коэффициент корреляции ϕ между оценками вероятности как выборки и ранжированием вероятности как совокупности был 0.75, что примерно столь же высоко, как корреляции ϕ между оценками по вероятности и ранжированием по подобию. Это подтверждает положение о том, что испытуемые в обеих группах, оценивающих вероятность, основывали свои упорядочения в значительной степени на подобию, оценка же была сделана третьей группой.

За счет каких особенностей выборки воспринимаются репрезентативными?

Олсон (Olson, 1976) указал, что, хотя “понятие суждения, основанное на оценке репрезентативности, подтверждается в значительной степени, как экспериментально, так и интроспективно, в большом количестве ситуаций”, этого недостаточно, если мы не можем определить “факторы, которые делают свойства определенной задачи или проблемы существенными на столько, чтобы относительно них оценивать репрезентативность” (с. 608). В некоторых контекстах, таких как исследование подобия и вероятности, они были идентифицированы. В качестве еще одного примера, оказывается, что ключ, определяющий репрезентативность в выборке для неупорядоченных выборок различного размера, взятых из некоторого распределения Бернулли, не соответствует пропорции “успехов” в выборке и соответствующему параметру совокупности (Kaheman и Tversky, 1972b, 3). Исследование, которое я сейчас опишу (Bar-Hillel, 1980b), предлагает общую методологию для идентификации особенностей выборок из любой популяции, определяющих репрезентативность. Это можно пояснить на примере попытки идентифицировать признаки, которые определяют репрезентативность выборок (трех наблюдений) взятых из функции распределения (например, нормального), по форме напоминающей колокол.

Основной принцип прост: если мы предполагаем, что суждения репрезентативности и вероятности определены в соответствии с одними и теми же существенными свойствами, то эти свойства могут быть обнаружены, за счет опроса, в котором испытуемым предлагается определить вероятности для некоторого набора выборок. Эта процедура меняет роли вероятности и репрезентативности. Другими словами, оценки вероятности используются не для того, чтобы подтвердить репрезентативность, а чтобы ее вывести. Однако если попросить испытуемых оценить выборки непосредственно на основе репрезентативности, то это может рассматриваться как спорный вопрос. Они могут возразить: “Что Вы подразумеваете под репрезентативностью?” Вместо этого, предложенная стратегия заменяет собой ясный, однозначный вопрос (“Какая из этих выборок, более вероятно, взята наугад из этой совокупности?”), который, согласно стандартным посылкам, имеет нормативный ответ.

Испытуемым предложили подобную задачу:

Средний рост студентов американских колледжей - 175 см. Из офиса секретаря колледжа были наугад взяты три файла, принадлежащие Джону, Майку и Бобу. Какой результат является более вероятным для роста трех молодых людей?

Джон – 178 см Майк – 170 см Боб – 176 см

или

Джон – 177 см Майк – 177 см Боб – 177 см

Фактические данные отличались в различных документах и их можно найти в колонках стимулов Таблицы 1 (вышеупомянутый пример соответствует строке 1). Систематически изменяя используемые числа, можно было проверить только то, какие особенности выборок, и в каком порядке, определяли ответы испытуемых. Прежде, чем я предложу вашему вниманию обобщенную картину, вот несколько пояснений чтобы, помочь читателю понять таблицу.

1. Выборка обозначалась как А всегда, когда она воспринималась как более вероятная, чем выборка, с которой она шла в паре. В первоначальных формах вопросника, выборки А появлялись справа так же часто, как и слева. Точная пропорция испытуемых, которые выбрали А, а не В, приведена в колонках ответов.

2. Выборку, которая является более вероятной с нормативной точки зрения, можно увидеть в колонке статистических измерений. Выборка А более вероятна, чем выборка В тогда и только тогда, когда $P(A)/P(B) > 1$, то есть, когда

$$\frac{f(X_1) \cdot f(X_2) \cdot f(X_3)}{f(Y_1) \cdot f(Y_2) \cdot f(Y_3)} > 1$$

где $A = [X_1, X_2, X_3]$, $B = [Y_1, Y_2, Y_3]$. Это отношение было вычислено для функции с нормальной плотностью при $\mu=175$ см, $\sigma=6$ см. Так как статистически правильный ответ появляется в столбцах А и В одинаково часто, ответы испытуемых четко связаны с этим не более, чем на случайном уровне.

Возможно наиболее поразительная особенность результатов, сведенных в Таблице 1 – тот факт, что не существует никакого признака, на который полагаются испытуемые, упорядочивая выборки. Например, иногда они отбирают выборку, среднее значение которой ближе к среднему значению совокупности (например, строка 1), а иногда они выбирают совершенно противоположное (например, строка 2). Иногда испытуемые отбирают выборку, содержащую более крайние наблюдения (например, строка 13), а иногда напротив (например, строка 14). Действительно, кажется, что набор признаков является важным и рассматривается некоторым последовательным способом, как изображено на Рисунке 3 (количество идентичных наблюде-

Таблица 1 Оценки и статистические измерения относительного подобия пар выборок (Бар-Хиллел, 1980).

Стимулы для выборки А		Стимулы для выборки В						Ответы			Правильные ответы		Статистика Колмогорова-Смирнова					
		Джон	Майк	Боб	Джон	Майк	Боб	А (%)	В (%)	n					P(B)	P(A)	А	В
Три различия против трех сходств	1	178	170	176	177	177	177	95	5	188	1.39	В	0.31	0.63				
	2	181	165	177	175	175	175	93	7	29	6.42	В	0.30	0.50				
	3	178	170	176	175	175	175	97	3	30	1.59	В	0.31	0.50				
	4	178	180	176	175	175	175	85	15	26	1.59	В	0.57	0.50#				
	5	178	180	176	177	177	177	80	20	25	1.39	В	0.57	0.63				
	6	191	183	189	175	175	175	72	28	25	1300	В	0.91	0.50#				
Три различия против двух сходств	7	178	170	176	176	170	176	72	28	25	1.12	В	0.31	0.43				
	8	178	170	176	178	170	178	76	24	25	0.68	А	0.31	0.36				
	9	176	174	175	178	170	178	57	43	29	0.56	А	0.43	0.36#				
Предпочтителен больший диапазон	10	178	170	176	179	174	175	62	38	29	1.58	В	0.31	0.43				
	11	178	170	176	177	173	176	71	29	31	1.43	В	0.31	0.37				
	12	178	170	175	178	172	177	54	46	28	1.14	В	0.31	0.31				
	13	178	170	176	177	171	175	54	46	28	1.14	В	0.31	0.37				
Предпочтителен меньший диапазон	14	178	170	176	179	169	177	54	46	26	0.75	А	0.31	0.30#				
	15	178	170	176	181	165	177	77	23	30	0.23	А	0.31	0.30#				
	16	178	170	176	180	170	175	64	36	25	0.81	А	0.31	0.20#				
	17	178	170	176	181	169	175	59	41	27	0.60	А	0.31	0.17#				

Двусторонние выборки против односторонних	18	178	170	176	172	170	174	78	22	27	1.00	и тот, и другой	0.31	0.57
	19	178	170	176	178	180	176	64	36	25	1.00	и тот, и другой	0.31	0.57
	20	178	170	176	184	176	182	76	24	25	0.26	A	0.31	0.57
	21	181	165	177	178	180	176	63	37	32	4.29	B	0.57	0.57
	22	176	174	175	184	176	182	73	27	30	0.87	A	0.37	0.57
	23	177	171	175	178	180	176	93	7	28	0.17	A	0.43	0.57
	24	176	174	175	178	180	176	79	21	37	1.58	B	0.43	0.57
Две односторонних выборки	25	183	175	177	184	176	182	87	13	30	0.42	A	0.50	0.57
	26	185	175	177	184	176	182	85	15	27	0.68	A	0.50	0.57
	27	178	180	176	184	176	182	91	9	34	0.26	A	0.57	0.57
	28	178	180	176	181	179	183	81	19	26	0.32	A	0.57	0.75

Источник: Бар-Хиллер (1980b)

ний о том, представлены ли обе стороны среднего значения выборки, среднее значение и диапазон). Это усиливает мнение о том, что в независимости от определения, предполагаемое подобие выборки, составляет собственную концепцию. Если бы испытуемые использовали единственный признак, не было бы необходимости называть это *репрезентативностью*. Известно, что оценки испытуемых, кажется, основаны на сложных соображениях о структуре выборок, представляющих основания для выражения (скорее, концептуализации) их эвристического *суждения на основе репрезентативности*.

Может ли блок-схема на Рисунке 3 рассматриваться больше, чем просто удобное и емкое резюме результатов Таблицы 1? Насколько серьезно можно считать ее процессной моделью фактического рассуждения испытуемых в описанной задаче? С одной стороны, очевидно, что не все особенности модели будут соответствовать – другим выборкам, другим размерам выборок, другим распределениям. Я бы, например, не утверждала, что любая выборка с дисперсией будет всегда оценена как более вероятная, чем любая выборка без дисперсии – хотя это и сохраняется для выборок, использованных в моем эксперименте. Упорядоченная модель, таким образом, является только приближением фактической когнитивной стратегии людей. Однако, будучи таковой, она обладает весомым психологическим преимуществом, являясь правдоподобной, так как она содержит в себе лишь маленькое число простых в вычислительном отношении признаков.

В этом пункте предлагается интересное наблюдение, которое связывает описанное выше исследование и настоящее. В исследовании гистограмм, два вида упорядочения по вероятности сравнивались с упорядочением по подобию: (а) упорядочение двух выборок по признаку более или менее вероятных исходов некоторой совокупности, обозначенной $P(s/p)$; (б) упорядочение двух совокупностей в качестве более или менее вероятных источников некоторой выборки, обозначенной $P(p/s)$ (иногда называемое обратными вероятностями). Эти две задачи, привели к очень близким результатам, согласно которым они обе были сделаны очень похожим образом – действительно, в соответствии с гипотезой, что они были обе решены на основе репрезентативности (то есть, в данном случае, визуального подобия). Существует статистическая концепция, которая является аналогичной психологической концепции репрезентативности – концепции степени соответствия. Измерение степени соответствия типично используется в проверке гипотезы, когда неизвестна совокупность, породившая некоторую известную выборку, при отсутствии информации о предполагаемой вероятности. Таким образом, при оценке совокупностей, репрезентативность, кажется, более соответствует концепции степени соответствия между выборкой и совокупностью, чем концепции вероятности совокупности, условной для некоторой выборки, $P(p/s)$.

В исследовании распределения роста, однако, была предложена совокупность, которая являлась источником для выборок. Поэтому, испытуемым не приходилось оценивать ни обратные вероятности, ни степень соответ-

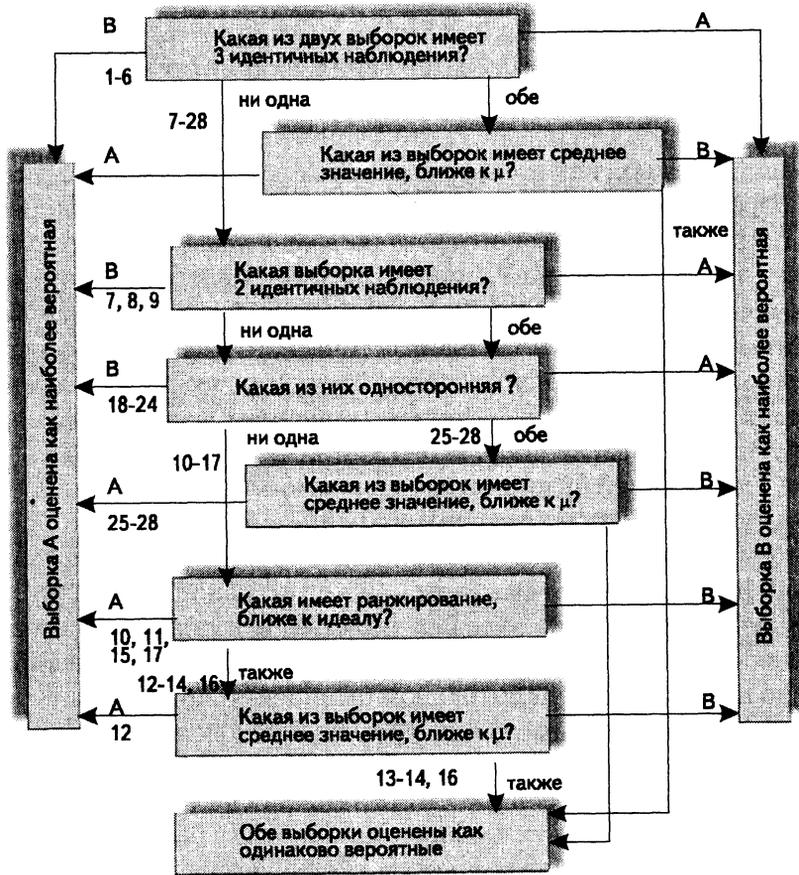


Рис. 3. Блок-схема, суммирующая признаки и последовательность, в которой они использовались для упорядочения выборок по вероятности.

(Источник: Бар-Хиллел, 1980b.)

ствия. Однако, интересно, смогут ли их оценки, так или иначе, быть пере-строены в качестве интуитивной попытки оценить последнее.

Одно из возможных измерений степени соответствия между выборками из трех элементов и нормальным распределением - одно-выборочная, дву-сторонняя статистика Колмогорова-Смирнова (КС) (см., например, Siegel, 1956, стр. 47-52). Эта статистика интерпретируется так, что чем больше значение (максимальное расстояние между функциями распределения выборки и совокупности), тем менее вероятно, что эта выборка взята наугад из совокупности. В последних двух столбцах Таблицы 1 приведена статистика КС для каждой из исследуемых выборок. В 18 парах, распределение выбо-рок по статистике КС согласовывалось с их распределением большинством

испытуемых; в 3 парах статистика КС была выровнена (в 1 из которых, строка 12, испытуемые также были фактически единодушны), а в 7 парах (отмеченных в Таблице 1 звездочками) распределения различны.

Это мало влияет на сравнение степени согласованности между распределением КС и распределением, произведенным испытуемыми, со степенью согласованности между нормативным распределением и распределением испытуемых (в последнем пришли к согласию в 13 из 28 стимулов), так как последние цифры – только побочные продукты отдельно взятых стимулов, которые не были наугад *отобраны* из всех возможных стимулов. Намного важнее рассмотреть, где модель иерархических особенностей отличается от модели “степени соответствия”. Случаи, где распределение КС было согласовано, а распределение испытуемых не было (строки 21, 27, и в небольшой степени 12) могут быть приписаны значениям КС, вычисленным со стандартным отклонением в 6 см. Это привело к появлению выборки (181, 169, 175) (выборка В, строка 17), которая имеет диапазон 12 см, при самом маленьком значении КС, когда “идеальный” (то есть, наиболее репрезентативный) диапазон в оценках испытуемых был несколько меньшим – от 8 до 10 см. Так как мы никак не можем, на основе имеющихся данных, установить стандартное отклонение субъективного распределения роста, становится непонятно, как поступить с этим несоответствием.

Другой вид несоответствия, к которому относятся строки 4, 6, 9, и с 14-17, происходит, потому что модель КС, а не иерархическая модель, предусматривает компенсацию. Другими словами, если некоторая выборка имеет “недостаток” (как, например, она может состоять из трех идентичных наблюдений, или представлять только половину функции совокупности), испытуемые считают, что она менее репрезентативна, чем любая выборка без недостатка, но модель КС рассматривает также факторы, которые находятся ниже в иерархической модели. Как я упомянула выше, иерархический характер предложенной модели процесса должен быть расценен как приближение, поэтому рассматриваемое несоответствие является скорее несоответствием степени, чем вида.

Другие измерения степени соответствия вычислены статистиками. То, что они не пришли к согласию, указывает, что свойство, которое они намереваются измерять, не ясно и не точно определено. Хотя у опытного статистика интуиция более развита, чем у неподготовленного испытуемого, они оба могут пытаться искать одно и то же базовое понятие. До настоящего времени, даже усилия статистиков в этом направлении не увенчались большим успехом.

Роль размера в оценке выборки

Некоторые характеристики выборок соответствуют некоторым параметрам совокупности; эти характеристики называются статистикой выборки. К ним обычно применялась эвристика репрезентативности, когда репрезентатив-

ность эквивалентна *подобию* или *приближенности* статистики выборки к соответствующим параметрам популяции. Я назову это первичным смыслом репрезентативности. Но выборки могут быть описаны в терминах процедуры, посредством которой они также были получены. Например, выборка может быть осуществлена с возвращением или без; из известной совокупности или неизвестной; и т.д. Возможно, к ним также могло применяться понятие репрезентативности. Таким образом, люди могут (и имеется курьезное свидетельство, что это так) оценивать выборку, взятую наугад, как менее репрезентативную, чем стратифицированную выборку, или же большую выборку, как более репрезентативную, чем маленькую (Bar-Hillel, 1980b). Я назову это вторичным смыслом репрезентативности. Так как это касается методов осуществления выборки, а не последующей статистики выборок, репрезентативность во вторичном смысле отражает ожидания относительно репрезентативности в первичном смысле. Другими словами, чтобы оценить большую выборку как более репрезентативную, чем маленькую, стоит ожидать, что ее “существенные особенности” или “существенные свойства” лучше отразят “существенные особенности” или “существенные свойства” совокупности, до того, как будет сказано, каковы они на самом деле.

Экспериментальные проблемы, которые я опишу в последней главе, были посвящены рассмотрению вторичного смысла репрезентативности, по отношению к размеру выборки.

Проблема 1: Две фирмы, специализирующиеся в изучении общественного мнения, проводят наблюдение с целью определения пропорции избирателей, намеревающихся проголосовать положительно на некоем референдуме. Фирма А рассматривает выборку из 400 человек. Фирма В рассматривает выборку из 1.000 человек. В чьей оценке Вы были бы более уверены?

Фирмы А _____ Фирмы В _____ Приблизительно поровну _____

Эта задача и два ее варианта (в которых, соответственно, была добавлена информация, что полный размер рассмотренного общества был 8.000 и 50.000 соответственно), была предоставлена 72 испытуемым. Более чем 80% из них были больше уверены в достоверности большей выборки, по сравнению только с 4%, отдавшим предпочтение меньшей выборке.

Если просто сделать вывод о том, что чем больше выборка, тем более репрезентативной она кажется (то есть, ожидается, что она будет более точной или надежной), то эти результаты были бы слишком тривиальны, чтобы о них говорить. Вовсе не размер выборки сам по себе определяет ее репрезентативность, а скорее что-то более близкое к отношению между размером выборки и размером совокупности. Если рассматриваемые выборки, взяты из одной и той же совокупности, линейно связаны, то они будут одинаково распределены. Если размер совокупности изменяется также, как и размер выборки, различие становится очевидным. Рассмотрим следующую задачу:

Задача 2: Две фирмы, специализирующиеся в изучении общественного мнения, проводят наблюдение с целью определения пропорции избирателей, намеревающихся проголосовать положительно на некоем референдуме каждая в своем городе.

Фирма А работает в городе, где 1 миллион избирателей.

Фирма В работает в городе, где 50. 000 избирателей.

Обе фирмы формируют выборку, выбирая одного человека из каждых 1. 000 избирателей. В чьей оценке Вы были бы более уверены? _____

Здесь, хотя фирма А имеет выборку из 1. 000 человек, а Фирма В только из 50, процент испытуемых, которые выразили больше уверенности в достоверности большей выборки, понизился до 50% , тогда как 29% указали равную уверенность относительно обеих выборок. В другой группе испытуемых, которым сказали, что обе фирмы сделали выборку 1. 000 человек (нежели чем 1 из каждой 1. 000), 9% проявили одинаковую уверенность в достоверности обеих выборок, по сравнению с 62% , которые были более уверены в выборке людей, голосовавших в меньшем городе. Пояснительные тексты, составленные испытуемым, подтвердили то, что они рассматривали *пропорциональные* размеры выборок, как и абсолютные размеры.

Иногда отношение “выборка-совокупность” является действительно важным для оценивания выборки – например, когда формирование выборки сделано без возвращения и особенно, когда выборка представляет собой достаточно большую часть совокупности. Однако если формирование выборки сделано с возвращением (то есть, совокупность бесконечно велика), или если в выборку берется только малая часть большой совокупности (которая отражает совокупность, для всех прагматических целей), тогда рассмотрение относительного размера теряет смысл наряду с рассмотрением абсолютного размера.

Задача 3: Даны два закрытых сосуда. В обоих – смесь красных и зеленых бусинок. Количество бусинок различно в двух сосудах – в маленьком 10 бусинок, а в большом – 100 бусинок. Процент красных и зеленых бусинок один и тот же в обоих сосудах. Выборка осуществляется следующим образом: Вы вслепую достаете бусинку из сосуда, запоминаете ее цвет и возвращаете на место. Вы перемешиваете бусинки, снова достаете вслепую, и снова запоминаете цвет. В общем, Вы тянете бусинку из маленького сосуда 9 раз, а из большого – 15 раз. Как Вы думаете, в каком случае у Вас больше возможностей угадать преобладающий цвет?

Так как процедура формирования выборки, описанная в задаче 3 – с возвращением, то число бусинок в этих двух сосудах с нормативной точки зрения абсолютно не важно. Испытуемые в своих выборах должны были однозначно обратить внимание на большую выборку в 15 бусинок. Вместо этого, 72 из 110 испытуемых выбрали меньшую выборку из 9 бусинок. Это можно объяснить только тем, что отношение размера выборки к размеру совокупности - 90% в последнем случае и только 15% в первом. Другие результаты, к которым пришла Бар-Хиллел (Bar-Hillel, 1979), подтверждают этот же вывод.

Таким образом, согласно статистике выборок, если кандидаты на роль “основных свойств” неизвестны, эту роль может принять на себя размер

выборки. Люди понимают, что размер – важное свойство выборок. Почему же они не в состоянии перенести это понимание на первичный смысл репрезентативности? Другими словами, если люди полагают, что большие выборки будут более репрезентативны по отношению к совокупности, из которой они получены, чем меньшие, почему они не в состоянии оценить некоторые ненормативные результаты (например, пропорция выборки 60%, взятой из популяции, характеризующейся 50% пропорциональностью) как менее вероятные в больших, а не в меньших выборках?

Возможно оценка того, что некоторый результат формирования выборки является менее вероятным в больших выборках, зависит от восприятия его как нерепрезентативного. С нормативной точки зрения, только та выборка является “точной”, среднее значение которой идентично среднему значению совокупности. Все другие неточны, хотя несколько точнее, чем другие. С психологической точки зрения, результаты формирования выборок могут быть оценены как репрезентативные, даже если они отклоняются по некоторым параметрам от параметров совокупности. Возможно, значение 60% не воспринимается как нерепрезентативное и поэтому не вызывает у испытуемых вторичного чувства репрезентативности.

Чтобы проверить эту возможность, я дала испытуемым несколько вариантов проблемы рождаемости согласно Канеману и Тверскому (1972b, 3)

Некоторый город обслуживается двумя больницами. В большей больнице каждый день рождается приблизительно 45 младенцев, а в меньшей больнице, приблизительно 15 младенцев. Как Вы знаете, приблизительно 50% всех младенцев – мальчики. Точный процент младенцев-мальчиков, однако, изменяется каждый день. Иногда он может быть выше, чем 50%, иногда ниже.

В течение 1 года, каждая больница делала учет дней, на которые приходится (больше / меньше) 60% рожденных младенцев-мальчиков. Какая больница, как Вы думаете, сделала запись большего количества таких дней? (1972b, с. 443)

В отдельных задачах, выполняемых отдельными группами, пропорция мальчиков была изменена с 60% до 70%, 80% и 100%. В таблице 2 показано влияние этого изменения на распределение ответов испытуемых.

Вторая колонка точно повторяет, и в задаче, и в результатах, версию Канемана и Тверского. Но заметьте, что, как только пропорция мальчиков превышает 70%, модальным ответом становится “меньшая больница”, что является правильным ответом. А когда сказано, что все младенцы – мальчики, “меньшая больница” является даже ответом у большинства. Согласно другому набору задач, пропорция была меньше чем 60%, 70% или 80%. В этом случае везде был модальный ответ “одинаково”.

Эти результаты говорят о том, что, как только результат формирования выборок воспринят как нерепрезентативный, размер выборки будет играть надлежащую роль в оценке выборки людьми, возможно, потому что меньший размер более совместим с нерепрезентативными результатами, чем больший. Однако если результат формирования нельзя назвать нерепрезентативным, люди не принимают во внимание размер выборки. Размеры, боль-

ше чем 70%, кажется, закодированы как нерепрезентативные. Размеры 50% и включительно – нет.

Таблица 2. Пропорция тестируемых, выполнивших задачу о рождаемости (различные варианты).

	<i>более 60%^a</i>	<i>более 60%^b</i>	<i>более 70%^b</i>	<i>более 80%^b</i>	<i>100%^c</i>
<i>большая</i>	24%	20%	25%	26%	19%
<i>меньшая</i>	20%	20%	43%	42%	54%
<i>одинаково</i>	56%	60%	60%	32%	27%
<i>N</i>	50	40	28	27	41

	<i>менее 60%^a</i>	<i>менее 70%^b</i>	<i>менее 80%^b</i>
<i>большая</i>	20%	31%	29%
<i>меньшая</i>	24%	28%	25%
<i>одинаково</i>	56%	41%	46%
<i>N</i>	45	29	28

^aКанеман и Тверски

^b ранее неопубликованные данные, Бар-Хиллер

^c из Бар-Хиллер (1979). Здесь, большая и маленькая больницы по 15 и 5 младенцев, соответственно

Обсуждение

В этой главе представлены выводы о трех исследованиях оценок субъективной вероятности, которые рассматривают понятие репрезентативности. Однако каждое из них касается его в разной мере.

Первое исследование – это прямая демонстрация того, что оценки подобию и вероятности связаны. В отличие от многих других задач, которые проверяли эту гипотезу (см., напр., Kaneman & Tversky, 1972b, 3; 1973, 4), в данном случае испытуемых нельзя обвинить ни в неспособности учитывать другие соображения (или, хотя бы очевидные или простые), ни в том, что они придают слишком большое значение ошибочным или нерелевантным соображениям. Кстати, исключая подсчет вероятности, которую их просили оценить, испытуемым ничего не оставалось, как использовать ту стратегию, которой они пользовались. Это иногда уменьшает показательную силу. Если учитывать и другие результаты, можно увидеть, насколько непреодолимыми могут быть соображения подобию – даже простого визуального типа – при оценивании сложных событий, путем подсчетов или концептуально.

До исследования характерных свойств выборок, тесты репрезентативности изначально рассматривали совокупности с одним параметром или признаком с одним измерением. Это позволило сформировать непосредственное и априорное оценивание стимулов по подобию, не вызывая необходимости в независимых оценках, полученных экспериментальным путем. (Как бы там ни было, некоторые “очевидные” признаки позже были подвергнуты сомнению (см., напр., Olson, 1976)). Но когда стимулы усложняются, отношения подобия между ними становятся менее очевидными, так как многие признаки влияют на общую оценку. В этом исследовании использовалась установленная к тому времени корреляция между вероятностью и репрезентативностью, чтобы показать, как последняя определяется в выборках, характеризующихся более чем одним признаком.

Часто лучшим способом обнаружить, насколько подобны два стимула, является прямо спросить о подобии. Это немного проблематично, если необходимо узнать, насколько выборка S репрезентативна совокупности P . Однако полезно уметь оценить это косвенно, например, спросив, насколько вероятно то, что выборка S взята из совокупности P . Второе исследование в этой главе показывает жизнеспособность этого подхода, давая описание того, какие свойства выборок делают их репрезентативными в некотором контексте.

Наконец, в третьем исследовании подчеркивается важное различие между репрезентативностью выборки и репрезентативностью результата формирования выборки. В первичном смысле, репрезентативность – это оценка, которая приписывается результатам формирования выборки. Другими словами, чтобы оценить, насколько выборка репрезентативна популяции в этом смысле, должны быть известны характеристики обеих. Таким образом, имеет смысл оценивать репрезентативность выборки из 15 новорожденных, 9 из которых мальчики, имея перед глазами общее распределение пола среди новорожденных (известную пропорцию мальчиков). Нет смысла оценивать репрезентативность этой выборки, если пропорция мальчиков ни в выборке, ни в популяции не известна. Но существует и другой смысл репрезентативности, который сделает значимой также и вторую оценку. В этом смысле, выборки более репрезентативны, если их оценивают как более вероятные в первом смысле. Другими словами, независимо от средней величины совокупности и до того как становится известна средняя величина выборки, выборка из 45 новорожденных более репрезентативна (во втором смысле), чем выборка из 15, потому что она с большей вероятностью будет напоминать популяцию, если станет известна в ней пропорция мальчиков. Следовательно, в выборке из 45 может оказаться 65% мальчиков, а в выборке из 15 – 53%. Но если четко различать эти два смысла, путаницы не возникнет.

В третьем исследовании показано, что люди реагируют на размер выборки, когда делают вторичные оценки репрезентативности, даже если они часто не видят подтекст, который обычно бывает у первичных и стандартных оценок репрезентативности.

6. Оценки репрезентативности и на основе репрезентативности*

Амос Тверски и Даниель Канеман

Несколько лет назад, мы представили анализ принятия решений в условиях неопределенности, который связал субъективные вероятности и интуитивные прогнозирования по поводу ожиданий и впечатлений о репрезентативности. В эту концепцию включили две различные гипотезы: (i) люди ожидают, что выборки будут подобными своей родительской совокупности, а также отобразят случайность процесса формирования выборки (Tversky & Kahneman, 1971, 2; 1974,1); (ii) люди часто полагаются на репрезентативность как на эвристику для суждения и прогнозирования (Tversky & Kahneman, 1972b, 3; 1973, 4).

Первая гипотеза была выдвинута для объяснения общего мнения, что случайные процессы самокорректируются, то есть наблюдается преувеличенная вера в стабильность исходов, наблюдаемых в малых выборках, проигрышей в казино и связанные с этим предубеждения относительно оценок случайности. Мы предположили, что околонуточная концепция случая стимулирует веру в закон малых чисел, согласно которому даже небольшие выборки являются высоко репрезентативными по отношению к их родительским совокупностям (Tversky & Kahneman, 1971, 2). Подобная гипотеза могла также объяснить общую тенденцию к преувеличению непротиворечивости и прогнозированной ценности черт характера личности (Mischel, 1979) и слишком высокому оцениванию корреляции между подобными переменными (см. Главу 15) и типами поведения (Shweder и D'Andrad, 1980). Люди, кажется, верят в голограммоподобную модель личности, в которой любой акт поведения представляет истинный характер человека (Kahneman, & Tversky 1973, 4).

Гипотеза, согласно которой люди ожидают, что выборки будут высоко репрезентативны по отношению к их родительской совокупности, концеп-

* Эта работа была поддержана Службой Военно-морских Исследований согласно Контракту N00014-79-C-0077 с Университетом Стэнфорда.

туально независима от другой гипотезы, согласно которой люди часто используют эвристику репрезентативности, чтобы предсказывать и оценивать вероятности. То есть они часто оценивают вероятность неопределенного события или выборки «степенью, в которой они (i) подобны в существенных свойствах своей родительской совокупности и (ii) отражают существенные особенности процесса, который послужил основанием для их возникновения» (Kahneman & Tversky, 1972b, с. 431, 3). Эта гипотеза изучалась в нескольких контекстах, включая интуитивные статистические оценки и прогнозирование выбора профессии (Kahneman & Tversky, 1972b, 3; 1973, 4).

Обе гипотезы репрезентативности использовались для объяснения разнообразия наблюдений, таких как относительная неэффективность информации консенсуса и использования подобия в интерпретации проективных тестов (Nisbett и Ross, 1980). Эти гипотезы также определили направление поиска существенных нарушений нормативных правил в интуитивных суждениях, что дало хорошие результаты. Большая часть этого исследования касалась оценок, сделанных *на основе* репрезентативности, то есть роли репрезентативности в прогнозировании и при составлении заключений. Относительно мало работ было посвящено оценкам *репрезентативности*, то есть характеру этого отношения и его показателей, вне контекста случайного осуществления выборки (Bar-Hillel, 1980b). Первая глава этой части книги описывает природу репрезентативности, а также условия, в которых концепция репрезентативности позволяет объяснить интуитивные прогнозы и оценки вероятности. Во второй главе мы проиллюстрируем различие между логикой репрезентативности и логикой вероятности в оценке вероятности сложных событий.

Отношение репрезентативности

Репрезентативность – это соотношение между процессом или моделью M и некоторым случаем или событием X , связанным с этой моделью. Репрезентативность, как и подобие, может быть определена опытным путем, например, если попросить людей оценить, какое из двух событий, X_1 или X_2 , является более репрезентативным по отношению к некоторой модели M , или является ли событие X более репрезентативным по отношению к M_1 или M_2 . Рассматриваемая модель могла относиться к человеку, монете или мировой экономике, а соответствующие исходы могли быть описанием, последовательностью выпадения орлов и решек или ценой на золото.

Репрезентативность – направленное соотношение: мы говорим, что выборка более или менее репрезентативна по отношению к отдельно взятой совокупности, и что действие является репрезентативным по отношению к человеку. Мы обычно не говорим, что совокупность репрезентативна по отношению к выборке, или что человек репрезентативен по отношению к действию. В некоторых задачах, однако, можно полностью изменить роли модели и исхода. Например, можно оценивать, репрезентативен ли человек

стереотипу библиотекаря или репрезентативно ли занятие библиотекаря этому человеку.

Мы различаем четыре основных случая, для решения которых обычно используется концепция репрезентативности.

1. M — класс, а X — значение переменной, определенной в этом классе. Именно в этом смысле мы говорим о (более или менее) репрезентативных размерах дохода профессора колледжа или возрасте вступления в брак в какой-либо культуре. Естественно, наиболее репрезентативная величина будет близка к средней, медиане или способу распределения соответствующей переменной в классе M . Отношение репрезентативности, главным образом, определено в этом случае так, что испытуемый знает распределение частоты соответствующей переменной.

2. M — класс, а X — частный случай этого класса. Большинство читателей, вероятно, согласится, что Джон Апдайк — более репрезентативный американский писатель, чем Норман Майлер (Norman Mailer). Конечно, такое суждение не основано на частоте; оно отражает степень, в которой стили, темы и идеи этих авторов являются центральными в современной американской литературе. Подобные соображения определяют репрезентативность случаев, которые сами скорее являются классами, чем частными случаями. Например, дрозд считается более типичной птицей, чем курица, хотя он встречается реже (Rosh, 1978; Smith, Shoben и Rips, 1974). Таким образом, событие репрезентативно по отношению к категории, если оно имеет существенные особенности, которыми обладают элементы этой категории, и если у него отсутствуют отличительные черты, которых нет у членов этой категории (Rosh, 1975; Tversky, 1977).

Современные разработки в области формирования концепции (Rosh & Mervis, 1975; Mervis & Rosh, 1981), семантической памяти (Bransford & Franks, 1971) и распознавания образов (Posner & Keele, 1968) показали, что наиболее репрезентативные, или прототипные, элементы категории лучше запоминаются, легче вспоминаются и распознаются, чем элементы, которые являются более частыми, но менее репрезентативными. Кроме того, люди часто допускают ошибку, “распознавая” прототипный стимул, который никогда не демонстрировался. Поэтому репрезентативность может смещать память распознавания, а также оценки частоты.

Надо заметить, что существуют два способа, с помощью которых элемент может быть высоко репрезентативен по отношению к классу. Два смысла репрезентативности близко соответствуют отношениям типичности и прототипичности. Элемент высоко репрезентативен по отношению к категории, если он типичен или модален; он может также быть репрезентативен, если он является идеальным типом, который воплощает сущность категории. Например, Нью-Йорк является прототипом американского города, в то время как Цинциннати можно выбрать как типичный город. Точно так же наши понятия о прототипе и типичной француженке различаются. Первая — веро-

ятно молодая, изящная парижанка, в то время как вторая будет полной женщиной средних лет из провинции.

3. M — класс, а X — подкласс M . Большинство людей, вероятно, согласится, что население Флориды менее репрезентативно по отношению к американскому населению, чем население штата Иллинойс, и что студенты, изучающие астрономию, являются менее репрезентативными по отношению ко всему студенчеству, чем студенты, изучающие психологию. Критерии репрезентативности различны для подкласса и для отдельного события, потому что событие может давать представление о главной тенденции свойств, в то время как, подкласс может также представлять диапазон и изменчивость. Человек, чей рост, вес, возраст и доход соответствует средним величинам для американского населения, репрезентативен по отношению к этой совокупности. Группа из 100 человек с одними и теми же характеристиками не смогла бы представлять изменчивость признаков.

Если класс M состоит из таких отдельных групп, что изменчивость в пределах каждой группы является очень маленькой относительно изменчивости между группами, мы рассматриваем каждую группу скорее как представителя категории, чем как подкласс. Таким образом, естественно рассматривать “дрозда” как вид птиц или как частный случай категории “птица”, хотя множество “дрозды” является подклассом класса птиц. В общем смысле (2) может рассматриваться как частный случай (3), где подкласс X состоит из одного члена. Точно так же (1) может расцениваться как вариант (2) с одним измерением. Три типа репрезентативности различаются по сложности X , где (1) — одноэлементное событие с единственным признаком, (2) — одноэлементное событие с множеством признаков и (3) — многоэлементное событие с одним или более признаком.

Особенно важный пример репрезентативности подкласса имеет место, когда X является случайной выборкой из указанной совокупности. Случайная выборка должна представлять случайность процесса выбора, а не только существенные особенности совокупности, из которой она взята. Когда 100 человек отобраны наугад, например, выборка из 53 мужчин и 47 женщин может показаться более репрезентативной, чем выборка из 50 мужчин и 50 женщин, потому что первая представляет неравномерность формирования выборки, в то время как последняя — нет (Tversky & Kahneman, 1972b, 3). Статистическая концепция репрезентативной выборки исследована Крускалом (Kruskal) и Мостеллером (Mosteller) (1979a, 1979b).

4. M — (причинная) система, а X — (возможное) последствие. Этот случай отличается от предшествующих, тем, что M больше не является классом объектов или случаев, а скорее системой, которая порождает различные эффекты. Например, M может представлять экономику США, а X — степень инфляции, или M может представлять человека, а X действие, выполненное M , например, развод, самоубийство, выбор профессии. Здесь, X репрезентативно по отношению к M , или потому что оно часто связано с M (например, высокая температура обычно сопровождается пневмонией) или

потому что люди верят, правильно или неправильно, что M — это причина X (например, высшая мера наказания предотвращает похищения людей). Тверски и Канеман (1980, 8) проиллюстрировали и исследовали включение причинных схем в оценки условных вероятностей.

Подводя итоги, отношение репрезентативности может быть определено для (1) величины и распределения, (2) события и категории, (3) выборки и совокупности (4) причины и следствия. Во всех четырех случаях, репрезентативность выражает степень соответствия между X и M , но ее показатели не одинаковы в этих четырех случаях. В случае (1) на репрезентативность влияет кажущаяся относительная частота или статистическая связь. В случаях (2) и (3) репрезентативность определена прежде всего подобием, например, одного случая другим, или статистик выборок соответствующим параметрам совокупности. Наконец, в случае (4) репрезентативность определяется в значительной степени причинными убеждениями (валидными или нет).

Репрезентативность и вероятность

Использование репрезентативности для объяснения оценок вероятности и интуитивных прогнозов опирается на предположения о том, что:

1. Отношение “ X — (очень, ..., несколько не) репрезентативно M ” может осмысленно оцениваться испытуемыми.
2. Эти оценки не должны основываться на впечатлениях от вероятности или частоты, которые должны быть объяснены репрезентативностью.
3. Отношение репрезентативности имеет собственную логику, которая систематически отклоняется от логики вероятности.

Когда эти допущения принимаются во внимание, становится интересно проверить, опосредованы ли оценки вероятности оценками репрезентативности.

Оценка вероятности неопределенного события или прогноза неизвестного числа — сложный процесс, который включает интерпретацию проблемы, поиск необходимой информации и выбор соответствующего ответа. Это можно сравнить с действием гибкой компьютерной программы, которая включает множество потенциально полезных подпрограмм. В терминах этой аналогии, эвристика репрезентативности — одна из процедур, которые могут использоваться, чтобы восстановить, интерпретировать и оценить информацию. Использование этой эвристики, конечно, не препятствует использованию других процедур, так же как использование образов как эвристики для припоминания не препятствует использованию других стратегий. Однако уверенность в эвристике приводит к характерным предубеждениям. Например, когда используются образы, чтобы вспомнить людей, присутствовавших на определенной встрече, ожидается, что легче будет вспом-

нить тех участников, которые находились в непосредственной видимости, чем тех, кого не было хорошо видно. Точно так же использование репрезентативности для оценки субъективной вероятности приводит к переоценке одних вероятностей и недооценке других.

Ранние исследования дали основание для крайней гипотезы, согласно которой некоторые оценки вероятности основаны исключительно на репрезентативности. Например, наблюдение, что субъективные выборочные распределения по существу не зависят от размера выборки (Tversky & Kahneman, 1972b, 3), привело к предположению, что люди оценивают вероятность выборки по подобию ее статистик соответствующим параметрам совокупности. Большинство доступных данных, однако, подтверждают более умеренную гипотезу, согласно которой интуитивные прогнозы и оценки вероятности подвержены влиянию репрезентативности, хотя полностью от нее не зависят. Таким образом, на субъективные вероятности оказывают существенное влияние (нормативно) незначимые факторы, которые затрагивают репрезентативность и они относительно независимы от (нормативно) значимых переменных, которые не влияют на репрезентативность. Важность предубеждений репрезентативности и влияние таких переменных как размер выборки, достоверность, и базовое значение зависит от характера задачи, плана эксперимента, опыта респондентов и наличия подсказок или других характеристик. Роль этих факторов в исследовании суждения обсуждается далее.

Если уверенность в репрезентативности приводит к систематическим ошибкам, почему люди используют его как основание для прогнозов и оценок? Ответ на этот вопрос можно разделить на три части. Во-первых, репрезентативность кажется легко доступной и легкой для оценки. Современное исследование категоризации (Mervis & Rosh, 1981; Rosh, 1978) полагает, что понятия часто организуются и обрабатываются в терминах прототипов или репрезентативных примеров. Следовательно, нам легче оценить репрезентативность события по отношению к классу, чем оценить его условную вероятность. Во-вторых, вероятные события обычно более репрезентативны, чем менее вероятные. Например, выборка, подобная совокупности, более вероятна, чем нетипичная выборка того же размера. В-третьих, мнение, что выборки вообще репрезентативны по отношению к их родительским совокупностям, приводит к тому, что люди переоценивают корреляцию между частотой и репрезентативностью или между статистической связью и коннотативным подобием. Таким образом, репрезентативность используется, потому что (i) она доступна, (ii) она часто связана с вероятностью и (iii) люди переоценивают эту связь. Уверенность в репрезентативности, однако, ведет к предсказуемым ошибкам суждения, потому что репрезентативность имеет собственную логику, которая отличается от логики вероятности.

Различие между репрезентативностью и вероятностью в большей степени очевидно (i), когда представлены недостоверные данные или (ii), когда контрольное событие специфично. В случае (i), исход, который является высоко репрезентативным по отношению к нашей модели, может быть ма-

ловоятным, если наша мысленная модель основана на данных ограниченной валидности. Рассмотрим, например, вероятность того, что кандидат, который произвел превосходное впечатление во время собеседования, преуспеет в очень трудной задаче. Поскольку известно, что впечатления от собеседования часто ошибочны, а успех или неудача на работе определяется многочисленными факторами, которые не возможно предугадать в краткой беседе, успех может быть маловероятен даже, когда он высоко репрезентативен по отношению к нашему впечатлению о кандидате.

В случае (ii) репрезентативный результат может быть маловероятным, потому что он крайне специфичен или детален. Вообще, событие может быть маловероятным либо потому, что оно нетипично, либо потому, что оно очень детализировано. Вес менее 135 фунтов является нетипичным для человека средних лет; а вес 157,625 фунтов является типичным, но слишком точно определенным. Действительно, последнее значение более репрезентативно для человека средних лет, хотя первое намного более вероятно. Из этого примера следует, что увеличение определенности вообще не ведет к уменьшению репрезентативности. Следовательно, сравнение событий, которые отличаются по степени точности, часто создает конфликт между репрезентативностью и вероятностью. Например, случайная выборка из четырех карт, состоящих из короля червей, туза пик, девятки бубен и четверки треф, кажется более репрезентативной, чем выборка, состоящая из четырех карт одной масти, хотя последняя гораздо более вероятна. Таким образом, предубеждения репрезентативности в суждениях вероятности должны быть наиболее очевидными в оценке событий, которые являются репрезентативными, но крайне специфическими. Такие предубеждения представлены в исследованиях оценок вероятности сложных событий, описанных в следующей главе.

Об оценке сложных событий

Существенное различие между вероятностью и репрезентативностью возникает при оценке сложных событий. Предположим, что нам дали некоторую информацию о человеке (например, краткое описание личности) и что мы размышляем о различных признаках или комбинациях признаков, которыми этот человек может обладать: род занятий, склонности или политические симпатии. Один из основных законов вероятности гласит, что детальность может только понизить вероятность. Таким образом, вероятность того, что данный человек является и республиканцем и художником одновременно, должна быть меньше, чем вероятность того, что человек является художником. Это условие основано не только на стандартном исчислении вероятности, но также и на ненормативных моделях (например, Shafer, 1976; Zadeh, 1978).

Однако, требование, что $P(A \text{ и } B) \leq P(B)$, которое можно называть правилом конъюнкции, не относится к подобию или репрезентативности. Синий

квадрат, например, может быть более подобен синему кругу, чем просто кругу, и человек может походить на наш образ республиканца и художника больше, чем на наш образ республиканца. Поскольку подобие объекта цели может быть увеличено, если добавить к цели особенности, которыми обладает и объект (см. Tversky, 1977), подобие или репрезентативность могут быть увеличены путем спецификации цели. Если оценки вероятности опосредованы репрезентативностью или подобием, то возможно строить задачи, где соединение исходов окажется более репрезентативным и, следовательно, более вероятным, чем один из его компонентов.

Эффект конъюнкции: исследование 1

Этот способ прогнозирования был впервые исследован в эксперименте, проведенном в Иерусалиме в 1974. Мы раздали 184 испытуемым четыре кратких описания личности. Каждое описание соответствовало стереотипу отдельно взятой профессии (например, водитель такси) и резко отличалось от стереотипа отдельно взятой политической партии (например, лейбористы), или наоборот. Следовательно, каждое описание (X) было репрезентативно по отношению к одной цели, обозначенной А, и нетипичным для другой цели, обозначенной В. Каждое описание сопровождалось списком из пяти или шести целевых событий: занятие, политические симпатии или конъюнкцией событий, например, водитель такси, который является членом партии лейбористов. Для каждого описания, половина испытуемых получила список, где была как цель А, так и цель В, в то время как другая половина получила список, включающий сложную цель (А и В). Оставшиеся четыре цели были идентичны в двух списках. Половину испытуемых попросили оценить цели согласно «степени, в которой X репрезентативен по отношению к соответствующему классу», а другая половина оценивала их согласно «вероятности того, что X является элементом соответствующего класса».

Структура исследования позволила провести косвенное сравнение репрезентативности и вероятности для события В и сложного события (А и В) относительно четырех постоянных альтернатив. Были получены следующие результаты. Во-первых, было оценено, что все четыре описания более репрезентативны по отношению к сложной цели (А и В), чем только к цели В. Во-вторых, упорядочение репрезентативности и вероятности каждого набора целей было почти идентично во всех случаях; средняя корреляция между средними значениями была 0.96. В частности, сложной цели (А и В) было приписано значительно более высокое среднее значение в упорядочении вероятности, чем простой цели В. Очевидно, уверенность в эвристике репрезентативности привела к тому, что испытуемые оценивали конъюнктивное событие как более вероятное, чем один из его компонентов, вопреки правилу конъюнкции в теории вероятности. Назовем этот вид оценок эффектом конъюнкции.

Исследование 2: Билл и Линда

Поскольку стимульный материал, используемый в ранних исследованиях, был крайне специфичен для израильской культуры, мы построили английскую версию задач и провели аналогичное исследование с несколькими существенными изменениями. Во-первых, мы сравнили результаты межгруппового плана (в котором каждый испытуемый сравнивал либо сложную цель (А и В), либо простую цель В с тем же самым набором альтернатив) с результатами внутригруппового плана, в котором каждый респондент сравнивал непосредственно две критических цели. Мы выдвинули гипотезу, что правило конъюнкции не действует в первом упомянутом исследовании, как в нашем исследовании, но мы ожидали, что частота нарушений значительно уменьшится в последнем проекте, где участников просили сравнить $P(A)$ с $P(A \text{ и } B)$. Во-вторых, мы ожидали, что даже ограниченный опыт испытуемых в области статистики не допустит многих нарушений правила конъюнкции, по крайней мере, при внутригрупповом плане.

Чтобы исследовать эти гипотезы, мы провели прямое и косвенное исследование с тем же самым стимульным материалом. Аналогичные исследования были проведены в трех группах испытуемых, которые обладали различным опытом в области статистики. Первая группа испытуемых состояла из студентов старших курсов Университета Британской Колумбии и Стэнфордского университета, не обладающих подготовкой в теории вероятности или статистике. Промежуточная группа состояла из аспирантов в области психологии и педагогики и студентов-медиков из Стэнфордского университета, которые прошли курсы по статистике и были ознакомлены с основными концепциями теории вероятности. Третья группа состояла из аспирантов в области принятия решений Бизнес-школы Стэнфорда, каждый из которых прошел курсы повышенной сложности по теории вероятности и статистике.

Были составлены два кратких описания личности. Каждому участнику был предоставлен один из этих набросков при прямом исследовании, а другой при косвенном исследовании. В первом тесте, описание личности сопровождалось восьмью возможными исходами, включая репрезентативный исход, нерепрезентативный исход и конъюнкцию этих двух. При косвенном исследовании список исходов включал либо два отдельных критических исхода, либо их конъюнкцию. Образец для этих двух задач при прямом исследовании приведен ниже. Числа в круглых скобках — средние значения, приписанные испытуемыми различным исходам.

Биллу 34 года. Он умен, но лишен воображения, обязателен, но в общем скучный. В школе он был силен в математике, но слаб в общественных и гуманитарных дисциплинах.

Пожалуйста, упорядочите следующие утверждения по их вероятности, присвоив наиболее вероятному утверждению 1 и наименее вероятному 8.

(4.1) Билл — терапевт, хобби которого — игра в покер.

(4.8) Билл — архитектор.

- (1.1) Билл — бухгалтер. (А)
- (6.2) Хобби Билла — играть джаз. (J)
- (5.7) Хобби Билла — серфинг.
- (5.3) Билл — репортер.
- (3.6) Билл — бухгалтер, хобби которого — играть джаз. (А&J)
- (5.4) Хобби Билла — альпинизм.

Линде 31 год, не замужем, искренняя и очень яркая натура. Она увлекается философией. Будучи студенткой, интересовалась проблемами дискриминации и правосудия, а также участвовала в антиядерных демонстрациях.

Пожалуйста, упорядочите следующие утверждения по их вероятности, присвоив наиболее вероятному утверждению 1 и наименее вероятному 8.

- (5.2) Линда — преподаватель в начальной школе.
- (3.3) Линда работает в книжном магазине и занимается йогой.
- (2.1) Линда — активист феминистского движения. (F)
- (3.1) Линда — психиатр.
- (5.4) Линды — член Лиги женщин-избирателей.
- (6.2) Линды — кассир в банке. (T)
- (6.4) Линды — страховой агент.
- (4.1) Линда — кассир в банке и активист феминистского движения. (T&F)

Как читатель, вероятно, догадался, описание Билла было построено таким образом, чтобы быть репрезентативным по отношению к бухгалтеру (А) и нетипичным по отношению к человеку, чье хобби — джаз (J). Описание Линды было построено таким образом, чтобы быть репрезентативным по отношению к активной феминистке (F) и нетипичным для кассира в банке (T). В соответствии с психологическими принципами подобия (Tversky, 1977), мы ожидали, что вероятности сложных целей, такие как бухгалтер, который играет джаз (V&J) и кассир в банке, являющаяся активисткой феминистского движения (T&F), попадали бы между вероятностями соответствующих простых целей.

Чтобы проверить этот прогноз, мы попросили первую группу из 88 испытуемых, упорядочить эти восемь целей «по степени, в которой Билл (Линда) подобен типичному элементу соответствующего класса». Распределения подобия подтвердили наши гипотезы относительно описаний. Доля испытуемых, которые показали прогнозируемый порядок для Билла ($A > A$ и $J > J$), составила 87%; процент испытуемых, показавших прогнозируемый порядок для Линды ($F > T$ и $F > T$), был 85%.

Все участники получили описание либо Билла, либо Линды в форме, составленной для внутригруппового плана, и упорядочили эти восемь целей согласно их вероятностям. Полученные данные приведены в верхней части Таблицы 1, где в строке «эффект конъюнкции (%)» приведен процент испытуемых в каждой группе, которые оценили сложную цель как более репрезентативную, чем простую цель. Строки «А и В» и «В» представляют, соответственно, средние ранги, приписанные сложной и менее репрезентативной простой целям. Средний ранг подобия и средний ранг вероятности при внутригрупповом плане отражены на Рисунке 1.

Таблица 1. Эффект конъюнкции.

	Группа 1 "Наивные испытуемые"		Группа 2 Испытуемые с промежуточными знаниями в области статистики		Группа 3 Опытные в области статистики испытуемые	
	Линда	Билл	Линда	Билл	Линда	Билл
Внутригрупповой план						
Эффект конъюнкции (%)	89	92	90	86	85	83
Средний ранг: А и В	4.2	3.6	3.9	3.5	4.0	3.4
Средний ранг: В	6.3	6.4	6.2	6.4	6.1	5.6
N	88	94	53	56	32	32
Межгрупповой план						
Средний ранг: А и В	3.3	2.3	2.9	2.4	3.1	2.5
Средний ранг: В	4.4	4.5	3.9	4.2	4.3	4.6
N	86	88	55	56	32	32

При межгрупповом плане, две версии каждой проблемы были построены путем удаления из списка либо сложной цели, либо двух простых целей. Краткое описание личности, указания и оставшиеся пять целей были точно такими же, как для внутригруппового плана. Результаты опроса по межгрупповому плану для всех групп испытуемых представлены в нижней части Таблицы 1.

Результаты, сведенные в Таблицу 1, показывают, что сложная цель была оценена как более вероятная, чем простая критическая цель, как в проекте 1, так и 2. Этот результат, сохраняется как для двух описаний, так и для всех групп. К нашему удивлению, опыт в статистике имел незначительное влияние на эффект конъюнкции, что показали более чем 80% испытуемых во всех трех группах.

В предыдущих исследованиях критические цели были представлены в большем количестве возможных исходов, которые могли скрыть отношение включения между ними. Поэтому интересно выяснить, нарушают ли люди правило конъюнкции даже тогда, когда логическое соотношение между целями довольно очевидно. Чтобы проверить эту гипотезу, мы предоставили новой группе испытуемых (не имеющих опыта в области статистики) описания Билла и Линды. Каждому испытуемому было дано одно из двух

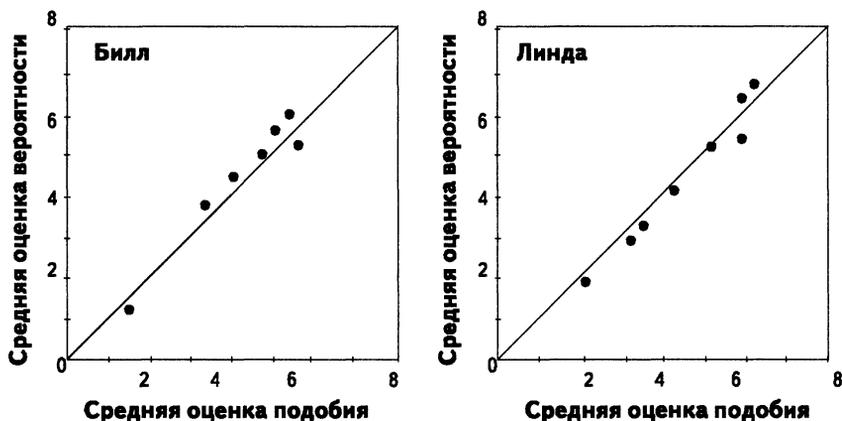


Рис. 1 График средних оценок для восьми результатов, оцененных по вероятности и по подобию для двух описаний.

описаний, и его спросили, какая из двух критических целей (то есть, J и (A и J), или T и (T и F)) более вероятна. Эта процедура не уменьшила эффект конъюнкции: в случае Билла сложная цель была выбрана 92% испытуемых ($N = 88$) и 87% испытуемых ($N = 86$) в случае Линды.

Общее игнорирование правила конъюнкции ставит перед нами интригующий вопрос относительно его нормативной достоверности. Чтобы исследовать этот вопрос, мы опросили 36 аспирантов промежуточной группы, которые участвовали в эксперименте. Их спросили, (1) как они упорядочили две критических категории, (2) почему они так сделали, и (3) рассмотреть утверждение, что “вероятность того, что Билл одновременно является бухгалтером и играет джаз, не может превышать вероятности того, что он играет джаз, потому что каждый член первой категории — всегда является членом последней”. Больше чем две трети испытуемых (1) сказали, что они выбрали сложную цель, (2) дали некую версию аргумента подобию или типичности как причины и (3) согласились, после некоторого размышления, что их ответ был неправильным, потому что он расходится с правилом конъюнкции. Только двое испытуемых утверждали, что упорядочение вероятности не должно согласовываться с принадлежностью классу, и только один сказал, что он не так понял вопрос. Хотя опрос, склонил испытуемых в пользу правила конъюнкции, результаты показывают, что испытуемые третьей группы (с опытом в области статистики), по крайней мере, расценивают нарушение этого правила как печальную ошибку.

Интерпретируя невыполнение правила конъюнкции, важно рассмотреть, относится ли эффект, полностью либо частично, к лингвистическим конвенциям или диалоговым правилам. Например, в раннем исследовании мы предоставили людям следующее описание, “Джон — 27 лет, является неординарной личностью. В колледже он был выдающимся атлетом, но не

демонстрировал способностей или интереса к интеллектуальным вопросам”. Мы обнаружили, что Джон, по оценкам испытуемых, более вероятно, является “преподавателем гимнастики”, чем просто “преподавателем”. Хотя каждый преподаватель гимнастики, в некотором смысле, преподаватель, можно поспорить, что термин *преподаватель* понимается здесь в смысле, который исключает значение преподавателя гимнастики или инструктора автошколы. При таком планировании этой проблемы удалось избежать, определив критический исход экстенсивно как пересечение двух множеств, например, бухгалтеров и джазовых музыкантов-любителей.

Нарушения правила конъюнкции также наблюдаются и в задачах на последовательность, где цель состоит из последовательности событий. Словик, Фишхофф и Лихтенштейн (Slovic, Fischhoff, Lichtenstein, 1976) предоставили испытуемым краткое описание личности человека, подходящего под стереотип инженера, а не журналиста. Их испытуемые приписали более низкую вероятность тому, что “Том В. выберет журналистику как основную специальность в колледже”, чем тому, что “Том В. выберет журналистику как основную специальность в колледже, но скоро разочаруется в своем выборе и переключится на технические науки”. Строго говоря, первый случай включает последний, и вышеупомянутое суждение нарушает правило конъюнкции. Однако относительно этого примера можно возразить, что, согласно нормативным правилам разговора, утверждение, что Том В. выбрал журналистику как основную специальность в колледже, подразумевает, что он также остался на этой специальности. Иначе, это утверждение ввело бы в заблуждение.

Подобные возражения могут также возникнуть относительно примеров Билла и Линды. Таким образом, можно поспорить, что испытуемые считают, например, категорию “кассир в банке” как “кассир в банке, не принимающая участия в феминистском движении” в отличие от данной категории “кассир в банке, которая является активисткой в феминистском движении”. Однако, наличие эффекта конъюнкции при межгрупповом плане, в котором критические цели не сравниваются непосредственно, указывает, что этот эффект нельзя адекватно объяснить в терминах переформулировки категорий цели, согласно стандартным диалоговым имплицативным формам. Скорее, наблюдаемые оценки показывают общую тенденцию оценивать вероятности соответствующих событий степенью, в которой Линда репрезентативна по отношению к типичным или прототипичным членам соответствующих категорий.

Кроме того, мы наблюдали эффект конъюнкции в нескольких задачах, в которых, отсутствовали диалоговые имплицативные формы. Следующие задачи, например, касаются прогнозирования будущих событий, где интерпретация В как (В и “не А”) кажется невероятной.

Исследование 3: Прогнозы на 1981 г.

Задачи, описанные в этой главе, были предназначены для того, чтобы проверить правило конъюнкции в прогнозировании событий реальной жизни, где испытуемые полагаются на свои общие знания. В исследовании этой проблемы в декабре 1980 принимала участие группа из 93 испытуемых, которые не имели опыта в области статистики. Группа получила следующие инструкции.

В этом анкетном опросе Вас просят оценить вероятность различных событий, которые могут произойти в течение 1981 г. Каждая задача включает четыре возможных события. Ваша задача состоит в том, чтобы упорядочить эти события по вероятности, присвоив 1 наиболее вероятному событию, 2 — второму, 3 — третьему и 4 — наименее вероятному событию.

Анкетный опрос включал шесть вопросов. Два вопроса приведены ниже. Результаты остальных вопросов оказались подобными. Числа в круглых скобках показывают средние оценки для каждого случая; мы также приводим процент испытуемых, которые оценили сложную цель как более вероятную, чем простую.

Теннис в 1981 г. (эффект конъюнкции: 72%)

Предположим, что Бьёрн Борг дойдет до финала в Уимблдоне в 1981. Пожалуйста, упорядочьте следующие результаты от наиболее до наименее вероятного.

- (1.7) Борг выиграет матч.
- (2.7) Борг проиграет в первом сете.
- (3.5) Борг выиграет в первом сете, но проиграет матч.
- (2.2) Борг проиграет в первом сете, но выиграет матч.

Политика США в 1981 г. (эффект конъюнкции: 68%)

Пожалуйста, упорядочьте следующие события по вероятности того, что они произойдут в 1981.

- (1.5) Рейган сократит федеральную поддержку местным органам власти.
- (3.3) Рейган обеспечит федеральную поддержку матерям-одиночкам.
- (2.7) Рейган увеличит бюджет на оборону меньше, чем на 5%.
- (2.9) Рейган обеспечит федеральную поддержку матерям-одиночкам и сократит федеральную поддержку местным органам власти.

Как и в предшествующих исследованиях, сложная категория была оценена как более вероятная, чем один из ее компонентов. Результат совместим с понятием репрезентативности, которая скорее касается в этом случае соотношения между причинно-следственной системой и ее исходами, чем подобия описания стереотипу. Во второй задаче, например, кажется нетипичным для президента Рейгана, обеспечить федеральную поддержку матерям-одиночкам и весьма репрезентативным для него сократить федеральную поддержку для местных органов власти. Конъюнкция этих действий оказывается промежуточной по репрезентативности, а оценки вероятности, очевидно, следуют тому же образцу.

В первой задаче большинство испытуемых оценили победу Борга на чемпионате как наиболее вероятное событие, а проигрыш Борга в первом сете как менее вероятное. Конъюнкция этих двух событий, а именно проигрыш Борга в первом сете, но победа в матче, оценивалась как менее вероятная, чем первая возможность, но более вероятная, чем вторая. Очевидно, испытуемые объединяли события скорее согласно принципам репрезентативности или причинно-следственного воздействия, чем согласно законам вероятности.

Обсуждение

Результаты, о которых сообщалось в предыдущих исследованиях, прямо подтверждают гипотезу, что люди оценивают вероятность событий по степени, в которой эти события репрезентативны по отношению к соответствующей модели или процессу. Поскольку репрезентативность события может быть увеличена за счет уточнения, сложная цель может быть оценена как более вероятная, чем один из ее компонентов. Этот прогноз был подтвержден исследованиями, использующими как внутригрупповой, так и межгрупповой план для совокупности испытуемых с различным уровнем подготовки в области статистики.

В отличие от других правил вероятности, таких как регресс к среднему, который труден для понимания испытуемыми с отсутствием опыта в области статистики, правило конъюнкции является и простым, и легко применимым. Большинство испытуемых проявило желание подтвердить его в абстрактной форме, хотя почти все они нарушили его на практике, когда оно находилось в противоречии с интуицией репрезентативности. Полученные результаты противоречат выводам Johnson-Laird и Wason (1977) относительно проверки утверждений “если то” (см. также Johnson-Laird, Legrenzi и Sonino-Legrenzi, 1972). Эти исследователи обнаружили, что большинство испытуемых потерпело неудачу в заданиях с абстрактным материалом, а не в конкретных примерах. С другой стороны, наши испытуемые подтвердили правило конъюнкции в абстрактной форме, но нарушили его в конкретных примерах (см. Главу 34).

Вывод о том, что конъюнкция часто кажется более вероятной, чем один из ее компонентов, может иметь далеко идущие последствия. Нет причин полагать, что суждения политических аналитиков, присяжных заседателей, судей и врачей независимы от эффекта конъюнкции. Этот эффект, вероятно, особенно негативно проявит себя при попытках дать прогноз относительно будущего, оценивая вероятности отдельно взятых сценариев. Словно смотря в хрустальный шар, политические деятели, футурологи, а также обыватели ищут образ того будущего, которое лучше всего представляет их модель развития настоящего. Этот поиск ведет к построению детальных сценариев, которые являются внутренне последовательными и высоко репрезентативными по отношению к нашей модели мира. Такие сценарии

часто оказываются менее вероятными, чем менее детальные прогнозы, которые фактически являются более вероятными. С увеличением детализации сценария, его вероятность может только устойчиво уменьшаться, но его репрезентативность, и, следовательно, его очевидная вероятность, может увеличиваться. Уверенность в репрезентативности, по нашему мнению, является первичной причиной безосновательного предпочтения детальных сценариев и иллюзорного смысла интуиции, который часто обеспечивают такие конструкции.

Путаница между суждениями вероятности и подобия имеет место не только при прогнозировании неопределенного будущего, но также и при реконструкции неопределенного прошлого, например, в истории или криминальных расследованиях. Здесь также, перечень прошлых событий часто объединяется в репрезентативный сценарий, который включает вероятные предположения относительно неизвестных событий. Включение таких предположений может только уменьшать вероятность того, что полный перечень событий является истинным, но это обеспечивает ощущение репрезентативности и последовательности, которое может увеличивать кажущуюся вероятность сценария. Например, гипотеза “обвиняемый покинул место преступления” может казаться менее вероятной, чем гипотеза “обвиняемый, покинул место преступления из страха быть обвинённым в убийстве”, хотя последнее менее вероятно, чем первое. Хороший рассказ часто менее вероятен, чем посредственный.

Наконец, важно понять, что эффект конъюнкции — это признак более фундаментальной проблемы. Он просто обнаруживает несогласованность между логикой вероятности и логикой репрезентативности, которая часто обуславливает мнения людей относительно неопределенных событий. Так как суждение человека неотделимо от решения волнующих проблем нашей жизни, конфликт между интуитивной концепцией вероятности и логической структурой этой концепции настоятельно требует разрешения. С одной стороны, мы не можем с готовностью отказаться от эвристики, которую мы используем, для оценки неопределенности, потому что многое из наших знаний о мире связано к ней. С другой стороны, мы не можем бросить вызов законам вероятности, потому что они определяют важные истины нашего мира. Нравится вам это или нет, А не может быть менее вероятно чем (А и В), и уверенность в противоположном — ошибочна. Наша задача состоит в том, чтобы сохранить всё полезное и эффективное в интуитивном суждении при исправлении ошибок и предубеждений, к которым оно склонно.

Часть III
**Причинность и
атрибуция**

7. Общепринятое положение: информация не обязательно информативна*

Ричард Е. Нисбетт, Юджин Борджида, Рик Крендалл, Харви Рид

Когнитивная теория, которая в настоящее время оказывает самое большое влияние на социопсихологов — это теория атрибуции, формализованная версия которой была представлена Гарольдом Келли (Harold Kelley) в 1967. Теория формулирует представление человека как ученого-любителя, который делает попытку вывести причины явлений, которые он наблюдает. Причины, которые он приписывает, определяют взгляд на его социальный мир, а эта точка зрения может определять его поведение. Чрезвычайно широкий диапазон явлений, начиная от исследования соответствия Аша (Asch) и заканчивая работой Шахтера (Schachter), посвященной эмоциям, может быть описан как частный случай процесса каузальной атрибуции в действии. Фактически, кажется вполне возможным, что наиболее важным вкладом Келли может, в конечном счете, считаться создание языка или карты дорог, с помощью которой можно описывать и искать взаимосвязи между разнообразными социально-психологическими явлениями.

В дополнение к своему содействию организации, Келли установил три формальных источника влияния на процесс каузальной атрибуции. В попытке приписать причины событиям типа “Человек реагирует способом X на ситуацию A”, исполнитель реагирует на три источника информации: *отличительность информации* (человек реагирует способом X во всех ситуациях общего типа или только в ситуации A?); *постоянство информации* (человек реагирует способом X всегда, при широком разнообразии обстоятельств или он реагирует способом X только иногда?); и *согласованность информации* (большинство людей реагирует способом X или реакция является нехарактерной). Приписывание причины будет зависеть от ответов на каждый из этих вопросов. Таким образом, человек считается первичной

*Это сокращенная версия книги, которая появилась в редакции J. S. Carroll и J. W. Payne., *Познание и социальное поведение*. Hillsdale, N.J.: Lawrence Erlbaum Assoc., Inc, 1976. Периздано в соответствии с разрешением.

причиной своей реакции настолько, насколько он реагирует подобным образом во всех ситуациях общего типа и настолько, насколько его реакции не демонстрируются другими.

Ситуация считается причиной настолько, насколько реакция человека уникальна по отношению к ситуации А и настолько, насколько его реакция типична.

Анализ процесса приписывания, проделанный Келли, как одобрялся, так и критиковался на том основании, что он банален. Нравится или нет это качество теории, к удивлению обнаруживается, что одна из ее фундаментальных аксиом не нашла фактически никакого подтверждения в последующем исследовании. Это положение о том, что люди отвечают на согласованность информации посредством локализации причины. Несмотря на теорию и здравый смысл, доказано, что знание поведения других людей мало влияет на приписывание причин. Знание того, что реакция человека глубоко типична, кажется, не приводит к выводу, что, скорее ситуация, чем исполнитель, является главным причинным фактором. И наоборот, знание того, что реакция человека является уникальной, не приводит к выводу, что скорее исполнитель, чем ситуация является главным причинным фактором.

Ниже мы рассматриваем доказательство того, что точка зрения, согласно которой люди используют согласованность информации при создании приписываний, не пользуется особой поддержкой. Это доказательство касается и случаев, где исполнителем является другой человек, а также случаев, взятых прежде всего из нашего собственного исследования, где исполнитель — это сам человек. Затем мы демонстрируем сходство между неспособностью согласованности информации повлиять на приписывание и доказательством Канемана и Тверского (1973, 4) того, что информация о базовом значении не в состоянии повлиять на прогнозы. Мы предлагаем объяснения двух неудач в терминах относительного влияния абстрактной (согласованность, базовое значение) информации против конкретной (исполнителя или связанной с целью) информации. Наконец, мы используем различие между абстрактной и конкретной информацией при вопросах общения и убеждения.

Согласованность информации и восприятие других людей

Существуют два исследования, которые изучают влияние согласованности информации на предсказание поведения других людей. Оба исследования показывают заметную слабость согласованности информации. Первое из них — исследование Л.З. МакАртура (L.Z. McArthur) (1972). Ее исследованием было прямое тестирование предположений Келли относительно эффектов влияния отличительности, постоянства и согласованности на каузальную атрибуцию. Испытуемым давали короткие описания типа “человек реагирует на стимул способом X” и дополнительно давалась информация по величинам относительно отличительности, постоянства и согласованности. Например, испытуемым могли сказать: “Во время танца Ральф наступал на

ноги Джоан”, и не указать дополнительно, что Ральф наступает на ноги почти всех девушек (или не наступает на ноги девушкам), или что Ральф почти всегда (или никогда не) наступает на ноги Джоан, и что почти все (или никто) не наступают на ноги Джоан. Затем испытуемых спросили, являлся ли случай наступания на ноги ошибкой Ральфа, ошибкой Джоан или виноваты обстоятельства. Испытуемых также попросили предсказать обобщенные реакции (насколько вероятно, что Ральф должен посоветовать Джоан записаться в танцевальный клуб?) и обобщения стимула (насколько вероятно, что Ральф наступал на ноги на ледяном тротуаре?).

Отличительность информация составила 10% разницы в причинно-следственном приписывании (подводящем итог по всем причинам) и 63% разницы в ожиданиях относительно обобщения стимула. Постоянство информации составило 20% разницы в причинно-следственных приписываниях и 14% разницы в ожиданиях относительно обобщения реакции. Напротив, согласованность информации составила меньше 3% разницы в любом из трех видов вывода. Эти результаты нарушают не только здравый смысл теории приписывания, но и здравый смысл вообще. Хотя испытуемые полагают, что важно знать, наступает ли Ральф на ноги большинству девушек и наступает ли он обычно на ноги Джоан, для них совершенно не важно, наступают ли на ноги Джоан другие люди!

Здравый смысл приписывания или любой другой его разновидности также нарушается в другом исследовании, касающемся восприятия других людей. Миллер, Гиллен, Шенкер и Редлав (Miller, Gillen, Schenker, Radlove, 1973) попросили студентов колледжа прочитать раздел классического исследования Милграма (Milgram, 1963), в котором изучается повиновение. Половине испытуемых дали фактические данные исследования Милграма, показывающие, что практически ко всем испытуемым был применён электрошок значительной силы, а к большинству был применён электрошок максимальной силы. Других же испытуемых уверили, что такое поведение будет редким. Затем всех испытуемых попросили оценить двух человек, к которым был применён электрошок максимальной силы. По 11 признакам с большой нагрузкой на оценочный компонент, например, привлекательность, доброжелательность, агрессивность. Только на одну из этих 11 оценок согласованность информации имела существенное влияние. Знание того, что применения электрошока максимальной силы было модальным поведением, фактически не повлияло на оценки людей, перенесших максимальный шок.

Согласованность информации и самовосприятие

Согласованность информации также имеет небольшое влияние на атрибуции, сделанные относительно самого человека. Бем (Bem, 1967) предложил, а Келли (Kelley, 1967) ввел в теорию атрибуции понятие того, что люди производят причинно-следственный анализ своего собственного поведения в

полном соответствии с их приписыванием относительно поведения других людей. Они наблюдают их реакцию, принимая во внимание ситуации, в которых она протекает, и делают выводы относительно их чувств и мотивов. Например, испытуемый в классическом эксперименте Шахтера и Зингера (Schachter и Singer, 1962), который знает, что ему был введен препарат, который возбуждает эмоции и затем смоделирована ситуация, вызывающая сильные эмоции, проводит своего рода причинно-следственный анализ. Он чувствует признаки пробуждения эмоций, которые обычно он может приписать эмоциональному воздействию ситуации, но вместо этого приписывает их препарату, который он принял. В результате он проявляет в поведении меньше признаков эмоций, чем испытуемые, которые не знали, что им был введен такой препарат, и даже чем контрольная группа испытуемых, которым препарат, возбуждающий эмоции, не был введен вообще. Значит, испытуемый воспринимает причину своих автономных реакций как “внешнюю” по отношению к себе и ведет себя соответственно.

Несколько лет назад, мы начали программу терапевтических вмешательств, основанных на точке зрения, что людей можно заставить искать причину их реакций во внешней среде. Предполагалось, что всякий раз, когда человек реагирует неадекватно или патологически, можно добиться определённого результата, убедив человека приписать его реакции чему-то внешнему по отношению к себе. Первое исследование, единственное успешное, было проведено Стормсом и Нисбеттом (Storms и Nisbett, 1970). Исследуемой патологией была бессонница. Мы попросили студентов, у которых были проблемы с засыпанием, за 15 минут перед сном принимать пилюлю (на самом деле «пустышку»), которая, как им сказали, ускоряла сердцебиение, делала дыхание быстрым и нерегулярным, обладала согревающим эффектом и приводила тело в общее состояние тревожности. Это – конечно признаки бессонницы. Как оказалось, испытуемые, принимавшие эти пилюли, как сообщалось, засыпали быстрее, в те ночи, когда они их принимали, чем когда не принимали, и быстрее, чем контрольная группа испытуемых, которые не принимали никаких пилюль. Стормс и Нисбетт пришли к выводу, что один или оба процесса приписывания могут объяснить результаты. Бессонница вызвана в значительной степени состоянием эмоционального возбуждения в ночное время, за счет любого количества причин, включая беспокойство относительно личных проблем, неудобного режима дня или хронического невроза. Поскольку человек лежит в кровати в состоянии эмоционального возбуждения, его сон включает мысли с эмоциональным содержанием. Эмоциональное возбуждение может связаться с эмоциональным переживанием и усиливать его. Результирующее усиленное эмоциональное состояние увеличивает эмоциональное возбуждение и так далее, по замкнутому кругу. Однако этот круг может быть разорван, за счет знания того, что возбуждение является внешним. В этом случае человек не делал бы выводов относительно того, как он взволнован по поводу экзамена или насколько сердит он был на соседа по комнате и из наблюдения собствен-

ного эмоционального состояния. В те ночи, когда человек принимает пилюли, эмоциональное возбуждение воспринимается неизбежным и, таким образом, не связанным с любыми мыслями в голове. Так как замкнутый круг усиленного эмоционального возбуждения, таким образом, был разорван, сон мог наступить.

Альтернативно, или дополнительно, мог бы действовать несколько другой процесс более общего плана. Наши страдающие бессонницей испытуемые сообщили, что они были весьма обеспокоены тем фактом, что они страдают бессонницей. Они воспринимали это как свидетельство более общей патологии и как нечто, плохо отражающееся на состоянии психического регулирования. Для испытуемого с такими переживаниями, знание о неизбежном, произведенном извне эмоциональном возбуждении должно оказать благотворное влияние. Испытуемый мог бы рассуждать, что, по крайней мере, сегодня вечером бессонница не может считаться свидетельством общей психопатологии. Поскольку подобное беспокойство частично является причиной бессонницы, сон должен наступать быстрее в ночи, когда применялись пилюли.

Попытки управлять депрессией

Вооруженные этим успехом в ситуации со страдающими бессонницей, мы начали попытки изменений состояния депрессии. Методикой во всех исследованиях была манипуляция согласованности, разработанная таким образом, чтобы перенести во внешнюю среду причину депрессивного аффекта, убеждая испытуемых, что она широко распространена. Поскольку это состояние разделяется другими людьми, само присутствие влияет на человека менее отрицательно. Должно казаться, что это состояние не определяет индивидуальные, возможно патологические, реакции испытуемого на отдельно взятые обстоятельства. С уменьшением волнения и беспокойства относительно способности испытуемого справляться с жизненными ситуациями, депрессия могла бы частично уменьшиться.

Исследование 1: Воскресная хандра. Многие студенты испытывают общее ухудшения настроения по воскресеньям. Хотя день может начинаться достаточно хорошо с поздним завтраком, кофе и воскресными газетами, ощущение беспокойства часто начинается днем. Необходимо много сделать на предстоящую неделю, слишком много, чтобы всерьез подумать о воскресном пикнике, хотя не достаточно много, чтобы начать работу сию минуту. К концу дня, оказывается, что не было ни активного отдыха, ни работы, воскресные газеты, причем возможно даже разделы о финансах и путешествиях, были тщательно изучены, а впереди маячит длинный вечер утомительной учебы. К вечеру, если все складывается плохо, как это часто бывает, работа идет с трудом, если идет вообще, и воцаряется хмурое настроение и неуверенность в себе.

Мы предположили, что, если бы это явление было общераспространенным и люди знали об этом, интенсивность воскресной хандры могла бы уменьшиться. Если бы отдельно взятый студент знал, что комнаты в общежитии полны людей в том же состоянии, то его собственные отрицательные эмоции должны быть несколько смягчены. Вместо того чтобы подумать, что он не создан для студенческой жизни или что у него никогда не будет серьезных отношений с женщиной, он может просто сказать себе, что люди имеют тенденцию грустить по воскресеньям и оставить это как есть.

Чтобы проверить это предположение, мы просили большое количество студентов мужского пола старших курсов в Йельском университете заполнить несколько шкал настроения в 16:00 и 22:00 по воскресеньям. Шкалами настроения были несколько шкал Вессмана и Рикса (Wessman и Ricks, 1966), с большим акцентом на факторе повышения–снижения настроения. Кроме того, испытуемых попросили заполнить анкетный лист в 22:00, относительно их успеваемости и социальной активности в течение дня и количества случаев, в которых они дали выход некоторому аффекту понижения настроения, например, когда они кричали или плакали. Наконец, испытуемые брали стопку кассет с мультфильмами и оценивали их по забавности.

После предварительных измерений по воскресеньям испытуемые делились на три группы, по 18 испытуемых в каждой. Одной группе, группе контроля, просто говорилось, что исследователи изучали настроение по воскресеньям, и участников просили в следующее воскресенье заполнить тот же самый пакет материалов. Для другой группы, подробно был описан синдром воскресной хандры, и испытуемым дали (ложную) статистику, чтобы указать, насколько широко она распространена среди студентов. Испытуемым говорили, что 92% студентов Йельского университета испытывали подобное явление по крайней мере иногда, в то время как 65% испытывали его в большинство воскресений. Третьей группе дали ту же самую согласованную информацию, как и второй группе, и, кроме того, дали теорию, объясняющую это явление. Испытуемым сказали, что оно вызвано “крушением эмоционального возбуждения” по воскресеньям: нормальное эмоциональное возбуждение буднего дня обычно сопровождается более высоким эмоциональным возбуждением в субботу; и тогда в воскресенье, наблюдается его резкий упадок. Этот недостаток эмоционального возбуждения часто интерпретируется как депрессия или преобразовывается в депрессию.

Ожидалось, что испытуемые в последних двух экспериментальных группах дадут иное толкование своим грустным переживаниям в воскресенье, менее принимая их на свой счет и становясь менее подавленными по поводу своей депрессии. Если это так, они должны были показать уменьшение аффекта понижения настроения по шкале с первого воскресенья на следующее; уменьшение поведения, вызванного понижением настроения, такого как вспышки гнева или плача; увеличение успеваемости и социальной активности; и более высокую среднюю оценку забавности мультфильмов в те-

чение второго воскресенья. Но ничего подобного не произошло. Настроение испытуемых не улучшилось ни на один показатель по сравнению с испытуемыми контрольной группы.

Исследование 2: Хроническая депрессия. Не слишком разочарованные, мы сделали попытку подобного вмешательства со студентами мужского пола старших курсов, которые описали себя как имеющих хроническую депрессию. Двадцать испытуемых были привлечены посредством рекламы в студенческой газете Мичиганского университета, которая призывала “студентов мужского пола старших курсов, страдающих депрессией, участвовать в исследовании депрессии Институтом социального исследования”.

По прибытии в лабораторию, испытуемых наугад разделили на две группы. Испытуемым из контрольной группы сказали, что разработали новую шкалу настроения и что необходимо, используя ее, получать ежедневные записи настроения людей, описывающих себя как страдающих от депрессии. Это помогло бы оценить валидность шкалы и ее способность обнаруживать изменения настроения.

Испытуемым экспериментальной группы дали ту же самую информацию и, кроме того, сказали, что экспериментаторы находились на заключительной стадии испытания теории депрессии на молодых людях мужского пола. Теория, основанная на фактах, по крайней мере, в частных случаях, гласила следующее. Испытуемым сообщили, что в течение некоторого времени известно, что изменения настроения у взрослых зависит частично от присутствия гонадальных гормонов – у мужчин от тестостерона. До недавнего времени считалось парадоксальным, однако, что дети почти никогда не испытывают депрессии. Так как дети имеют чрезвычайно низкий уровень всех гонадальных гормонов, это казалось противоречием общему правилу, что гормоны вызывают хорошее настроение. Этот парадокс недавно был разрешен открытием, что лимбические доли, (эмоциональный центр мозга) в юношеские годы становится зависимым от гонадального гормона, который регулирует настроение. У большинства мужчин, переключение рассчитано таким образом, чтобы соответствовать повышению уровня тестостерона, который достигает пика приблизительно в 25 лет. У многих молодых людей, однако, переключение завершается до того, как «топливо», если можно так выразиться, доступно в достаточных количествах, чтобы регулировать изменения настроения.

Большое значение придавалось тому, что испытуемый и был таким молодым человеком. Ожидалось, что манипуляция улучшит настроение по трем взаимосвязанным причинам:

1. Для депрессии предполагалось временное ограничение.
2. Отрицательное влияние, было смещено «наружу» в смысле, что оно теперь может быть приписано скорее неудачному биологическому инциденту, чем течению собственной жизни испытуемого и какой-либо патологической неспособности справиться с этим.

3. Отрицательный эффект и его причины наблюдались у многих других людей, таким образом, подразумевая распространённость проблем испытуемого и его реакции на них.

Всех испытуемых просили заполнять анкеты настроения Вессмана и Рикса в конце каждого дня в течение 2-недельного периода. Анкетный опрос также включил сообщение о том, как испытуемый спал в предыдущую ночь, потому что нарушения сна — частые признаки депрессии. Наконец, в конце семестра, в котором проходило исследование, были получены средние оценки успеваемости испытуемых.

Не имелось никаких различий в сообщениях о настроении испытуемых в экспериментальной группе и в группе контроля в любом пункте в течение 2-недельного периода, также не было различий в сообщениях о качестве сна. Не было фактически и намека на какую-либо тенденцию в направлении гипотезы по этим переменным. Испытуемые экспериментальной группы получили более высокие средние оценки, как и было спрогнозировано, но это не имело статистической значимости ($0.05 < p < 0.10$)...

Атрибуция и психология прогнозирования

Канеман и Тверски продемонстрировали логическую ошибку, которая похожа на ошибку, наблюдаемую в исследовании эффектов согласованности информации. В книге под названием «Психология прогнозирования» (1973, 4) эти исследователи показали, что люди не принимают во внимание базовое значение совокупности при прогнозировании категориальной принадлежности целевого элемента совокупности...

Если на испытуемых не влияет информация базового значения в их прогнозах относительно принадлежности целевого события категории, тогда едва ли вероятно, что на их приписывания, будет влиять согласованность информации. Согласованность информации — это то же самое, что информация базового значения. Это информация базового значения относительно скорее поведенческих реакций, чем принадлежности категории. Приписывание, кроме того, является более сложным и непрямым выводом, чем прогнозирование. Канеман и Тверски попросили испытуемых произвести довольно прямую и несложную цепь вывода: «если большинство членов популяции принадлежат некой категории, то тоже самое можно сказать и о целевых испытуемых». Их испытуемые были не в состоянии сделать такой вывод. В рассмотренном исследовании атрибуции, требуется еще более сложная цепь вывода: «если большинство элементов совокупности ведут себя некоторым образом, то ситуация должна оказывать сильное давление на такое поведение, и поэтому не стоит обращаться к личным отличительным особенностям, чтобы объяснить поведение целевого испытуемого, если оно модально».

Остаётся проверить, не желают ли испытуемые применять поведенческие базовые значения к прогнозированию целевых случаев. Если нет, то

необходимый нам вопрос должен быть изменен от «почему люди не могут изменять свои приписывания в соответствии с согласованной информацией» на более фундаментальный «почему люди обращаются с информацией базового значения, как будто она неинформативна».

Исследование 5: Поведенческие базовые значения, прогнозирование и атрибуция. Чтобы исследовать вопрос о желании людей изменять свои прогнозы, если им дана поведенческая информация базового значения, испытуемым были описаны два психологических эксперимента (Nisbett и Borgida, 1975). Студенты университета штата Мичиган прочитали детальные описания (а) эксперимента Нисбетта и Шахтера (1966), в котором испытуемых просили испытать на себе электрошок такой силы, какой они могли перенести и (б) эксперимента Дарли и Латане (Darley и Latane, 1968), относительно оказания помощи, в котором во время того, как несколько студентов обсуждали проблемы управления колледжем через наушники в отдельных кабинках, у одного из «испытуемых» начиналось что-то напоминающее приступ. Эти два эксперимента были выбраны, потому что в нашем опыте обучения, предположения студентов колледжа относительно поведенческого базового значения были известны. Несмотря на то, что студенты склонны предполагать, что немногие испытуемые получают сильный электрошок, модальное поведение — это вытерпеть самый большой разряд, который может выдать аппарат, то есть достаточный для того, чтобы заставить всю руку испытуемого непреднамеренно дергаться. И хотя студенты предполагают, что большинство людей быстро покинуло бы свои кабинки, чтобы помочь в ситуации приступа, типичный испытуемый никогда ничего не делает, чтобы помочь жертве в опыте, проведенном Латаном и Дарли с шестью испытуемыми.

Поскольку испытуемые не знали об истинных поведенческих основных значениях, было возможно дать некоторым из испытуемых фактическое базовое значение этих двух экспериментов, и, таким образом, создать дифференциальные условия базовой информации. Испытуемым со знанием базовой пропорции (условие согласованности информации) показали краткие видеосъемки интервью со студентами, описанными как испытуемые в первоначальных экспериментах (или в одном варианте условия согласованности информации, испытуемым показали краткие письменные описания биографии и личности студентов). Затем испытуемых из группы согласованности информации попросили предсказать, как целевые субъекты, о которых они просмотрели видеозапись или прочитали, вели бы себя. Поэтому можно сравнить прогнозы испытуемых группы согласованности информации, как с фактической информацией базового значения, которой они обладали, так и с предположениями относительно базового значения, сделанными испытуемыми, при отсутствии согласованности информации.

На Рисунке 1 показаны результаты эксперимента с электрошоком. Верхняя гистограмма показывает фактические данные базового значения, пре-

доставленные испытуемым из группы согласованности информации. Вторая гистограмма даёт оценки базового значения, составленные испытуемыми без знания базового значения. Очевидно, что оценки испытуемых этой группы весьма отличаются от фактических данных. Они приняли получение умеренного количества электрошока как модальное поведение. Третья гистограмма представляет предположения испытуемых группы согласованности информации относительно поведения целевых субъектов, о которых они просмотрели видеоматериал или прочитали. Хотя испытуемые были полностью осведомлены о базовом значении, может быть замечено, что распределение даже отдаленно не походит на фактическое базовое значение. Вместо этого, распределение подобно распределению базового значения, сделанного группой испытуемых, без знания базового значения. Результаты эксперимента об оказании помощи были полностью подобны.

Этот эксперимент дал возможность проверить другую гипотезу, предложенную Тверским и Канеманом (1971, 2) в статье под названием «Вера в закон малых чисел», где эти авторы доказывали, что даже ученые уделяют мало внимания размеру выборки и склонны делать необоснованные выводы относительно совокупности исходя из знания о небольшом количестве случаев. Чтобы проверить эту точку зрения в данном контексте, некоторым испытуемым не сообщили о базовых значениях в обоих экспериментах и показали им краткие видеointервью с двумя испытуемыми из каждого эксперимента. Испытуемым со знанием целевой информации сказали, что испытуемые в обоих экспериментах вели себя наиболее экстремальным образом, то есть, что оба испытуемых в эксперименте с электрошоком приняли максимально возможный удар током, и оба испытуемых в эксперименте об оказании помощи так и не помогли жертве. Затем испытуемых со знанием целевой информации попросили указать, каково, по их мнению, будет распределение всей совокупности испытуемых в экспериментах. В обоих экспериментах, испытуемые были склонны сделать вывод, что образ действий совокупности будет идентичен поведению двух испытуемых, которых они видели. В нижней гистограмме на Рисунке 1 можно увидеть, что оценки базового значения в эксперименте с электрошоком были подобны истинному базовому значению. Оценки не были настолько подобны для эксперимента с оказанием помощи, но они были довольно приближены к J-образной кривой базовой пропорции.

Испытуемые были склонны делать вывод, что образ действия совокупности был подобен поведению двух рассмотренных ими случаев, и когда процедура формирования выборки была неопределенна, и когда их внимание неоднократно привлекалось к тому факту, что эти два случая были отобраны наугад из шляпы, содержащей имена всех испытуемых.

Подведем итог, испытуемые не использовали информацию базового значения, когда их попросили, сделать прогноз относительно поведения целевых субъектов. Важно заметить что, помимо вопросов прогнозирования, было задано несколько вопросов по атрибуции, например, влияли ли обстоя-



Рис. 1. Базовое значение, оценки базового значения и прогнозы о целевых испытуемых в эксперименте с электрошоком.

ательства или личные склонности на поведение данного человека. Согласованность информации не имела существенного влияния на вопросы атрибуции. Последняя ошибка кажется фактически неизбежной при наличии предыдущей ошибки. Поэтому, вопрос относительно того, почему люди должны игнорировать согласованность информации в создании атрибуций, должен быть сокращен до более фундаментального вопроса относительно того, почему информация базового значения должна игнорироваться даже для такого простого вывода как прогнозирование. Любой ответ на этот фундаментальный вопрос относительно неспособности людей, извлечь информацию из информации базового значения в идеале должен одновременно относиться и к другому главному выводу в данном исследовании. Этот вывод состоит в том, что испытуемые, в действительности, “чрезмерно” пресыщены информацией целевого случая, они склонны принять то, что экстремальное поведение является модальным, если говорить, что только два испытуемых вели себя наиболее экстремальным способом.

Абстрактная информация против конкретной информации

Канеман и Тверски (1973, 4) не размышляли о причинах того, почему людям не удается использовать информацию базового значения. Их основное объяснение сосредотачивается на идее, что люди просто не очень хорошо умеют обращаться с вероятностными данными. Даже в сфере азартных игр, где люди имеют, по крайней мере, некоторое элементарное представление о том, как обращаться с вероятностями, они могут проявлять примечательную слепоту и предубеждения. Вне таких ситуаций люди могут быть абсолютно не в состоянии увидеть необходимость такой “простой” вероятностной информации как базовое значение. Или, непонимание того, как должным образом комбинировать информацию базового значения с информацией целевого случая, ведет к тому, что люди просто игнорируют информацию базового значения в целом.

Имеется, конечно, значительная доля правды в точке зрения, что люди не умеют работать с вероятностной информацией. Фактически, существует единственное исключение в литературе приписывания из правила, что люди игнорируют согласованность информации. Это — ясное свидетельство использования базового значения успехов и неудач при создании приписываний относительно способностей отдельно взятого человека (Weiner и другие, 1971). Если большинство людей терпит неудачу в отдельно взятой задаче, то целью человека воспринимается приобретение больших способностей; если люди преуспевают, то целью воспринимается приобретение более низких способностей. Конечно, на протяжении всей жизни мы приобретаем опыт относительно поведения других людей. Способность, фактически, является по определению выводом, полученным из базового значения.

Нам кажется, однако, что может также действовать другой принцип. По своей природе базовое значение или согласованность информации расплывчата, незначительна и абстрактна. Напротив, информация целевого случая ярка, значительна и конкретна. В исследовании депрессии, мы делали попытку противопоставить воспоминания о довольно сухой, статистической информации и живую, настойчивую реакцию на стимулы реального мира. В изучении дегустации крекера согласованность, абстрагированная от информации об уровне жидкости в бутылках, была противопоставлена чувственному впечатлению. В исследовании с электрошоком и об оказании помощи, данные о частоте были противопоставлены видеосвидетельству или письменному описанию реально существующего человека с родителями, планами относительно карьеры, увлечениями и индивидуальными особенностями. Несмотря на логическую уместность информации базового значения, она всё-таки не будет значительной для продолжения познавательного процесса.

Эта гипотеза не нова. В 1927 году Бертран Рассел (Bertrand Russell) предположил, что «общепринятая индукция зависит от эмоционального интереса случаев, но не от их числа» (стр. 269). В экспериментах, проведенных

Канеманом и Тверски, а также в исследованиях, проведенных нами и другими относительно эффектов согласованности информации, простое представление количества случаев было противопоставлено случаям эмоционального интереса. Согласно с гипотезой Рассела, эмоциональный интерес в каждом случае брал верх.

Мы предполагаем, что конкретная эмоционально интересная информация обладает большим потенциалом делать выводы из-за подобия «сценариев», вызывающих данную информацию, или схем, вовлекающих подобную информацию. И вывод далее следует по хорошо изученным строкам предыдущего сценария. Абстрактная информация менее богата потенциальными связями с ассоциативной сетью, посредством которой можно достичь сценариев. В соответствии с этим предположением, Нисбетт и Боргида (Nisbett и Borgida, 1975) обнаружили, что согласованность информации относительно поведения других людей в экспериментах с электрошоком и оказании помощи, не только не повлияла на предсказание испытуемых о том, как бы они себя вели, находясь в подобных условиях, но и не была ни разу упомянута в последующем опросе, почему они сделали такие предсказания. Вместо этого испытуемые заострили внимание на определённых деталях экспериментальной ситуации и соотнесли их с похожими ситуациями из своего жизненного опыта. «Уверен, я бы помог парню, потому что у меня был друг, чья сестра болела эпилепсией».

Гипотеза Рассела имеет несколько важных посылок к действию в повседневной жизни. В качестве иллюстрации рассмотрим простой пример. Предположим, Вам необходимо купить новую машину, и в целях экономии и долговечности Вы решили приобрести одну из таких солидных шведских машин среднего класса, как Volvo или Saab. Будучи осторожным покупателем, Вы идёте в *службу потребителей*, которая сообщает Вам, что по результатам экспертных исследований Volvo превосходит по механическим параметрам, а обыватели отмечают более высокую износостойчивость. Вооружённый информацией, Вы решаете обратиться к дилеру фирмы Volvo до конца недели. Между тем на одной из вечеринок Вы рассказываете знакомому о своём намерении, его реакция заставляет Вас задуматься: «Volvo! Должно быть ты шутишь. У моего шурина была Volvo. Сначала замысловатая компьютерная штука, обеспечивающая заправку топливом, вышла из строя. 250 баксов. Затем у него начались проблемы с задним мостом. Пришлось его заменить. Потом трансмиссия и сцепление. Через три года продали на запчасти». Логический статус этой информации таков, что количество из нескольких сот обывателей, владеющих Volvo из *службы потребителей*, возросло на единицу и что средняя частота осуществления ремонта понизилась на йоту по трём или четырём измерениям. Однако, каждый, кто утверждает, что он не примет во внимание мнение случайного собеседника, либо не искренен, либо совершенно себя не знает.

Исследование 6. Подверженность влиянию абстрактной информации против конкретной информации. Кажется, необходимо воплотить на практике мысленный эксперимент *Службы Потребителей* (Borgida и Nisbett, 1977). Так как нашими испытуемыми были студенты-психологи Университета штата Мичиган, мы избрали курсы по психологии в университете в качестве потребительского товара. Были выбраны десять лекционных курсов высшего уровня по психологии, различающиеся заявленным качеством. Группа студентов младших курсов, планирующих выбрать психологию как свою специализацию, была встречена в кабинете экспериментатором. Он сообщил студентам, что является членом комиссии факультета, которая занимается проблемами долговременного планирования работы отделения. Одна из проблем касается определения того, сколько студентов выберет тот или иной курс в будущем. Для того чтобы определить будущее распределение студентов по курсам, испытуемых просили заполнить экспериментальное расписание и отразить их планы в области психологии на старших курсах.

Испытуемых контрольной группы контроля просили посмотреть на учебный план (в действительности имитацию, состоящую из 27 курсов, исключая лабораторные занятия, занятия по статистике) и поставить галочку рядом с 5–10 курсами, которые они хотят слушать, и обвести кружком пометки для курсов, на которые они точно запишутся.

Двум экспериментальным группам сообщили, что с целью облегчения их выбора им дадут дополнительную информацию относительно лекционных курсов, которые выбирают большинство студентов. Для обеих групп эта информация состояла из детального описания, более понятного, чем аннотация в учебном плане, содержания и структуры каждого из десяти курсов. Затем для испытуемых одной экспериментальной группы организовали беседу со студентами старших курсов. Эти студенты коротко охарактеризовали каждый из десяти курсов, которые они прослушали. От одного до четырёх, человек обычно два или три, характеризовали каждый курс. Каждая характеристика начиналась с оценки, включающие следующие пять терминов: «отлично», «очень хорошо», «хорошо», «удовлетворительно», «плохо». Вот один из примеров:

Хотя курс включает большое количество материала, которое нужно пройти, он очень хорошо изложен. Вы имеете представление, о своем положении, что очень полезно при прохождении курса. Это очень обширная и важная область психологии, но причина, по которой я оценил этот курс, как «очень хорошо», а не «отлично» в том, что материал не особенно стимулирует мышление.

В другом экспериментальном условии (условие базового значения), испытуемым было сказано, что они будут читать средние оценки курса, основанного на шкалах, заполненных всеми студентами курса в конце предшествующего семестра. После описания каждого курса приводилась пятибалльная шкала, помеченная от «отлично» до «плохо». На каждой шкале была

Таблица 1. Среднее число выбранных курсов и взвешенная тенденция выбора

Условие	Рекомендованные курсы		Нерекомендованные курсы		Неупомянутые курсы	
	Количество	Изменная тенденция выбора	Количество	Изменная тенденция выбора	Количество	Изменная тенденция выбора
Беседа со старшекурсниками (N=22)	4.73 ^a	8.31 ^{a,b}	0.50 ^a	0.77 ^{a,b}	3.09 ^{a,b}	4.32 ^{a,b}
Наличие знаний базового значения (N=18)	4.11	6.33 ^b	0.94	1.56 ^b	4.17 ^b	5.89 ^b
Группа контроля (N=18)	3.33 ^a	5.22 ^a	1.39 ^a	2.17 ^a	5.39 ^{a,b}	7.17 ^a
F (2,55)	6.14 [*]	10.34 ^{**}	6.59 [*]	6.65 [*]	13.24 ^{**}	8.19 ^{**}

^a означает, что значения в колонке различаются одно от другого на уровне 0.01 по тесту Ньюмена-Келса (Newman-Keuls)

^b означает, что значения в колонке различаются одно от другого на уровне 0.05 по тесту Ньюмена-Келса

* $p < 0.005$, ** $p < 0.001$

помещена пометка, показывающая среднюю оценку, и было обозначено количество студентов, на мнение которых было основано среднее значение. Эти N располагались по порядку от 26 до 142. Средняя оценка каждого курса указывалась таким образом, чтобы быть идентичной средним оценкам, данным студентами-старшекурсниками, принимавшими участие в беседе.

Структура эксперимента позволяет сравнить эффективность рекомендаций, основанных на информации из первых рук, то есть кратких комментариях двух или трех студентов, которые проходили курс, с эффективностью более устойчивой, глубоко обоснованной информации. В таблице 1 представлено среднее количество рекомендованных (средние оценки 2.50 или лучше), нерекомендованных (средние оценки 3.75 или хуже) и неупомянутые курсы, выбранные этими тремя группами. Около каждой категории приведена оцененная тенденция выбора, индекс, который дает нулевую надбавку к курсу, если его не выбрали, 1, если он был помечен «галочкой» и 2, если он был обведен «кружком».

Можно заметить, что проведение беседы со старшеклассниками имело большее воздействие на выбор курса. Испытуемые этой группы были склонны пройти рекомендованные курсы и менее склонны пройти нерекомендо-

ванные или неупомянутые курсы, чем испытуемые контрольной группы. Напротив, метод базового значения повлиял только на выбор неупомянутых курсов.

Можно поспорить, что группа испытуемых, у которых была беседа со старшекурсниками, обладала большим количеством информации, чем группа базового значения. Одной из версий для этого аргумента является именно то, что мы хотели предложить. Наши студенты вели себя, как будто они извлекли больше информации из комментариев пары человек в личной беседе, чем из сухих, статистических данных целых совокупностей. Другая версия этого аргумента, – то, что комментарии, сделанные студентами при личной беседе, содержат очень ценную информацию, не доступную в условиях базового значения, относительно, например, организации курсов, процедур аттестации.

Чтобы задействовать последнее возражение, мы провели аналогичное исследование с одним важным изменением. Группе базового значения дали дословные письменные комментарии, сделанные студентами-старшекурсниками, участвующими в беседе. Кроме того, эти комментарии были явно описаны как *репрезентативные* взгляды студентов, проходящих курс, отобранные из общего количества оценок в конце семестра. В этом случае, на испытуемых, которым были доступны устойчивые средние оценки, основанные на больших и полных совокупностях, с дословными комментариями студентов, и со “знанием”, что они являются репрезентативными комментариями, было оказано меньше давления в их выборах, чем на испытуемых, которые просто слышали устные комментарии старшекурсников в личной беседе.

Коммуникация, направленная на конкретизацию

Не трудно понять выводы теории Бертрانا Расселла относительно общепринятой индукции и вышеупомянутой ее иллюстрации для общих вопросов общения и убеждения. Если на людей не влияют сухие, статистические данные, дорогие для сердец ученых и специалистов, занимающихся политическим планированием, то социальный и технологический прогресс должен приостановиться до тех пор, пока не будут найдены эффективные, конкретные, эмоционально интересные пути коммуникативных заключений. Мы собрали несколько исследований убеждения, которые, по нашему мнению, хорошо понятны в терминах различия между абстрактной и конкретной информацией. Мы представляем их ниже в надежде, что они могут служить источником вдохновения реальной жизни и руководством для исследования вопросов, касающихся природы информации и ее воздействия.

1. Ранняя версия Зеленой Революции стала возможной в начале 30-ых годов XX века благодаря прогрессу в методах ведения сельского хозяйства. Правительство должным образом продолжало информировать фермеров

страны об этих методах посредством сельскохозяйственных агентов по округам, распространяющих статистику и правительственные брошюры, и ждало сообщений об увеличении урожая. Таких сообщений не последовало, и скоро стало ясно, что фермеры не приобщились к новым методам. Правительство утвердило программу, предписывающую сельскохозяйственным агентам направляться на выбранные фермы и выращивать зерновые культуры вместе с фермерами, используя новые методы. Соседние фермеры, наблюдавшие увеличение урожая, немедленно приобщались к новым методам.

2. Длинные списки ожидания в клиниках по обнаружению рака растягивались на месяцы, и так было, начиная с осени 1974. Такая ситуация сложилась не из-за выпуска новой статистики главным хирургом страны, Американской медицинской ассоциацией или любой другой организацией. Появление длинных списков ожидания совпадает по времени с мастектомией, которую перенесли госпожа Форд и госпожа Рокфеллер.

3. Тимоти Краус, в своей книге, описывающей освещение в прессе предвыборной кампании президента в 1972, под названием *Мальчики в Автобусе* (1974), сообщила, что накануне выборов большая группа репортеров, освещающих кампанию МакГоверна, согласилась, что он не может проиграть больше чем на 10 пунктов. Эти люди были репортерами из агентства новостей, телевидения и главных газет и журналов новостей. Они знали, что во всех опросах МакГоверн отставал на 20 пунктов, и они знали, что за предыдущие 24 года ни один опрос не расходился с итогами выборов более чем на 3%. Однако, они видели своими собственными глазами, как восторженные толпы людей приветствуют МакГоверна.

4. *The New York Times* (Kaufman, 1973) недавно взял интервью у граффитиста метро в Нью-Йорке, который сильно обгорел в пожаре, вызванном искрой, которая подожгла его банки с распыляющейся краской. Мальчик, известный под псевдонимом "Али", признавался, что за 2 недели до несчастного случая он читал о мальчике по имени Бернанд Браун, которого задавило насмерть, когда он рисовал граффити на поездах. "Возможно, если бы мы знали имя, которое он использовал, например, 'Джо 146', это произвело бы впечатление, – сказал он, но я помню, как смеялся над этим, думая, что он, наверное, был какой-то пижон, который не знал, что делал..."

Мы полагаем, что настоящее исследование и примеры, взятые из повседневной жизни показывают, что некоторые виды информации, которую ученые рассматривают как подходящую и наводящую на выводы, обычно игнорируются людьми. Другие виды информации, которые являются логически слабее, наталкивают на размышления и приводят к действиям. Мы можем думать о ставшей бесполезной деятельности психологов, которые больше занимаются изучением обработки информации, чем обнаружением того, что их испытуемые расценивают как информацию, достойную обработки.

8. Причинные схемы при принятии решений в условиях неопределенности*

Амос Тверски и Даниель Канеман

Многие решения, которые мы принимаем, как в незначительных, так и в важных вопросах, зависят от очевидной вероятности событий, таких как выполнение обещания, успех предприятия или реакции на действие. Так как у нас нет адекватных формальных моделей для вычисления вероятности подобных событий, их оценка обязательно субъективна и интуитивна. То, как люди оценивают свидетельства, чтобы оценить вероятности, пробудило большой исследовательский интерес в недавние годы, например, В.Эдвардс (W.Edwards, 1968, 25); Канеман и Тверский (Kahneman Tversky, 1979a, 30); Словик (Slovic, 1972a); Словик и Фишхофф (Slovic FishHoff) и Лихтенштейн (Lichtenstein, 1977); Тверский и Канеман (Tversky и Kahneman, 1974, 1). Это исследование выделило несколько эвристик принятия решений, которые связаны с характерными ошибками и предубеждениями. Эта часть книги посвящена роли причинно-следственного рассуждения при принятии решений в условиях неопределенности и некоторым предубеждениям, которые связаны с этим способом мышления.

С психологической точки зрения банально то, что люди стремятся достигнуть последовательной интерпретации событий, которые окружают их, и что для достижения этой цели служит организация событий с помощью схем причинно-следственных отношений. Классическая работа Мичетта (Michotte, 1963) ярко продемонстрировала тенденцию воспринимать последовательности событий в терминах причинно-следственных отношений, даже когда человек полностью осознает, что отношение между событиями случайно и что приписанная причинно-следственная связь является иллюзорной. Распространенность причинно-следственных схем при восприятии элементарных социальных связей была выдвинута на первый план в 1958

* Данная глава – первая часть книги, которая появилась в редакции М. Фишбайн, Прогресс в социальной психологии. Hillsdale, N.J.: Lawrence Erlbaum Assoc., Inc, 1980. Переиздано в соответствии с разрешением.

конструктивной работе Хайдера (Heider), а исследование причинно-следственного приписывания стало одним из центральных в современной социальной психологии (Jones и другие, 1971; Ross, 1977).

В этой главе описывается роль причинно-следственных схем при принятии решений в условиях неопределенности. В частности, мы исследуем оценки условной вероятности $P(X/D)$ некоторого целевого события X , на основе некоторого свидетельства или данных D . Для психологического анализа воздействия свидетельства, полезно различать типы отношений, которые испытуемый может увидеть между D и X . Если D воспринимается как причина возникновения или невозникновения X , мы считаем D *причинно-следственными* данными. С другой стороны, если X воспринимается как возможная причина D , мы рассматриваем D как *диагностические* данные. Например, считается, что описание личности A обеспечивает причинно-следственные данные для прогнозирования его поведения, в то время как описание поведения A обеспечивает диагностическую информацию относительно его личности. Если D не является ни причиной, ни результатом X , но они оба воспринимаются как последствия другого фактора, мы считаем D показательными данными. Таким образом, поведение в одной ситуации обеспечивает показательные данные для поведения в другой ситуации, если обе манеры поведения расценены как проявления одной и той же черты. Наконец, если оказывается, что D и X , не связаны прямой или косвенной причинной связью, мы называем D *случайными* данными.

При нормативном рассмотрении теории условной вероятности, различия между типами отношения D к X незначительны, и воздействие данных зависит исключительно от их информативности. Напротив, мы предполагаем, что психологическое воздействие данных зависит от их роли в причинно-следственной схеме. В частности, мы выдвигаем гипотезу, что причинно-следственные данные имеют большее воздействие, чем другие данные такой же информативности; и что в присутствии данных, порождающих причинно-следственную схему, случайные данные, не соответствующие этой схеме, имеют маленькое значение, или вовсе его не имеют.

В этой части книги мы сравниваем влияние причинных и диагностических данных и показываем, что люди приписывают большее воздействие причинным, чем диагностическим данным равной информативности. Мы также исследуем класс проблем, где отдельно взятые данные имеют как причинно-следственное, так и диагностическое значение, и демонстрируем, что интуитивные оценки $P(X/D)$ зависят от прямого причинного воздействия D на X без учета диагностических соображений....

Причинное и диагностическое умозаключение

Логически выведенные асимметрии

Причинно-следственная схема носит естественный ход событий; он развивается от причин к следствиям. Мы предполагаем, что легче и более естественно следовать нормальной последовательности и рассуждать от причины к следствию, чем обратить эту последовательность и рассуждать от следствий к причине. Если причинно-следственные выводы действительно более легкие и естественные, чем диагностические выводы, то можно ожидать, что люди выведут результаты из причин с большей достоверностью, чем причины из результатов – даже, если результат и причина фактически дают одинаковое количество информации относительно друг друга. Мы проверили эту гипотезу, используя две различные величины: оценки условной вероятности и достоверность точности предсказаний.

В одном наборе вопросов мы попросили испытуемых сравнить две условные вероятности $P(Y/X)$ и $P(X/Y)$ для пары событий X и Y таких, что (1) X естественно рассматривается как причина Y ; и (2) $P(X) = P(Y)$, то есть предельные вероятности двух событий равны. Последнее условие подразумевает что $P(Y/X) = P(X/Y)$. Мы сделали прогноз, что большинство испытуемых сочтет причинно-следственное отношение более сильным, чем диагностическое, и ошибочно утвердит, что $P(Y/X) > P(X/Y)$.

В другом наборе вопросов мы попросили испытуемых сравнить достоверность прогнозов с использованием двух непрерывных переменных: зависимость от одной была дана, а другую необходимо было спрогнозировать. Как и ранее, задачи построены таким образом, чтобы одна из переменных рассматривалась как причинная относительно другой. Если две переменные обладают подобными предельными распределениями, не существует надежной статистической причины ожидать различие в точности, с которой одна переменная может быть спрогнозирована исходя из другой. Однако мы выдвинули гипотезу, согласно которой многие испытуемые посчитают, что прогнозирование от причины к результату может быть сделано с большей достоверностью, чем предсказание от результата к причине.

Прогнозируемая асимметрия между причинными и диагностическими выводами наблюдалась с двумя типами вопросов. Эффект проиллюстрирован следующими задачами, где числа в круглых скобках указывают количество студентов (Университета штата Орегон), которые выбрали каждый ответ.¹

Задача 1: Какое из следующих событий является более вероятным?

(А), что у девочки голубые глаза, если у ее матери голубые глаза. (N = 69)

¹ В более раннем варианте этой книги была приведена неправильная - низкая пропорция ответов “поровну”, основанных на выборке из Израиля. Мы обязаны Анне М. Б. Гонзалес и Майклу Кубовому за это наблюдение. Настоящие данные, собранные в Университете штата Орегон, совпадают с результатами, полученными Гонзалес в Йельском университете.

- (В), что у матери голубые глаза, если у ее дочери голубые глаза. (N = 21)
(-) два события одинаково вероятны. (N = 75)

Задача 2: При осмотре старшеклассников в некотором городе, рост мальчиков сравнили с ростом их отцов. Какой из прогнозов является более достоверным?

- (А) прогнозирование роста отца по росту сына. (N = 23)
(В) предсказание роста сына по росту отца. (N = 68)
(-) равная достоверность. (N = 76)

Очевидно, что распределение роста или цвета глаз – по существу одно и то же в следующих друг за другом поколениях. Чтобы проверить восприятие этого факта, мы спросили другую группу из 91 испытуемых, является ли пропорция матерей с голубыми глазами в большой выборке матерей и дочерей большей (N = 15), равной (N = 64) или меньшей (N = 12), чем пропорция дочерей с голубыми глазами. Таким образом, хотя испытуемые оценили предполагаемые вероятности как равные, они оценили “причинную” условную вероятность выше, чем “диагностическую”.

Строго говоря, рост отца не обуславливает рост его сына. В обиходе, однако, весьма приемлемо говорить, что мальчик высок, потому что рост его отца – 6 футов 4 дюйма, в то время как утверждение, что рост отца – 6 футов 4 дюйма, потому что его сын высок, совершенно аномально. В общем, мы ожидаем асимметрию вывода относительно двух переменных всякий раз, когда первая объясняет вторую лучше, чем вторая объясняет первую. Для иллюстрации, рассмотрим следующие задачи, которые были представлены двум различным группам испытуемых:

Задача 3: Какое из следующих утверждений имеет больше смысла?

- (А) Том тяжелый, потому что он высокий. (N = 63)
(В) Том высокий, потому что он тяжелый. (N = 7)

Задача 4: Какое из предсказаний более достоверно?

- (А) прогнозирование роста человека исходя из его веса. (N = 16)
(В) прогнозирование веса человека исходя его роста. (N = 78)

Хотя рост и вес не считаются причинами друг друга, большинство испытуемых посчитало, что высокий рост человека лучше объясняет большой вес, чем наоборот, возможно, потому что прототипичный высокий человек обладает большим весом, в то время как прототипичный тяжелый человек не всегда высокий. Соответственно, большинство испытуемых выразило большую уверенность в прогнозировании веса человека исходя из его роста, чем в прогнозировании роста человека исходя из его веса. Такая асимметрия, безусловно, не может быть оправдана на статистических основаниях.

Задачи 3 и 4 свидетельствуют о том, что асимметрия вывода происходит даже при отсутствии прямой причинной связи между этими двумя переменными – если одна из них (например, рост) более естественно рассматривается как объяснение другой (например, вес). Следующие две задачи иллюстрируют случай, где две переменные рассматриваются как признаки или проявле-

ния некоторой основной черты. Одна из переменных, однако, обеспечивает более прямое проявление или более надежную меру основной черты. Мы ожидаем, что выводы от более сильного к более слабому признаку будут сделаны с большей достоверностью, чем выводы в обратном направлении.

Задача 5: Какое из следующих событий является более вероятным?

(А), то, что атлет выиграл десятиборье, если он выиграл первое состязание в десятиборье. (N=21)

(В), то, что атлет выиграл первое состязание в десятиборье, если он выиграл десятиборье. (N = 75)

(-) два события одинаково вероятны. (N = 70)

Задача 6: Два теста на интеллектуальные способности проводились в большой группе студентов: всестороннее испытание длительностью 1 час и 10-минутная сокращенная версия. Какой прогноз является более достоверным?

(А) прогнозирование результата студента в коротком тесте на основании его результата во всестороннем тесте. (N = 80)

(В) прогнозирование результата студента во всестороннем тесте исходя из его результатов в коротком тесте. (N = 47)

(-) равная достоверность. (N = 39)

Снова, правильный ответ “поровну” в обеих задачах. В задаче 5, предполагаемая вероятность того, что атлет победит в десятиборье равна $1/N$, где N – количество конкурентов. Этому также равна априорная вероятность того, что атлет победит в первом состязании. Следовательно, две условные вероятности должны быть равны. В задаче 6, стандартное предположение о линейных регрессиях влечет за собой точность в предсказании одного испытания исходя из другого. Ответы на обе задачи, однако, показывают явное предпочтение одному направлению прогнозирования над другим.

В задачах 5 и 6 речь идет о двух признаках одной и той же основной черты, которые отличаются по выраженности. Победа в десятиборье и победа в отдельном состязании – оба являются проявлениями спортивного превосходства, но первое обеспечивает более выраженный признак превосходства, чем последнее. Точно так же результаты тестов на уровень интеллекта отражают основную черту интеллекта, и всесторонний тест отражает более высокую степень выраженности этой черты, чем сокращенная версия. Результаты подтверждают гипотезу, что прогнозирование от более сильного признака к более слабому связано с большей достоверностью, чем обратное предсказание.

Асимметрии вывода, наблюдаемые в предшествующих задачах, связаны с асимметриями отношений приближенности, которые исследованы в работах Тверского (1977). Эмпирические исследования показывают, что оцененное подобие заметного объекта или прототипа менее заметному объекту или варианту меньше, чем подобие варианта прототипу. Например, насыщенный красный менее подобен бледно-красному, чем наоборот (Ропш, 1975), “хорошая” форма менее подобна “плохой” форме, чем наоборот, и подобие зна-

менитой страны (например, коммунистический Китай) менее знаменитой стране (например, Северная Корея) меньше, чем обратное подобие (Tversky, 1977). Асимметрии прогнозирования, видимо, следуют тому же правилу. Таким образом, мы воспринимаем сына как более подобного своему отцу, чем наоборот, и мы также приписываем свойства отца сыну с большей достоверностью, чем наоборот. Тот же самый процесс, поэтому, может лежать в основе, как асимметрии подобия, так и асимметрии вывода.

Причинная и диагностическая значимость свидетельства

В предыдущей главе описывалось, что влияние причинно-следственных данных на оцененную вероятность последствия больше, чем воздействие диагностических данных на оцененную вероятность причины. В настоящей главе исследуются вопросы, в которых свидетельство имеет, как причинное, так и диагностическое значение по отношению к целевому событию. Мы исследуем гипотезу, согласно которой люди склонны сосредотачиваться на причинном воздействии данных, касающихся будущего, и имеют тенденцию пренебрегать диагностическими выводами, касающимися прошлого. Сначала мы обсудим класс задач, в которых преобладание причинных соображений над диагностическими приводит к появлению непоследовательных и парадоксальных оценок вероятности. Следующая пара задач была представлена Туроффом (1972) в обсуждении метода прогнозирования с взаимным влиянием фактов.

Задача 7a: Какая из следующих двух вероятностей выше?

- (1) вероятность того, что в течение следующих пяти лет Конгресс примет закон о предотвращении ртутного загрязнения, если количество смертельных случаев, приписанных отравлению ртутью, в течение следующих пяти лет превысит 500 человек.
- (2) вероятность того, что в течение следующих пяти лет Конгресс примет закон о предотвращении ртутного загрязнения, если количество смертельных случаев, приписанных отравлению ртутью, в течение следующих пяти лет не превысит 500 человек.

Задача 7b: Какая из следующих двух вероятностей выше?

- (1) вероятность того, что количество смертельных случаев, приписанных отравлению ртутью, в течение следующих пяти лет превысит 500 человек, если Конгресс примет в течение следующих пяти лет закон о предотвращении ртутного загрязнения.
- (2) вероятность того, что количество смертельных случаев, приписанных отравлению ртутью, в течение следующих пяти лет превысит 500 человек, если Конгресс не примет в течение следующих пяти лет закон о предотвращении ртутного загрязнения.

Пусть C – событие, что в течение следующих 5 лет Конгресс примет закон о предотвращении ртутного загрязнения, и пусть D – событие, что в течение следующих 5 лет, количество смертельных случаев, приписанных отравле-

нию ртутью, превысит 500. Пусть \bar{C} и \bar{D} обозначают события, противоположные событиям C и D , соответственно.

Большинство испытуемых считают, что, скорее всего, Конгресс примет закон, ограничивающий ртутное загрязнение, если список жертв превысит 500 человек, чем если не превысит, то есть $P(C/D) > P(C/\bar{D})$. Большое количество испытуемых также оценило как менее вероятное событие, что список жертв достигнет 500 человек, если закон будет принят в течение следующих пяти лет чем, если не будет принят, то есть $P(D/C) > P(D/\bar{C})$. Эти оценки отражают причинные убеждения, что большой список жертв повлиял бы на более скорое принятие мер по предотвращению загрязнения, и что такая мера будет эффективной для предотвращения ртутного отравления. В выборке из 166 студентов, 140 выбрали модальный ответ на оба вопроса. Этот внешне верный образец суждений нарушает наиболее элементарные правила условной вероятности.

Ясно, $P(C/D) > P(C/\bar{D})$ подразумевает $P(C/D) > P(C)$. Кроме того, неравенство

$$P(C/D) = \frac{P(C \& D)}{P(D)} > P(C)$$

сохраняется, тогда и только тогда, когда $P(C \& D) > P(C)P(D)$, что верно тогда и только тогда, когда

$$P(C) < \frac{P(C \& D)}{P(C)} = P(D/C)$$

что в свою очередь подразумевает $P(D/C) > P(D/\bar{C})$, если $P(C)$ и $P(D)$ не равны нулю. Следовательно, подразумевает $P(C/D) > P(C/\bar{D})$, вопреки $P(D/C) > P(D/\bar{C})$ преобладающему образцу суждений.

Легко построить дополнительные примеры подобного типа, в которых интуиция людей нарушает вычисление вероятности. Такие примеры состоят из пары событий, A и B , таких, что возникновение B увеличивает вероятность последующего возникновения A , в то время как возникновение A уменьшает вероятность последующего возникновения B . Например, рассмотрите следующую задачу.

Задача 8: Пусть A – событие, что до окончания следующего года, Питер установит систему сигнализации в своем доме. Пусть B – событие, что дом Питера будет обокраден до конца следующего года. Пусть \bar{A} и \bar{B} события, противоположные событиям A и B , соответственно².

Вопрос: Какая из двух условных вероятностей, $P(A/B)$ или $P(A/\bar{B})$, выше?

Вопрос: Какая из двух условных вероятностей, $P(B/A)$ или $P(B/\bar{A})$, выше?

² символы A , B , и т.д. введены, чтобы облегчить построение выводов. Испытуемым предлагали устные описания событий.

Подавляющее большинство испытуемых (132 из 161) заявило, что $P(A/B) > P(A/\bar{B})$ и что $P(B/A) < P(B/\bar{A})$, вопреки законам вероятности. Мы интерпретируем этот образец суждений как другой признак господства причинных соображений над диагностическими. Чтобы оценить природу этого эффекта, давайте проанализируем структуру задачи 8.

Для начала рассмотрим $P(A/B)$, условную вероятность того, что Питер установит сигнализацию в своем доме до конца следующего года, учитывая, что его дом будет обокраден в течение этого периода. Сигнальная система могла быть установлена либо до, либо после ограбления. Информация, переданная этим условием, то есть предположением о краже, имеет причинное значение относительно будущего и диагностическое значение относительно прошлого. Определенно, кража является причиной для последующей установки сигнализации, что, в свою очередь, является диагностическим признаком того, что дом не был оборудован ею во время кражи. Таким образом, причинное воздействие кражи увеличивает вероятность установления сигнальной системы, в то время как диагностическое воздействие кражи уменьшает эту вероятность. Почти единодушные оценки, что $P(A/B) > P(A/\bar{B})$ указывают, что причинное воздействие B преобладает над его диагностическим воздействием.

Такой же анализ применяется к $P(B/A)$: вероятности, что дом Питера будет обокраден до конца следующего года, учитывая, что он установит сигнализацию в течение этого периода. Наличие сигнализации причинно влияет на сокращение вероятности последующей кражи; это также дает нам диагностический признак того, что кража могла побудить Питера установить сигнализацию.

Причинное влияние приобретения сигнализации уменьшает вероятность кражи; диагностическое влияние установки сигнализации увеличивает эту вероятность. В данном случае, распространенность суждения, что $P(B/A) < P(B/\bar{A})$ указывает, что причинное влияние A преобладает над его диагностическим влиянием. Вместо сравнения причинных и диагностических воздействий свидетельства, люди, очевидно, оценивают условные вероятности $P(A/B)$ и $P(B/A)$ прежде всего в терминах прямого причинного эффекта состояния, что приводит к противоречиям в задачах этого типа.

Существенная особенность задач Туроффа (Turoff) – неопределенное временное отношение между событием, создающим условие и целевым событием. Даже при отсутствии временной неопределенности, часто создающее условие событие имеет как причинное, так и диагностическое значение. Проведенный анализ приводит к гипотезе, что на оценки условных вероятностей влияют причинно-следственные соображения, даже когда временное отношение между событиями полностью определено.

Задача 9: Какая из следующих вероятностей выше?

$P(R/H)$ вероятность того, что в течение 90-ых годов станет необходимым нормирование топлива для частных потребителей в США, учитывая, что в течение 80-ых произой-

дет значительное увеличение использования солнечной энергии для отопления домов. $P(R/\bar{H})$. Вероятность того, что в течение 90-ых станет необходимым нормирование топлива для частных потребителей в США, учитывая, что в течение 80-ых не произойдет никакого увеличения использования солнечной энергии для отопления домов.

Поучительно рассмотреть нормативный подход (Байеса) к этой проблеме, с точки зрения различия, которое мы наблюдали между причинными и диагностическими соображениями. Событие H , что в течение 80-ых произойдет значительное увеличение использования солнечной энергии для отопления домов, имеет, как причинное, так и диагностическое значение. Прямое причинное воздействие H на R является отрицательным. При равном значении других параметров, значительное увеличение в использовании солнечной энергии может только облегчить топливный кризис в более поздние годы. Однако значительное увеличение в использовании солнечной энергии в течение 80-ых также является весомым признаком надвигающегося энергетического кризиса. В частности, это говорит о том, что цены на топливо в 80-ых достаточно высоки, чтобы производить инвестиции в использование солнечной энергии для отопления домов, экономичного для большого количества потребителей. Высокие топливные цены в 80-ых, в свою очередь, предполагают нехватку ископаемого топлива, что увеличивает вероятность нормирования топлива в последующем десятилетии. Таким образом, прямое причинно-следственное воздействие H на R уменьшает вероятность R , принимая во внимание, что диагностические значения H косвенно увеличивают вероятность R .

Хотя вопрос относительного значения этих факторов не может быть разрешен формально, мы утверждаем, что диагностические значения H могли перевесить его причинное воздействие. Количество топлива, которое может быть сэкономлено путем увеличения использования солнечной энергии для отопления домов, вряд ли, будет достаточно большим, чтобы предотвратить надвигающийся кризис. С другой стороны, дефицит топлива, который подразумевает событие H , является явным показателем предстоящего энергетического кризиса. Согласно этой линии рассуждения, $P(R/H) > P(R/\bar{H})$, где \bar{H} – событие, противоположное H .

Гипотеза, выдвинутая в этой главе, однако, состояла в том, что люди вообще придают больше значения прямому причинному влиянию события, создающего условия, в оценках условных вероятностей и не уделяют достаточного внимания его диагностическому значению. В задаче 9, эта гипотеза предполагает, что увеличение использования солнечной энергии для отопления в 80-ых должно уменьшить оцененную вероятность нормирования топлива в 90-ых. Действительно, 68 из 83 испытуемых заявили, что $P(R/H) < P(R/\bar{H})$. Тот же самый образец суждений наблюдается в других задачах этого типа, где косвенные диагностические значения условия противоречат его прямым причинным значениям. Хотя этот образец суждений не нарушает правил вероятности, как в задачах Туроффа, он отражает, по

нашему мнению, общую тенденцию пренебрегать диагностическим значением события, создающего условия, в оценках условной вероятности.

Прогнозы, объяснение и пересмотр

В предшествующих главах мы представили некоторые факты в поддержку гипотезы, что причинные выводы имеют большую эффективность, чем диагностические выводы. Во-первых, мы показали, что выводы от причин к следствиям делаются с большей достоверностью, чем выводы от следствий к причинам. Во-вторых, мы показали, что, когда одни и те же данные имеют, как причинное, так и диагностическое значение, первому обычно придается больше значения, чем последнему в оценках условной вероятности.

Мы обратимся теперь к более общему вопросу соотношения между образом, моделью или схемой системы, например, энергетической ситуации или личности человека и некоторого исхода или проявления этой системы, например, увеличение использование солнечной энергии или проявление враждебности. Модели или схемы обычно используются, чтобы спрогнозировать и объяснить исходы, которые в свою очередь служат для модификации или модернизации модели. Таким образом, человек может применять модель, чтобы прогнозировать исход или оценивать его вероятность; он может также использовать модель, чтобы объяснить возникновение определенного события или последствия. Наконец, он может использовать информацию: полученную при возникновении определенного события, при исправлении или пересмотре модели.

Предсказание и объяснение представляют собой два различных типа причинного вывода, в то время как пересмотр модели – это пример диагностического вывода. При прогнозировании испытуемый выбирает тот исход, который наиболее соответствует модели системы. При объяснении, испытуемый определяет те особенности модели, которые, скорее всего, вызовут указанный исход. С другой стороны, при пересмотре модели, испытуемый исправляет или дополняет элементы модели, которые наименее соответствуют данным.

Большинство выводов в повседневной жизни опираются на модели или схемы, которые являются неточными, неполными и иногда неправильными. Люди признают это и часто стремятся подтвердить, что их модели систем, такие, как намерения человека или энергетическая ситуация, ошибочны. Присутствие неопределенности в точности модели имеет значение для надлежащего прогнозирования, объяснения и пересмотра. Если в модели встречаются ошибки, то прогнозы, исходящие из этой модели должны быть умеренны или регрессивны, то есть они не должны отклоняться от прогнозирования базового значения. Например, следует больше воздерживаться от прогнозов, что человек будет вести себя нехарактерным или необычным образом, если информация об этом человеке исходит из ненадежного источника, чем если та же информация взята из более правдоподобного источника.

Объяснения, которые основаны на неопределенных моделях, должны использоваться с осторожностью, так как причинные факторы, которые используются в объяснении, могут не существовать в действительности. Кроме того, объяснение в условиях неопределенности должно всегда осуществляться с пересмотром модели. Например, если человек занимается деятельностью, которая кажется несовместимой с нашим представлением о его личности, мы должны серьезно рассмотреть возможность того, что наше представление было неправильным, и что оно должно быть пересмотрено в свете новых данных. Чем больше неопределенность модели и чем более удивительно поведение, тем серьезнее должен быть пересмотр. Адекватное объяснение должно принимать во внимание изменения в модели, которые подразумеваются или предложены событием, требующим объяснения. С нормативной точки зрения, объяснение в условиях неопределенности относительно модели включает как диагностические, так и причинные выводы.

Предыдущее исследование показало, что люди обычно “преувеличенно прогнозируют” исходя из высоко неопределенных моделей. Например, испытуемые уверенно прогнозируют профессиональный выбор или академическую успеваемость человека на основе краткого описания его личности, даже если это описание взято из ненадежного источника (Kahneman и Tversky, 1973, 4). Намерения и черты, выведенные из краткого описания личности, обычно рассматриваются как причины таких исходов, как профессиональный выбор или успех в учебе. “Преувеличенное прогнозирование”, которое наблюдается в таких задачах, совместимо с высоким влиянием причинно-следственных данных, что было проиллюстрировано в предыдущих главах.

В контексте объяснения и пересмотра, преобладание причинного и недостатка диагностического рассуждения – очевидны благодаря большой легкости, с которой люди закладывают причинные основания для исходов, которые они не могли спрогнозировать, и благодаря трудностям, с которыми они сталкиваются при пересмотре неопределенных моделей для размещения новых данных. Оказалось, что легче приспособить новый факт к существующей причинной модели, чем пересмотреть модель в свете этого факта. Кроме того, пересмотры, которые сделаны с целью размещения новых фактов, часто минимальны по объему и локальны по характеру.

Чтобы проиллюстрировать это предположение, мы обратимся к наблюдениям из более раннего исследования интуитивного прогнозирования (Kahneman и Tversky, 1973, 4), о которых ранее не сообщалось. В этом исследовании, 114 аспирантам-психологам предоставили описание личности аспиранта Тома В., предположительно составленное психологом-клиницистом, когда Том был старшекласником, на основе проективных тестов. Давалось следующее описание:

Том В. обладает высоким интеллектом, но ему не хватает творческого потенциала. Он любит порядок и ясность и нуждается в системе, где каждая мелочь находится на своем

месте. Манера письма — однообразная и механическая, иногда оживляется несколько банальными каламбурами и вспышками воображения подобно произведениям научной фантастики. Имеет сильную тягу к устойчивому материальному положению. Проявляет мало сочувствия и симпатии к другим людям, не любит взаимодействовать с другими. Будучи эгоцентричным, тем не менее, имеет глубокое чувство морали.

Испытуемых сначала попросили предсказать область специализации Тома В., оценивая девять возможностей в терминах их вероятности. Большинство испытуемых пришло к согласию, что Том, скорее всего, займется информатикой или будет инженером, и менее вероятно, что он будет специализироваться в общественных дисциплинах и социальной работе или в гуманитарных науках и педагогике. Ответ на дополнительный вопрос также показал единодушие в том, что проективные тесты не являются надежным источником информации для прогнозирования выбора профессии. После завершения задачи предсказания, испытуемым задали следующий вопрос.

На самом деле, Том В. - аспирант на факультете педагогики, и он зарегистрирован в специальной программе подготовки преподавателей для обучения детей-инвалидов. Пожалуйста, сформулируйте очень кратко теорию, которая, как Вы полагаете, наиболее вероятно объясняет отношение между личностью Тома В. и его выбором карьеры.

Каков надлежащий подход к этому вопросу? Испытуемых поставили перед очевидным конфликтом между свершенным фактом, выбор карьеры Томом В., и детальным, но ненадежным описанием его личности. Достоверность, с которой люди прогнозируют выбор профессии исходя из описания личности, подразумевает веру в корреляцию между личностью и выбором профессии. Это мнение, в свою очередь, подразумевает, что выбор профессии диагностичен по отношению к личности. В вышеупомянутом примере, выбор профессии Томом В. маловероятен ввиду описания его личности, и это описание взято из сомнительного источника. Поэтому разумный диагностический вывод должен привести к существенному пересмотру образа Тома В., чтобы привести его в большее соответствие со стереотипом выбранной профессии. Если полагать, что студенты, получающие специальное образование, в целом склонны к состраданию, то выбор профессии Томом В. должен вызвать сомнения относительно того, что он “проявляет мало сочувствия и симпатии к другим людям”, как заявлено в сообщении психолога. Адекватный ответ на задачу должен, по крайней мере, допустить возможность, что личность Тома В. отличается от описания, и что он является на самом деле добрее и гуманнее.

Наши испытуемые не следовали этому подходу. Только незначительное меньшинство (21%) упомянуло какие-либо оговорки относительно надежности описания. Подавляющее большинство испытуемых, включая скептиков, решило конфликт либо путем выбора подходящих аспектов описания личности Тома В. (например, его глубокое чувство морали), либо путем новой интерпретации психологической значимости его выбора (например, как выражение потребности в господстве).

Можно поспорить, что неспособность наших испытуемых пересмотреть образ Тома В. просто отражает характеристики задачи, которая была им предоставлена, а именно “объяснять отношение между личностью Тома В. и выбором карьеры.” Согласно этой точке зрения, задача интерпретируется как призыв к попытке соотнести выбор профессии Тома В. с *описанием* его личности, не подвергая сомнению его валидность. Мы считаем, однако, что распространенная тенденция рассматривать образ Тома В., как полностью достоверный, несмотря на серьезные сомнения, иллюстрирует более широкое явление: тенденцию объяснять без пересмотра, даже, если модель, которая используется в объяснении, высоко неопределенна.

По нашему представлению, ответы испытуемых иллюстрируют, как нежелание пересмотреть подробную и последовательную модель, пусть и неопределенную, так и непринужденность, с которой такая модель может использоваться для объяснения новых фактов, пусть и неожиданных. Нас поразила легкость, с которой наши испытуемые строили причинные соображения по поводу неожиданного выбора профессии Томом В., и у нас нет причин полагать, что они были бы менее поверхностны в объяснении других неожиданных действий с его стороны.

Высоко развитые навыки к объяснению способствуют общеизвестной устойчивости и стабильности впечатлений, моделей, концепций и парадигм при столкновении с неординарным свидетельством (Abelson, 1959; Novland, 1959; Janis, 1972; Jervis, 1975; Kuhn, 1962). Стимул для пересмотра модели может исходить только из признания несоответствия между этой моделью и некоторым новым свидетельством. Если люди могут объяснять большинство явлений к своему собственному удовлетворению при минимальных изменениях локального характера в своих концепциях, они редко будут чувствовать потребность в решительном пересмотре этих концепций. Таким образом, легкость причинного размышления не допускает процесса диагностического пересмотра....

9. Недостатки процесса атрибуции: о происхождении и исправлении ошибочных социальных оценок*

Ли Росс и Крег А. Андерсон

Введение в теорию атрибуции и ошибки атрибуции

Теория атрибуции и интуитивная психология

Теория атрибуции, в самом широком смысле, рассматривает попытки обычных людей понять причины и значения событий, свидетелями которых они являются. Она анализирует их “наивную психологию” как способ интерпретировать собственное поведение и действия других. Таким образом, настоящее господство в социальной психологии теории атрибуции является кульминацией длительной борьбы за улучшение этой дисциплинарной концепции человека. Не подчиняясь больше закономерности «стимул–реакция» (S-R) радикального бихевиоризма, выйдя за границы информационного процессора и перестав быть искателем когнитивной последовательности, человек в психологии получил статус, равный статусу ученого, который его исследует. С точки зрения теории атрибуции, люди являются интуитивными психологами, которые стремятся объяснить поведение других и сделать выводы об их социальном окружении.

Чтобы лучше понимать восприятие и действия этого интуитивного ученого, мы должны исследовать методы, которые он использует. Сначала, подобно ученому-психологу, он руководствуется набором предположений относительно природы человеческого Я и его поведения – например, что стремление получить удовольствие и избежать боли – распространенные и сильные мотивы поведения, или что подчинение желаниям и ожиданиям других людей, более обычно и требует меньше дальнейшей интерпретации,

* В этой главе многое заимствовано, как по содержанию, так и по структуре, из статьи первого автора в книге “Advances in Experimental Social Psychology” (1977). Разрешение на использование этих материалов подтверждено.

чем неподчинение. Психолог-любитель, подобно профессиональному психологу, также всецело полагается на данные, хотя они редко соответствуют формальным требованиям случайности и репрезентативности.

Иногда эти данные из первых рук; чаще, они являются результатом неофициального общения, получены из средств массовой информации или других косвенных источников. Далее интуитивный психолог должен выбрать или разработать методы кодирования, хранения и восстановления данных. Наконец, он должен использовать различные стратегии для подведения итогов, анализа и интерпретации данных – то есть он должен использовать правила, эвристики или схемы, которые позволят ему формировать новые выводы. Способность интуитивного ученого исследовать свое социальное окружение, соответственно, будет зависеть от точности и адекватности его гипотез, доказательств и анализов. Наоборот, любые систематические ошибки в существующих теориях, предубеждения в доступных данных или несоответствия в методах анализа, влекут за собой серьезные последствия, как для психолога-любителя, так и для общества, которое он создает. Эти недостатки, исследуемые с точки зрения современной теории атрибуции, подробно анализируются в этой главе.¹

Основные принципы теории атрибуции были предложены Хайдером (Heider, 1944, 1958) и развиты более детально Джонсом и Девисом (Jones & Davis, 1965), Келли (Kelley, 1967, 1971, 1973) и их коллегами (см. Jones и другие, 1971; Weiner, 1974). В центре внимания этих ученых находились две родственные задачи, с которыми сталкивается ученый-любитель. Первая – задача причинных суждений: исследователь стремится определить причину или набор причин, к которым можно *приписать* отдельно взятый эффект (то есть, некоторое действие или исход). Вторая – задача социального предположения: наблюдатель события делает выводы относительно атрибуции или намерений действующих лиц, а также относительно признаков или свойств ситуаций, на которые они реагировали.

И причинные оценки, и задачи социального предположения были предметом интенсивного теоретического и эмпирического исследования и, до недавнего времени, фактически доминировали в теории атрибуции. В последнее время, однако, начала получать некоторое внимание третья задача интуитивного психолога – прогнозирование или оценка исходов и поведения. Интуитивный психолог должен не только искать объяснения и делать выводы о намерениях; он должен также формировать ожидания и высказывать предположения относительно действий и исходов, неизвестных в настоящее время или которые произойдут в будущем. Например, когда кандидат в президенты обещает «облегчить бремя среднестатистического налогоплательщика», мы рассматриваем возможные причины для этого заявления и делаем выводы относительно личных свойств кандидата. (Возмож-

¹ Для более полного и систематического объяснения параллели обыватель-ученый, советуем читателю обратиться к работе Нисбетта и Росса, 1980.

но, обещание просто отвечало требованиям политической целесообразности? Можем ли мы сделать какие-то выводы об истинных убеждениях кандидата?) Но мы также, вероятно, будем размышлять о последующем поведении этого кандидата и его взглядах, которые касаются еще не исследованных проблем. (Если его выберут, снизит ли он налог на прибыль с недвижимости? Одобрят ли он сокращение программ социального обеспечения?) Таким образом, психология интуитивного предсказания является естественным продолжением теории атрибуции.

Три задачи атрибуции взаимозависимы согласно логики и психологии. Объяснения события и выводы о людях, производящих действия, а также объектах, которые вовлечены в действие, глубоко связаны. Все вместе они обеспечивают основание для размышления относительно характера событий, которые являются в настоящее время неизвестными или, вероятно, произойдут в будущем. Каждая задача, однако, предлагает уникальные возможности (и уникальные задачи интерпретации и методологии; см. Ross, 1977, стр. 175-179) для раскрытия предположений и стратегий, которые лежат в основе поведения интуитивного ученого. Стоит отметить, что в настоящее время стало популярным использование оценок и прогнозов как зависимых переменных в исследованиях предположений обывателей. Одна из причин этой увеличившейся популярности особенно важна. Оценки или предсказания новых или неизвестных событий могут часто выполняться относительно их *точности* в отличие от причинных оценочных суждений, которые исходят из анализа события. То есть можно сравнивать прогнозы и оценки событий с фактическими наблюдениями или измерениями. Это позволяет делать оценку как относительной адекватности стратегии атрибуции, которую использует интуитивный ученый, так и направления отдельно взятых ошибок и предубеждений.

Логические принципы атрибуции или эгоцентрические предубеждения

Современная теория атрибуции преследовала две отличительных, но взаимодополняющих цели. Во-первых, продемонстрировать, что оценки и выводы человека согласуются с некоторой логической или рациональной моделью. Во-вторых, проиллюстрировать и объяснить источники предубеждений или ошибки, которые искажают эти, по большому счету, правильные оценки и выводы. Мы кратко рассмотрим так называемые логические или рациональные правила, используемые интуитивным психологом, а затем остаток главы посвятим источникам ошибок в его попытках понимания, прогнозирования и управления событиями, которые разворачиваются вокруг него.

Принципы “ковариации” и “дисконтирования”. Люди должны, в большинстве случаев, разделять общее понимание социальных действий и исхо-

дов, которые их затрагивают, поскольку без такого согласия, социальное взаимодействие было бы хаотичным, непредсказуемым и неуправляемым. Самоанализ, проведенный теоретиками в области атрибуции, подкрепленный лабораторными доказательствами, привел к утверждению набора “правил”, которые могут использоваться в интерпретации поведения и исходов. Эти правила “здорового смысла” или схемы аналогичны, в некотором смысле более формальным правилам и процедурам, которым следуют социологи и статистики в своем анализе и интерпретации данных.

Келли, Джонс и их помощники выделили два случая, в которых могут применяться логические правила или схемы. В случае *множественного* наблюдения приписывающий имеет доступ к поведенческим данным, которые могли бы быть представлены как строки или колонки матрицы ответов: *действующий субъект* × *объект* × *ситуация* (или случай). Как правило, в этой области исследований участникам даны скорее заключительные утверждения, чем фактические ответы. Таким образом, потенциальный приписывающий узнает, что “Большинству театралов нравится новая пьеса Пинтера”, или “Мэри очень жалеет беспризорных животных” или “единственная телевизионная программа, которую смотрит Энн – Masterpiece Theatre.” В случае *единичного* наблюдения приписывающий должен рассматривать поведение единственного человека в одном случае. Например, он может видеть, что Сэм исполняет просьбу экспериментатора нанести болезненный удар другому человеку, или он может узнать, что “Луи поставил все свои деньги в Pimlico.”

Логические правила или принципы, управляющие атрибуциями в этих двух случаях, довольно различны (Kelley, 1967, 1971, 1973). В случае *множественного* наблюдения приписывающий применяет “принцип ковариации”; то есть он оценивает степень, в которой наблюдаемое может происходить в отсутствие каждого рассматриваемого причинного фактора. Соответственно, приписывающий делает вывод, что новая пьеса Пинтера – хороша в той степени, что она нравится широкому кругу театралов, вызвала похвалу людей, которым нравится мало пьес (например, “критикам”), и что ей рукоплескали так же энергично на девяностый день ее постановки, как и на девятый.

В случае *единичного* наблюдения стратегия оценки приписывающего включает применение “принципа дисконтирования”, с помощью которого социальный наблюдатель “обесценивает” роль любого причинного фактора в объяснении события настолько, насколько другие вероятные причины или детерминанты могут быть идентифицированы. Этот принцип атрибуции может быть по-новому сформулирован скорее в терминах социальных выводов, чем причинных атрибуций: насколько ситуативные или внешние факторы предоставляют “достаточное” объяснение события, настолько это событие приписано ситуации, и логически невозможно сделать никакого вывода (возможно, и опытным путем) относительно поведения человека. Наоборот, в какой степени действие или исход происходит вопреки, а не

благодаря сопутствующим ситуативным силам, настолько соответствующее событие приписывается человеку и делается “соответствующий вывод” (Jones и Davis, 1965), то есть приписывающий выводит существование и влияние некоторой черты, способности, намерения, чувства или другого свойства, которое могло объяснить действие актора или исход. Таким образом, мы не согласны с выводом, что крайне рискованная азартная игра Луи в Pimlico настолько отражала его устойчивые личные качества, что необходимо упомянуть такие факторы, как хорошие чаевые, глубокий финансовый кризис или то, что до игры он выпил семь рюмок мартины. С другой стороны, мы сочтем Луи неисправимым азартным игроком, если узнаем, что он играл, несмотря на угрозу жены уйти от него, если он когда-нибудь снова потратит зарплату на скачках, несмотря на то, что он знал, что не сможет оплатить жилье, если проиграет, и, несмотря на подслушанное замечание эксперта, что фаворит на скачках проявляет себя «даже лучше, чем согласно предварительным оценкам».

Стоит отметить, что применение двух различных принципов предъявляет различные требования к интуитивному ученому. Принцип ковариации требует, чтобы приписывающий применял правила, которые являются логическими или статистическими по своей природе, и не требует никакого дальнейшего понимания характеристик рассматриваемых объектов. Применение принципа дисконтирования, наоборот, требует значительного понимания природы человека и таких ситуативных сил, как потребность в деньгах, потребление алкоголя и угроза супруги развестись. В некотором смысле, принцип ковариации может применяться простым “статистиком”, в то время как принцип дисконтирования требует знаний психологии, чтобы оценить роль различных социальных воздействий и ситуативных сил и даже отличать преднамеренные действия и исходы от непреднамеренных (ср. Jones и Davis, 1965).

Доказательство систематического использования принципов атрибуции здравого смысла можно обнаружить, прежде всего, в исследовании с применением анкетного опроса, в котором испытуемые читали и интерпретировали краткие рассказы о реакциях одного или нескольких людей на некоторые объекты при указанных обстоятельствах (например, L. Z. McArthur, 1972, 1976). Проведенные по этому поводу исследования меньшего масштаба предоставили приписывающему, по-видимому, подлинные ответы, события и исходы (например, Jones, Davis и Gergen, 1961; Jones и DeCharms, 1957; Jones и Harris, 1967; Strickland, 1958; Thibaut и Riecken, 1955). Это исследование показало, что приписывающие могут использовать, по крайней мере, некоторые гипотетические принципы или практические методы. То, что методологии, используемые до настоящего времени, оставляли без внимания, так это *степень* точности обывателей, а также величину и направление их ошибок.

Эгоцентричные мотивационные предубеждения в атрибуции. В размышлении относительно возможных искажений в другой логической системе атрибуции, теоретики быстро сформулировали предубеждения “эго-защиты”, с помощью которого приписывающие сохраняли или усиливали чувство собственного достоинства или положительное мнение о своем поведении и способностях (Heider, 1958; Jones и Davis, 1965; Kelley, 1967). Попытки доказать существование такого мотивационного предубеждения включало проявление асимметрии в атрибуции положительных и отрицательных исходов, в особенности, тенденция для действующих лиц приписывать “успехи” своим собственным усилиям, способностям или поведению, а “ошибки” – удаче, трудности задач или другим внешним факторам. Задачи на успеваемость (например, Davis и Davis, 1972; Feather, 1969; Fitch, 1970; Wolosin, Sherman и Till, 1973) и обучающее поведение (например, Beckman, 1970; Freize и Weiner, 1971; Johnson, Feigenbaum и Weiby, 1964) предоставили яркое свидетельство этой асимметрии. Также было показано, что действующие лица могут больше хвалить себя за успех и меньше обвинять за неудачи, чем наблюдатель, оценивающий те же самые исходы (Beckman, 1970; Gross, 1966; Polefka, 1965).

Критики, скептически относящиеся к распространенным мотивационным предубеждениям, однако, испытали трудности в экспериментальном подтверждении этого (см. Miller, Ross, 1975, также Nisbett и Ross, 1980, см. гл. 10, для детального рассмотрения). Наиболее веские аргументы против исследования, показывающего мотивационные предубеждения – это очевидное несоответствие между личным восприятием испытуемых и их высказанными оценками. Можно без труда создать ситуации, в которых человек будет публично отрицать ответственность (или заявлять о ней) за случай, который он на самом деле принимает (или не принимает) как свою ответственность. В то время как высказанные оценки могут быть эгоцентричны в смысле сохранения общественного мнения о себе, они не подразумевают действие предубеждений эго-защиты для сохранения репутации в своих собственных глазах (Miller, 1978).

Кроме того, асимметрии в личных приписываниях успеха и неудачи (независимо от того, были ли они доступны исследователям) и различия в оценках самих людей и наблюдателей могут отражать другие немотивационные источники предубеждений. Как отметили некоторые исследователи, успех, по крайней мере, в экспериментальных ситуациях, ожидается и не противоречит прошлому опыту актора, в то время как неудача может быть непредвиденной и неожиданной. Точно так же успешные исходы преднамеренны и являются объектами планов и действий человека, в то время как, неудачи – это случайные события, которые возникают, несмотря на планы человека и его усилия. Наблюдатели, кроме того, редко знают о прошлом опыте или настоящих ожиданиях и намерениях акторов, свидетелями чьих результатов они являются.

Попытки опровергнуть существование глубоких предубеждений эго-защиты, были как эмпирическими, так и концептуальными. Так, испытуемые в некоторых исследованиях проявляют “контрзащитные” или уменьшающие оценку, предубеждения. Например, Росс, Бьербрауер и Полли (Ross, Bierbrauer и Polly, 1974), используя достоверную парадигму учитель – ученик, обнаружили, что учителя оценивали свои собственные действия и способности как более важные детерминанты скорее неудачи, чем успеха. И наоборот, учителя оценивали усилия и способности своих учеников как более важные детерминанты успеха, чем неудачи. В этом же исследовании эти на вид контрзащитные тенденции атрибуции признавались больше среди профессиональных преподавателей, чем среди неопытных студентов старших курсов, — результат, который противоречил очевидному выводу из теории эго-защиты: то, чему непосредственно угрожает неудача, должно быть более защищенным.

Исследователи, настаивающие на существовании эгоцентричных мотивационных предубеждений, могут, конечно, предоставить альтернативные интерпретации исследований, в которых не проявляются мотивационные или контрзащитные предубеждения (ср. Bradley, 1978). Действительно, во многих отношениях споры между сторонниками этой теории и скептиками стали напоминать более ранние и обширные споры о теории обучения и базового восприятия, в которых бесплодность поиска “решающего” эксперимента, касающегося мотивационных влияний (то есть, такого, который не может интерпретироваться “другой стороной”), стала более очевидной, чем во время накопления данных и усиления концептуального анализа.

Одной из ответных реакций на такое положение дел должен был стать временный отказ от мотивационных конструкций и сосредоточение внимания на немотивационных факторах (то есть, информационных и познавательных, а также факторах восприятия), которые влияют и потенциально искажают оценки атрибуции. Помимо существующих концептуальных трудностей, смешанных результатов, полученных опытным путем, и исторических уроков, препятствующих исследователям, которые искали бы мотивационные предубеждения, существуют две дополнительные причины для изменений в настоящий момент. Во-первых, все возрастает убеждение, что более серьезная оценка немотивационного влияния могла бы привести к пониманию и предвосхищению тех обстоятельств, в которых приписывания ответственности увеличили бы чувство собственного достоинства приписывающего, а также обстоятельств, в которых такие приписывания уменьшили бы его (ср. Miller и Ross, 1975). Во-вторых, все больше признается тот факт, что точные атрибуции вообще являются более “эгоцентричными”, чем неточные, то есть, что искажения причинного суждения способны приводить организм в состояние, *плохо* подготовленное к задаче долгосрочного выживания, какими бы благоприятными ни были немедленные последствия неточных восприятия и влияний.

Остальная часть этой главы рассматривает ограниченное количество таких немотивационных предубеждений (см. Nisbett и Ross, 1980, для более

полного обзора). В ней также обсуждается явление, которое увеличивает “удельный вес” таких предубеждений – сохранится тенденция ошибочных впечатлений, оценок и даже более обширных теорий, несмотря на наличие логически мощных данных, противоречащих этим убеждениям. Давайте признаем с самого начала, что ошибки и предубеждения, с которыми мы сталкиваемся вовсе не являются необъяснимыми погрешностями со стороны интуитивного ученого. Как правило, они отражают действие механизмов и стратегий, которые вполне хорошо удовлетворяют потребности организма во многих обстоятельствах; иначе они не пережили бы историю обучения человека или эволюционную историю вида. Эти ошибки и предубеждения могут справедливо быть расценены как “определенные сферы” сбой логически выведенных стратегий и тактик, которые целесообразны (и вероятно, также весьма точны) с точки зрения опыта организма в целом.

Немотивационные предубеждения атрибуции

Фундаментальная ошибка атрибуции

Идентифицированное первым (Heider, 1958) и наиболее часто упоминаемое немотивационное предубеждение, которое мы назовем *фундаментальной ошибкой атрибуции*, — это тенденция наблюдателей недооценивать воздействие ситуативных факторов и переоценивать роль диспозиционных факторов на поведение других. Будучи “*интуитивными*” психологами, мы, вероятно, слишком часто оказываемся нативистами или сторонниками индивидуальных различий, и слишком редко бихевиористами. С большой готовностью мы рассуждаем о чертах характера и ожидаем последовательности в поведении или исходах в совершенно несопоставимых ситуациях и контекстах. Мы делаем поспешные выводы после наблюдения поведения других людей, не замечая воздействия соответствующих сил окружающей среды и ограничений.

Общие факты, свидетельствующие о фундаментальной ошибке атрибуции. Помимо коротких рассказов и обращения к опыту, наиболее часто упоминаемое свидетельство для этого предубеждения (например, Jones и Nisbett, 1971; Kelley, 1971) заключается в очевидной готовности наблюдателей делать “*соответствующие*” личные выводы относительно участников исследования, реагирующих на явное давление ситуации. Например, Джонс и Харрис (Jones и Harris, 1967) обнаружили, что слушатели находили некоторое соответствие между замечаниями участников дискуссии в поддержку Кастро и их личными установками даже, если слушатели знали, что участники дискуссии повиновались экспериментатору, и у них не было выбора.

Более прямое доказательство того, что наблюдатели могут игнорировать или недооценивать давление ситуации, предоставил Бирбрауэр (Bierbrauer, 1973), когда изучал впечатления испытуемых от воздействия в классическом эксперименте Милграма (1963). В исследовании Бирбрауэра, участники были свидетелями дословного воспроизведения акта “повиновения” од-

ного из испытуемых, готового нанести максимальный разряд электрошока предполагаемой жертве. Независимо от того, почему и насколько произошла задержка в выставлении оценок, независимо от того, обладали они законным правом делать оценки или просто наблюдали, участники исследования Бирбрауэра допустили фундаментальную ошибку атрибуции. То есть они последовательно и серьезно недооценивали степень, в которой испытуемые уступили бы давлению ситуации, которое вызвало повиновение в эксперименте Милграма (см. Рисунок 1). Другими словами, они предполагали, что повиновение отдельно взятого испытуемого выражало скорее его личные характеристики, чем возможности ситуативных воздействий и ограничений, под влиянием которых оказались все испытуемые.

Особый случай преимущества, обусловленного ролью, при самопрезентации. Тенденция социальных наблюдателей недооценивать возможности влияния и ограничений ситуации и переоценивать роль индивидуальных черт часто наблюдалась в стратегии, концептуальных исследованиях и даже профессиональных спорах представителей современной социальной психологии (см. Nisbett и Ross, 1980; Ross, 1977). Особые случаи этой фундаментальной ошибки атрибуции помогают сосредоточить наше внимание на процессах-посредниках и более специфических ошибках интуитивного психолога. Важным с этой точки зрения является эксперимент Росса, Эмебайл и Стейнмец (Ross, Amabile и Steinmetz, 1977), в ходе которого анализировались оценки, сделанные относительно участников, роли которых (случайно назначенные) давали или не давали им преимущество. Росс и др. рассматривали роли ведущего и игрока в викторине по выявлению общих знаний. Роль ведущего обязывала испытуемого составить набор вопросов по общим знаниям, изложить эти вопросы игроку и обеспечивать точную обратную связь после каждого ответа игрока. Роль игрока была ограничена ответом или попыткой ответить на соответствующие вопросы. От двух участников (и, в последующем воспроизведении, также от наблюдателей) требовалось оценить общие знания как ведущего, так и игрока.

Произвольное назначение и выполнение этих ролей заставило участников и наблюдателей сталкиваться с очевидно нерепрезентативными или искаженными “выборками” знаний ведущих и игроков. Роль ведущего способствовала демонстрации скрытых знаний и давала возможность избежать ситуаций, показывающих их невежество; игроки же были лишены таких преимуществ в самопрезентации. Действительно, не было сомнения в произвольности назначения роли или в различии преимуществ, связанных с каждой ролью, в отличие от ситуаций реальной жизни, в которых социальные роли так же предоставляют преимущества и неудобства в самопрезентации. Однако соревнование в неравных условиях между ведущими и игроками последовательно вело к предубеждениям и ошибочным впечатлениям. Участники, в некотором смысле, просто не сумели сделать скидку на преимущества и неудобства соответствующих ролей, обусловленных ситу-

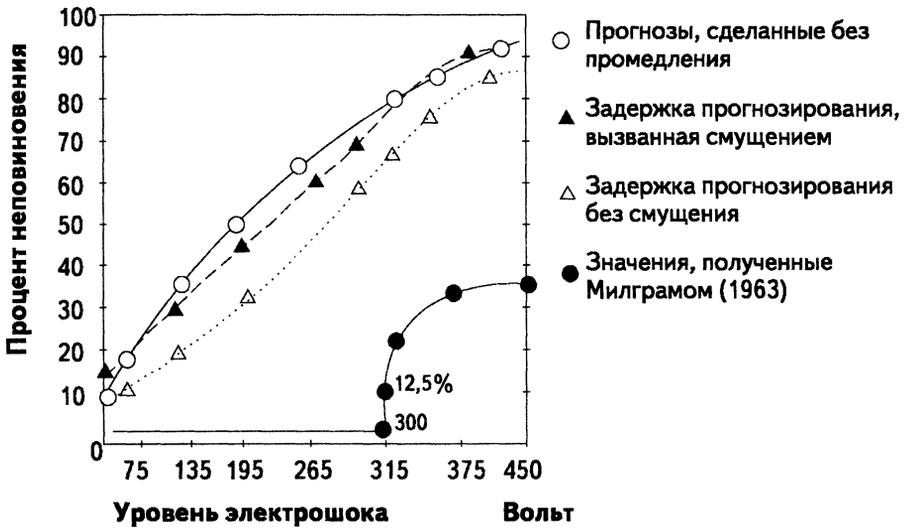


Рис. 1. Сравнение спрогнозированных и фактических значений неповиновения

ацией. Таким образом, игроки оценили своих ведущих выше себя, а независимые наблюдатели согласились с этим (см. Рисунок 2). Наблюдатели, вооруженные знанием, что ведущие не смогли бы ответить на заданные вопросы лучше, чем игроки, должны были признать, что уровень знаний игроков не был таким уж низким. Вместо этого, они сделали вывод, что ведущие обладали действительно выдающимися знаниями. Интересно, что самих ведущих не ввела в заблуждение сложившаяся ситуация. Этот факт смещает наше внимание с существования фундаментальной ошибки атрибуции и определенного воздействия социальных ролей к особым "выборкам данных", на которые полагались различные участники в своих выводах.

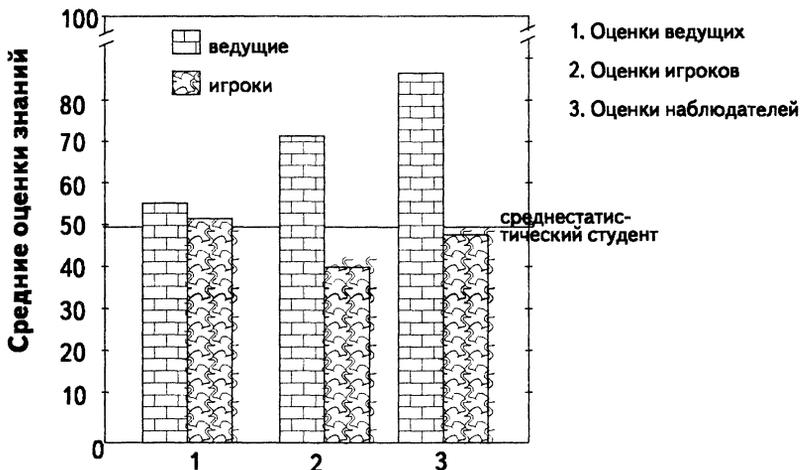


Рис. 2. Оценки общих знаний ведущих и игроков

В отличие от игроков и наблюдателей, ведущие не были вынуждены полагаться исключительно на “искаженные” выборки общих знаний. Возможно, они обладали рядом дополнительных доказательств степени и ограниченности собственных общих знаний, а также относительно нерепрезентативности той информации, которую они показали в своих вопросах; следовательно, они оценили, и свои знания, и знания игроков как “средние”.

И социальное, и теоретическое значение эксперимента Росса, Эмбайл и Стейнмец вполне понятно. Он побуждает нас рассматривать бесчисленные социальные контексты, в которых официальные или неофициальные роли ограничивают межличностное взаимодействие и, таким образом, искажают впечатления участников настолько, что кажутся оправданными привилегиями и ограничениями, наложенными выгодными или невыгодными ролями. Он также побуждает нас заострить внимание на одной из ошибок интуитивного ученого – его кажущейся нечувствительности к ограниченному значению искаженных выборок данных (см. также Hamill, Wilson и Nisbett, 1980; Nisbett и Ross, 1980, Гл. 4).

Предубеждения заметности и доступности

Возможно, наиболее исследованная область предубеждения атрибуции, которая включает эффекты внимания, а также когнитивных факторов и факторов восприятия, которые опосредуют внимание. Оказывается, что всякий раз, когда некоторый аспект окружающей среды становится непропорционально существенным или “доступным” для человека, воспринимающего его (ср. Tversky и Kahneman, 1973, 11), этому аспекту придают больше значения в причинно-следственном приписывании. Таким образом, когда участник визуально привлекает на себя внимание из-за особенного расового или сексуального статуса в пределах большей группы (Taylor и другие, 1976), из-за некоторой поразительной особенности внешнего вида или одежды (McArthur и Post, 1977; McArthur и Soloman, 1978), из-за поставленной учителем задачи (Regan и Totten, 1975), или даже из-за размещения людей на местах или других показателей визуальной перспективы (например, Storms, 1973; Taylor и Fiske, 1975), то на этого человека налагается непропорционально большая ответственность за любой результат его действий. (См. Taylor и Fiske, 1978, для более полного обзора.) Действительно, множество исследований, полученных на основе теории “объективного самосознания” (Duval и Wicklund, 1972; Wicklund, 1975) показали, что на восприятие людьми их собственных причинных ролей можно также повлиять с помощью простых манипуляций, которые направляют их внимание на себя как социальных объектов или наоборот (например, Duval и Hensley, 1976; Ellis и Holmes, 1979).

Признание и понимание того, как факторы заметности или доступности влияют на процесс атрибуции, могут помочь лучше понять основы многих знакомых предубеждений атрибуции, а также логически выведенных пре-

дубеждений, возможно даже отнеся их к особым случаям. Фундаментальная ошибка атрибуции, например, может отражать тот факт, что участники являются просто более заметными, чем условия окружающей среды и поэтому более вероятно, что они будут первоначально замечены при поиске причинных факторов. Действительно, когда ситуативные факторы и ограничения становятся непропорционально заметными для наблюдателя, мы можем ожидать ошибок атрибуции, которые кажутся противоположными так называемой фундаментальной ошибке. Таким образом, инспектор может ошибочно приписать действительно хорошее выполнение задания рабочим внешнему фактору, то есть своему наблюдению, когда этот внешний фактор становится чрезвычайно существенным (см. Strickland, 1958). Точно также, внутренний интерес человека при выполнении данной задачи может быть подорван (например, Deci, 1971; Lepper и Greene, 1975, 1978; Lepper, Greene, и Nisbett, 1973), если его внимание сосредотачивается на внешнем стимуле или ограничении, которое, как кажется, поощряет, но не обязательно, выполнение этой задачи.

Рассмотрим также эмпирическое обобщение Джонса и Нисбетта (Jones & Nibbett, 1971), согласно которому акторы, объясняя свое поведение, больше склонны сослаться на ситуативные факторы и менее – на факторы, относящиеся к характеру самого человека, чем наблюдатели такого поведения. Насколько участники и наблюдатели проявляют соответствующие различия в том, что находится в их центре внимания, то есть участники проявляют интерес к соответствующим особенностям окружающей среды, в то время как наблюдатели сосредотачивают свое внимание непосредственно на участниках, – настолько обобщение Джонса и Нисбетта становится особым случаем обобщения внимания/атрибуции. Действительно, экспериментально подтверждено что, перемещая центр внимания акторов и наблюдателей, их тенденции сослаться на ситуативные или на внутренние причины, могут аналогично меняться (например, Storms, 1973; Taylor и Fiske, 1975).

Наконец, рассмотрим недостаток, отмеченный пронизательным вымышленным детективом-психологом Шерлоком Холмсом – тенденцию недооценивать значение не-происшествий (non-occurence). Информативные не-происшествия – это события или действия, которые *не* произошли в некотором контексте, но которые, таким образом, содержат потенциально важную информацию (ср. Росс, 1977). Определенная релевантность низкой когнитивной доступности не-происшествий должна быть очевидна. При исследовании или рассмотрении причинных факторов, не-происшествия вряд ли являются высоко заметными или вряд ли им уделяется внимание; соответственно, невозможно придать им достаточно значения в объяснении наблюдаемых действий и исходов. Джилл, скорее всего, припишет гнев Джека чему-то, что она «совершила», чем чему-то, чего она не смогла сделать, просто потому, что первое является более заметным для нее, чем последнее. Действительно, принимая во внимание, что вина упущения является менее су-

пещественной, чем вина совершения поступка, Джек способен допустить ту же самую ошибку в объяснении причин собственного гнева.

Ложный консенсус или эгоцентрическое предубеждение атрибуции

Последнее немотивационное или “информационное” предубеждение, рассматриваемое в этой главе, касается оценок человеком социального согласия – кажущаяся обычность или необычность различных реакций, которые они наблюдают. В отличие от профессионального психолога, который полагается на четкие методы формирования выборки и статистические процедуры для проведения таких оценок, обыватель, должен полагаться на интуицию и субъективные впечатления, основанные на соответствующих данных, доступ к которым у него ограничен. Возможностей для предубеждения в таких оценках и в различных социальных выводах или приписываниях, которые отражают такие оценки, очень много. Определенное предубеждение приписывания, которое мы здесь рассмотрим, касается тенденции людей чувствовать “ложный консенсус”, то есть считать свои собственные поведенческие выборы и оценки как относительно обычные и соответствующие данным обстоятельствам, рассматривая альтернативные ответы как необычные, ненормативные и несоответствующие.

Ссылки на “эгоцентрическую атрибуцию” (Heider, 1958; Jones и Nisbett, 1971), “атрибутивное проектирование” (Holmes, 1968) и на определенные выводы и явления, связанные с предубеждениями ложного консенсуса, случайно появились в литературе, посвященной социальному восприятию и приписыванию (ср. Katz и Allport, 1931; Kelley и Stahelski, 1970). Возможно, наиболее веское свидетельство можно найти в ряде исследований Росса, Грина и Хауса (Ross, Greene & House, 1977).

В первом упомянутом исследовании, испытуемые рассматривали описания гипотетических конфликтных ситуаций, с которыми они могли бы столкнуться и должны были (а) оценить обычность двух возможных альтернатив реакции; (б) указать альтернативу, которую они сами бы выбрали; (в) оценить черты “типичного” человека, который выбрал бы каждую из двух указанных альтернатив.

Оценки продемонстрировали эффект “ложного консенсуса”; испытуемые рассудили, что альтернатива, которую они выбрали, будет более обычной, чем невыбранная альтернатива. Очевидное заключение к предположению ложного консенсуса – то, что интуитивный психолог оценивает те реакции, которые отличаются от его собственных, как раскрывающие устойчивые черты характера действующего лица в более полном объеме, чем реакции, подобные его собственным. Данные Росса, Грина и Хауса (1977) подтверждают этот прогноз; испытуемые делали сравнительно более достоверные и крайние прогнозы относительно типичного человека, который осуществил бы невыбранную испытуемым альтернативу, чем относительно того, кто осуществил бы выбранную альтернативу.

Термин *относительный* является критическим в этой формулировке предубеждения ложного консенсуса и требует некоторого разъяснения. Очевидно, человек, который идет по канату натянутому между двумя небоскребами, устраивает революцию или принимает обет безбрачия, признает, что немногие люди разделяли бы его выбор. Предубеждение ложного консенсуса, однако, привело бы его к пониманию своего выбора как более нормального, чем показалось бы людям, которые не стали бы ходить по канату, устраивать революцию или принимать обет безбрачия. Точно так же тезис признает, что для некоторых категорий реакции оценки всех оценивающих могут иметь отклонения в равной степени. Случаи жестокого обращения с детьми, например, могут быть одинаково недооценены как родителями, жестоко обращающимися со своими детьми, так и обычными родителями. Относительные термины гипотезы ложного консенсуса приводят только к прогнозированию того, что родители, жестоко обращающиеся со своими детьми, оценят это явление как более обычное и менее раскрывающее личные качества, чем обычные родители.

При заключительной демонстрации Росса, Грина и Хауса (1977) не использовалась гипотетическая методология анкетного опроса, а испытуемые были поставлены в условия реальной и последовательной ситуации конфликта. Испытуемых попросили походить по студенческому городку в течение 30 минут с рекламным щитом, на котором было написано «Обедайте в кафе “У Джо”». Экспериментатор пояснил испытуемым, что они могут легко отказаться от участия в исследовании, но ему бы хотелось, чтобы они участвовали и “помогая научно-исследовательской работе, узнали кое-что интересное”. Затем испытуемых просили принять решение относительно участия в исследовании, оценить вероятные решения других и сделать выводы относительно черт характера людей, согласившихся или отказавшихся участвовать.

Результаты этой “реальной” конфликтной ситуации (Табл. 1) подтвердили выводы более ранних исследований в форме анкетного опроса, касающихся гипотетических ответов. В целом, испытуемые, согласившиеся носить рекламный плакат, оценили, что 62% человек сделает такой же выбор. Испытуемые, которые отказались это делать, оценили, что только 33% людей исполнит просьбу экспериментатора. Кроме того, как и было спрогнозировано, согласившиеся и не согласившиеся испытуемые резко расходились в относительной силе выводов, которые они хотели сделать о человеке, согласившемся и отказавшемся носить рекламный плакат. Согласившиеся испытуемые сделали более уверенные и экстремальные выводы относительно личных характеристик отказавшегося человека; не согласившиеся испытуемые сделали более резкие выводы относительно согласившегося человека.

Таблица 1. Эффект ложного консенсуса: оценки обычности и выводы относительно черт характера с точки зрения двух альтернатив поведения

	Оцененная обычность согласия (%)	Оцененная обычность отказа (%)	Сила выводов относительно черт характера ^а	
			О тестируемых, согласившихся носить плакат	О тестируемых, отказавшихся носить плакат
Испытуемые, согласившиеся носить плакат	62	38	120,1	125,3
Испытуемые, отказавшиеся носить плакат	33	67	139,7	106,8

а) сумма оценок для четырех черт; большее число обозначает большую уверенность и более крайние выводы

Источник: обобщено из Ross, Greene и House (1977).

Общие выводы по результатам Росса, Грина и Хауса (1977) для нашей концепции интуитивного психолога очевидны. Оценки отклонения и нормы, и совокупность социальных выводов и межличностных реакций, которые сопровождают такие оценки, являются смещенными в соответствии с собственными поведенческими выборами обывателей. Очевидно, что анализ приписывания может быть искажен не только ошибками в анализе социальных данных интуитивным психологом, но также и более ранними предубеждениями в формировании выборки или оценке таких данных.

Некоторые немотивационные факторы играют роль в создании явлений ложного консенсуса. Главные среди них: (а) факторы выборочного воздействия и доступности; (b) факторы, имеющие отношение к разрешению ситуативной неоднозначности.

Факторы выборочного воздействия, лежащие в основе ложного консенсуса, довольно просты. Очевидно, мы знаем и общаемся с людьми, которые разделяют наш опыт, интересы, ценности и взгляды. Эти люди действительно реагируют так, как отреагировали бы мы в широком разнообразии обстоятельств. Действительно, наше окружение частично определено чувствами общего консенсуса, и мы склонны избегать тех, кто вряд ли разделяет наши оценки и реакции. Тот факт, что мы выступаем перед пристрастной выборкой людей и поведения не требует, чтобы мы допускали ошибку в наших оценках относительно соответствующих совокупностей, но оно делает такие ошибки вероятными. Факторы, усиливающие нашу способность вспо-

минать, визуализировать или воображать образцовые случаи поведения являются более тонкими и когнитивными по характеру. В данной ситуации определенные типы поведения, которые мы выбрали или выбрали бы могут быть с большей легкостью восстановлены из памяти и более легко представлены, чем противоположные типы поведения. Согласно Канеману и Тверскому (1973, 4) поведенческие выборы, которым мы отдаем предпочтение, могут быть более когнитивно “доступны”, и нас можно ввести в заблуждение этой легкостью или трудностью доступа при оценивании вероятности соответствующих поведенческих выборов.

Второй немотивационный источник эффекта ложного консенсуса является результатом реакции интуитивного психолога на неоднозначность – как относительно природы и величины ситуативных сил, так и относительно значения и последствий различных альтернатив реакции. Попытки снять такую неоднозначность включают интерпретацию, оценку и догадки, причем все это может оказывать параллельный эффект на собственные выборы поведения наблюдателей, а также на их прогнозы и выводы относительно выборов других людей. Таким образом, испытуемые, которые ожидали и боялись насмешек других людей, потому что на них была табличка «Обедайте в кафе “У ДЖО”», а также те, которые сочли пожелания и ожидания экспериментатора банальными, вероятно, откажутся носить рекламный щит, поддержат подобные отказы других людей и сделают соответствующие выводы относительно черт характера любого испытуемого, который согласился носить табличку. Противоположные приоритеты, конечно, привели бы к противоположным личным выборам и противоположным социальным оценкам и выводам.

Итак, предубеждение ложного консенсуса, как отражает, так и создает искажения в процессе атрибуции. Оно следует из неслучайного формирования выборки и поиска свидетельства и от необычного решения неоднозначных ситуативных факторов и сил. В свою очередь, оно смещает оценки относительно нормы и отклоняется и увеличивает расхождения и ошибки в интерпретации социальных явлений.

Устойчивость убеждений вопреки опыту

Различные недостатки интуитивного психолога, те, которые описаны в этой главе и ряде других работ (см. Nisbett и Ross, 1980), могут привести его к тому, что он будет придерживаться убеждений о себе, других людях или даже природе социального мира, которые являются преждевременными и ошибочными во многих случаях. Пока о них никто не знает и никто на них не влияет, такие убеждения могут казаться непоследовательными – просто предварительными по сути и адаптивными к новым входным данным. Постепенно увеличивающийся объем теории и исследования, однако, может теперь привести к совершенно противоположному выводу.

Оказывается, что убеждения, от относительно узких личных впечатлений до более широких социальных теорий, являются очень устойчивыми вопреки опыту, который кажется логически разрушительным. Две парадигмы иллюстрируют эту устойчивость. Первая включает способность убеждения продолжать существовать и усиливаться новыми данными, которые, с нормативной точки зрения, должны приводить к тому, что убеждения становятся более умеренными. Вторая включает сохранение убеждений после отрицания их первоначальных оснований.

Устойчивость убеждений и поляризация в свете появления новых данных

Люди, общественные организации, группы интересов и даже нации часто обладают различающимися убеждениями о важных социальных или политических проблемах. Такие расхождения во мнениях едва ли являются удивительными. Учитывая неофициальное и часто чисто интуитивное основание, на котором такие мнения базируются, а также роль, которую социальные коммуникации (часто очень необъективные) играют в формировании наших убеждений, разногласия являются неизбежными. Но что происходит, когда сторонникам различных точек зрения, позволяют проанализировать определенное свидетельство, особенно, когда оно носит формальный характер и идентично для всех заинтересованных сторон?

Оптимистическое ожидание состоит в том, что соперничающие группировки нашли бы компромисс между своими убеждениями. Этот компромисс может включать смещение в сторону позиции, определенной соответствующим свидетельством, если оно последовательно и неоспоримо; с другой стороны, он мог бы состоять из изменения в сторону большей умеренности или взаимной терпимости, если свидетельство неоднозначно или неокончательно. Менее оптимистическое ожидание состоит в том, что соперничающие группировки остались бы при своих точках зрения; то есть они проигнорировали бы новое свидетельство и устойчиво придерживались бы своих первоначальных позиций. Недавний эксперимент, проведенный Лордом, Леппером и Россом (Lord, Lepper & Ross, 1979), приводит к более удручающему выводу (по крайней мере, для тех, кто надеется или ожидает, что объективные данные ученого в области социальных наук урегулируют социальные споры).

Лорд и другие (1979) сначала отобрали испытуемых, которые либо поддерживали высшую меру наказания, считая ее эффективной превентивной мерой (сторонники), либо выступали против высшей меры наказания и полагали, что она не является сдерживающей (противники). Испытуемых ознакомили с двумя подлинными эмпирическими исследованиями. Одно говорило в поддержку их позиции; другое опровергало их мнение. При чтении этих исследований, две группы дали оценки относительно этих двух исследований и изменений в своих собственных отношениях и мнениях. Эти оценки ярко показали способность сторонников определенной теории интерпретиро-

вать новое свидетельство так, чтобы усилить и подтвердить свою теорию. Во-первых, как сторонники, так и противники высшей меры наказания последовательно оценили исследование, поддерживающее их убеждения, как “более убедительное” и “лучше проведенное”, чем исследование, которое их опровергало. Во-вторых, и в отличие от любой нормативной стратегии, которую можно вообразить для включения нового свидетельства в убеждения человека, влияние чтения двух исследований должно было еще больше поляризовать убеждения противников и сторонников смертной казни. То, как произошла эта поляризация, было особенно очевидно (см. Рисунок 3). После чтения краткого отчета о результатах, который поддерживал их точку зрения, убеждения испытуемых стали значительно более экстремальными; эти изменения поддерживались или увеличивались при рассмотрении подробностей относительно процедуры и данных. Наоборот, после чтения краткого отчета о результатах, который опровергал их собственную точку зрения, мнения испытуемых стали менее экстремальными; а после чтения соответствующих подробностей относительно процедур и данных испытуемые возвращались к убеждениям, которые у них были до ознакомления с исследованием. Фактически, большое количество испытуемых, которые прочитали, как резюме результатов, так и процедурные подробности исследования, опровергающего их мнение, в конечном счете, стали более уверены в правильности этого убеждения! Таких эффектов не наблюдалось, когда те же самые результаты и процедуры были прочитаны испытуемыми, чьи начальные представления поддерживались.

Очевидно, профессиональные ученые часто виновны в той же ошибке, что и интуитивные. Снова и снова каждый видит как соперничающие группировки, которые вовлечены в академические споры – является ли их темой происхождение Вселенной, эволюция человека или существование предубеждений атрибуции эго-защиты – получают поддержку для своих различающихся мнений из одного и того же источника. Далее в этой главе мы подробно рассмотрим процесс, лежащий в основе таких явлений и более конкретно прокомментируем нормативный статус готовности ученого перерабатывать свидетельство в свете существующих теорий и ожиданий. Сначала необходимо рассмотреть второй общий класс явлений устойчивости.

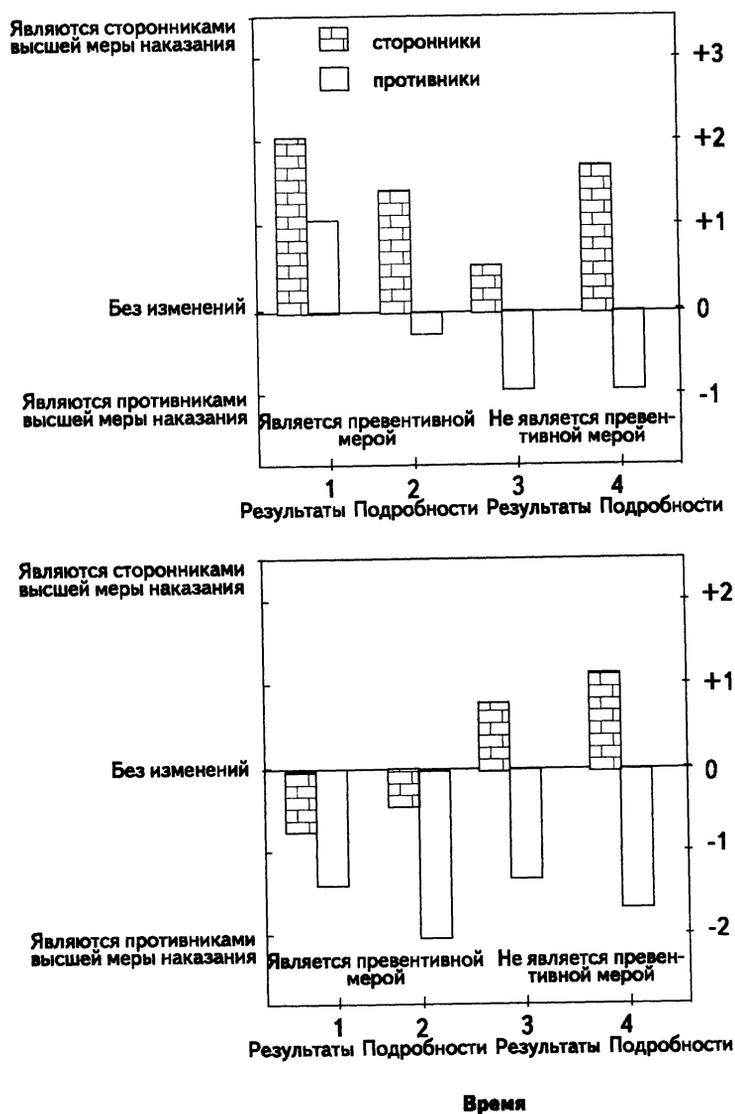


Рис. 3. Верхний рисунок: изменение отношения к высшей мере наказания, сообщаемое испытуемыми, в течение эксперимента, которые сначала рассматривали исследование, подтверждающее эффективность смертной казни в предотвращении преступлений. Нижний рисунок: изменение отношения к высшей мере наказания, сообщаемое испытуемыми, в течение эксперимента, которые сначала рассматривали исследование, опровергающее эффективность смертной казни в предотвращении преступлений

Устойчивость убеждений после очевидной дискредитации

Иногда, убеждениям угрожают не новые данные, а скорее требования привести свидетельство формирования такого убеждения, то есть прежде всего информацию или анализ, который привел к формированию этого убеждения. На уровне коротких рассказов легко привести такие примеры. Салли терпит неудачу в своих первых попытках научиться кататься на коньках и затем выясняет, что коньки, которые она позаимствовала, были слишком большие, чтобы обеспечивать ей поддержку в области лодыжек. Мэри предполагает, что букет цветов Джона отражает заботу, романтический характер и следование традициям, но затем узнает, что отец Джона – владелец цветочного магазина. Более широкие теории или убеждения о мире могут оспариваться подобным образом. Приходящая няня на основе опыта с единственным младенцем, который кричал всю ночь, решает, что кормление из бутылки приводит к коликам у младенцев, а затем обнаруживает, что этот ребенок страдал от высокой температуры. Или ученый обнаруживает, что классический эксперимент, на основании которого появилась некоторая теория, не является чистым из-за внешних факторов или прямого мошенничества.

Гипотеза устойчивости, в своей общей формулировке, гласит, что люди, о которых упоминалось выше, отстаивали бы свои начальные убеждения до степени ничем неоправданной. Но термины такого утверждения очевидно слишком общие и неопределенные, чтобы их можно было проверить. Когда именно мы сможем сделать вывод, что субъект “неуместно” упорствует в своих впечатлениях или убеждениях, хотя их основание было подорвано? Чтобы исследовать гипотезу устойчивости экспериментально, нам точно потребуются парадигма, которая позволит точно определять степень устойчивости и изменений.

Одна такая парадигма в форме дилеммы была предложена социальными психологами, которые использовали обман в ходе эксперимента, а затем опрашивали испытуемых, которые являлись целью такого обмана. Психологи показывали надуманный и недостоверный характер предоставленной информации, предполагая, что подобное рассмотрение эксперимента устранит любое влияние, которое такая информация, могла бы оказать на мнения или убеждения испытуемых. Многие ученые, однако, выразили общее беспокойство, что такие экспериментальные обманы могут нанести большой вред, который не будет полностью устранен с помощью обычных процедур рассмотрения (например, Kelman, 1972; A.G. Miller, 1972; Orne, 1972; Silverman, 1965).

Ряд экспериментов, проведенных Леппером, Россом и их коллегами (см. также более ранние исследования Walster и других, 1967; Valins, 1974) использовали общую парадигму дисконтирования или рассмотрения, с целью исследования явлений устойчивости убеждений при столкновении с очевидной дискредитацией. Мы начнем обсуждение этой работы, выделив пару экспериментов Росса, Леппера и Хаббарда (Ross, Lepper & Hubbard, 1975),

которые касаются впечатлений испытуемых после рассмотрения обмана о своих способностях в некоторой задаче или о способностях других людей.

Устойчивость личных впечатлений после их опровержения. Процедура, используемая Россом и другими (1975) была достаточно прямолинейной. Испытуемые сначала получали продолжительную ложную обратную связь во время выполнения задачи на проницательность (то есть, отличие подлинных сообщений о самоубийстве от вымышленных). В первом эксперименте было сообщено, что целью исследования является анализ того, как испытуемые управляют восприятием собственного поведения и способностей. Во втором эксперименте были привлечены наблюдатели, которые формировали социальные впечатления, когда были свидетелями управления ложной обратной связью. В обоих экспериментах после того, как манипуляция первыми впечатлениями была закончена, экспериментатор полностью дискредитировал «сведения», на которых были основаны впечатления участников и наблюдателей. В частности, участник (реплику которого в Эксперименте 2 нечаянно услышал наблюдатель), получил стандартное опровержение данных, благодаря чему он узнал, что предполагаемый исход был predetermined и что его обратная связь была абсолютно не связана с фактическим поведением. Прежде чем значения зависимых переменных были представлены, каждого испытуемого подвели к явному подтверждению его понимания характера и цели экспериментальной фальсификации.

После этой полной дискредитации первоначальной информации, испытуемые заполняли анкету с зависимыми переменными, описывающими действия людей и их способности. Устойчивость впечатления была очевидна как для участников, так и для наблюдателей. Фактически при каждом измерении (то есть, объективные оценки только что произведенного акта поведения участника, оценки поведения относительно будущего набора задач на проницательность и субъективные оценки его способностей) полностью дискредитированная манипуляция первичного исхода произвела существенные «остаточные» эффекты на участников и оценки наблюдателей (см. таб. 2).

Последующие эксперименты показали, что разнообразие необоснованных личных впечатлений, когда-то вызванных в соответствии с экспериментальными процедурами, может сохраниться даже после дискредитирующих процедур. Например, Дженнингс, Леппер и Росс (Jennings, Lepper & Ross, 1980) продемонстрировали, что впечатления испытуемых относительно их способности в межличностном убеждении (успех или неудача убедить конфедерата стать донором крови) могут сохраняться после того, как они узнали, что начальный исход был полностью недостоверен. Точно так же в двух сходных экспериментах Леппер, Росс и Лау (Lepper, Ross & Lau, 1979) показали, что ошибочные впечатления студентов относительно их «способности решать логические задачи» (и их выборы специализации с последующей оценкой через два месяца) упорно сохранялись даже после того, как

Таблица 2. Восприятие поведения и способностей участника после опровержения

	Личное восприятие участника			Восприятие участников наблюдателями		
	успех	неудача	t	успех	неудача	t
Оцененное первоначальное кол-во	18.33	12.83	5.91***	19.00	12.42	4.43***
Прогнозирование будущего числа	18.33	14.25	4.23***	19.08	14.50	2.68*
Оцененные способности	5.00	3.83	2.65*	5.33	4.00	3.36**

* $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$.

Источник: Данные второго эксперимента, проведенного Россом, Леппером и Хаббардом (1975).

они узнали, что хорошие или плохие методы обучения явились объяснением успехов или неудач, служивших основанием для таких впечатлений.

Устойчивость дискредитированных теорий после опровержения. Недавние эксперименты Андерсена, Леппера и Росса (1980) расширили область проявлений устойчивости от личных впечатлений до более обширных убеждений о реальном мире. В исследованиях Андерсена и других сначала создавались, а затем дискредитировались теории испытуемых относительно функциональных отношений между двумя измеряемыми переменными: адекватностью профессиональных действий пожарных и количеством очков, набранных в тесте на предпочтение риска. В одном варианте формирующие сведения состояли из пары определенных случаев, то есть, один состоявшийся и один несостоявшийся пожарный с соответствующим противоречивым количеством очков в тестах на предпочтение риска. Интересно, такие минимальные данные были достаточными, чтобы породить устойчивые теории, со стороны испытуемых, относительно вероятных связей между соответствующими измерениями. Более важным был вывод, что такие теории пережили опровержение – рассматриваемые случаи были полностью фиктивными, и различные испытуемые, фактически, получили противоположные пары очков рискованности и результатов работы. Действительно, когда сравнения были сделаны между испытуемыми, с которыми эксперимент проанализирован, и теми, с которыми нет, оказалось, что более чем 50% начального эффекта информации в рамках “случая” осталось после опровержения.

Итак, ясно, что убеждения могут сохраняться, несмотря на сильные логические или эмпирические сомнения. Они могут сохраняться и даже усиливаться за счет сведений, которые, как согласилось бы большинство нейтральных наблюдателей, логически должны приводить к некоторому ослаблению таких убеждений. Они могут даже сохраняться после полного разрушения их первоначальных основ. Необходимо сделать многое для определения точных пределов и исследования неизбежных исключений к таким явлениям. Ясно, что издержки предубеждений атрибуции обывателей и другие недостатки не подвержены коррекции, но вместе с тем сглаживаются за счет последующего опыта и обсуждений. Вопрос, который необходимо наконец задать – это *как и почему* такая устойчивость имеет место? То есть, какие когнитивные механизмы лежат в основе необоснованного постоянства наших впечатлений, убеждений и более широких социальных теорий?

Возможные механизмы, лежащие в основе устойчивости убеждений

Пристрастный поиск, воспоминание и ассимиляция информации. Не приходится сомневаться, что наши убеждения влияют на процессы, с помощью которых мы ищем, храним и интерпретируем соответствующую информацию. Действительно, без предшествующего знания и соответствующих предубеждений, наше понимание ежедневного опыта требовало бы значительно большего времени и усилий, и по всей вероятности, в результате снизилось бы понимание. Но неизбежное последствие нашей готовности обрабатывать сведения в свете наших априорных убеждений – это тенденция воспринимать большее подтверждение для этих убеждений, чем фактически существует в данных сведениях.

Такие “предубеждения подтверждения” (см. Einhorn и Hogarth, 1978; Hamilton, 1979; Hastie и Kumar, 1979; Wason и Johnson-Laird, 1972) отмечались философами на протяжении продолжительного времени (например, Васон, 1620/1960). Возможно, наиболее примечательна реакция сторонника теории на сомнительные или неоднозначные данные. Как утверждал Лорд и другие (1979), потенциально подтверждающие сведения могут быть приняты номинально (как таковые), в то время как потенциально опровергающие сведения подлежат критическому и скептическому исследованию. Таким образом, существуют два следствия: во-первых, любой образец сведений, обработанных этим способом, даже сведения, которое являются по существу случайными имеет тенденцию *поддерживать* начальные убеждения. Во-вторых, как только сведения были обработаны таким способом, они получают способность *усиливать* априорные убеждения, когда это убеждение подвергается новому эмпирическому опровержению или нападением на его первоначальное очевидное основание.

Роль предвзятой ассимиляции была убедительно показана для случая, где сторонник теории встречается с новыми данными (т.е. Lord и другие, 1979). Но роль этого механизма в опровержении парадигмы, возможно, менее очевидна, и мы вынуждены положиться скорее на предположение, чем на твердые данные. Мы предполагаем, что испытуемый, который формирует начальное впечатление о себе, другом человеке или некоторых функциональных отношениях, способен искать дополнительные данные, соответствующие этому впечатлению, в своей памяти и в непосредственной ситуации. Такие данные могут вспоминаться и рассматриваться как подходящие или подтверждающие только настолько, насколько они доказывают настоящее впечатление. Таким образом, испытуемый, который преуспел или потерпел неудачу в данной задаче, вспоминает подобные успехи или неудачи в похожих задачах – и принимает решения по поводу их релевантности данному случаю – на основе соответствия исходов. Точно так же испытуемый, решив, что переменные X и Y функционально связаны, вспомнит и придаст правдоподобность случаям, которые, скорее подтверждают, чем оспаривают это предположение. Опять же, такой предвзятый поиск, вспоминание и ассимиляция не только поддерживают первоначальное убеждение, они также создают образец доказательства, который остается доступным для поддержания рассматриваемого убеждения, когда его *первоначальное* основание атаковано или даже разрушено. Критическое предположение здесь заключается в том, что люди не постоянно совершенствуют или переоценивают сведения, соответствующие их убеждениям. Они обычно не приходят к решению “теперь, когда моя предшествующая гипотеза была несколько подорвана, я должен возвратиться и переоценить все сведения, которые я когда-либо рассматривал в свете этой гипотезы”.

Формирование причинных объяснений. Люди более чем просто замечают сведения, соответствующие их впечатлениям или убеждениям. Они также участвуют в причинном анализе или объяснении (Heider, 1958). То есть они пробуют *объяснить* характеристики себя или других людей, или функциональные отношения, в существование которых они верят. Таким образом, испытуемая, которая считает, что она лучше или хуже отличает сообщения о самоубийстве в исследованиях Росса (1975), могла бы найти аспект своей биографии, который объяснит наличие или отсутствие такого таланта. Точно так же испытуемый, который вынужден верить в положительные или отрицательные отношения между способностью к пожарному делу и предпочтением риска, будет иметь мало трудностей в определении логического основания для любых отношений. Опять же, этот процесс не только поддерживает начальное впечатление или убеждение, он также способен сохранить это впечатление или убеждение при возникновении сомнений или критике.

Свидетельство действия этого механизма устойчивости мы можем найти в двух исследованиях опровержения, демонстрирующих, что, когда испытуемым необходимо сформулировать такие объяснения до опровержения,

величина эффекта устойчивости растет. В исследовании Андерсона и других (1980), одну группу испытуемых попросили объяснить положительные или отрицательные связи в двух случаях с пожарными. Как и было спрогнозировано, вмешательство сильно усилило эффект устойчивости. Фактически, испытуемые, которые объясняли основание для положительных или отрицательных связей до опровержения, были только чуть менее уверены в них, чем испытуемые, которые не получили никакого опровержения. Подобные результаты были получены Россом, Лешпером, Стреком и Стейнмец (1977), обнаружившими, что испытуемые, которым было необходимо объяснить дальнейшую жизнь пациентов клиники (чьи истории болезни они читали), продолжили расценивать такие исходы как относительно вероятные даже, когда они узнали, что события были недостоверны и выдуманы экспериментатором.

Поведенческое подтверждение или гипотеза “самоисполнения” (self-fulfilment). Две парадигмы исследования, используемые Россом, Лешпером и их коллегами для исследования явления устойчивости испытывают недостаток одного элемента, который может быть критическим во многих ежедневных ситуациях. Определенно у испытуемых в этих исследованиях не было возможности воздействовать на свои убеждения. Такие действия важны частично, потому что они могут увеличивать психологические затраты или “диссонанс” (Festinger, 1957), связанный с изменением убеждений (ср. Ashmore и Collins, 1968; Collins и Hoyt, 1972; Hovland, Campbell, и Brock, 1957). Кроме того, такие действия создают новые данные, соответствующие этим убеждениям. Эти новые данные могут быть обработаны предвзято, кроме того, они могут быть смещены в направлении, подтверждающем соответствующую гипотезу.

Идея гипотезы само-подтверждения или самоисполнения не нова для социологов. Известные, но спорные исследования, например “эффект Пигмалиона” (Rosenthal и Jacobson, 1968), связанные с воздействием ожиданий преподавателей относительно “скачка” в способностях и поведении их студентов, являются актуальными. Однако, недавние исследования, проведенные Шнайдер (Snyder) и его коллегами, значительно продвинули нашу оценку и понимание таких явлений, демонстрируя, как ожидания испытуемых или гипотезы, которые они должны проверить, могут порождать их “объективную поддержку” (например, Snyder и Swann, 1973a, 1978b; Snyder, Tanke и Berscheid, 1977).

Заключительные замечания: Убеждения меняются!

Наше предшествующее обсуждение явлений и механизмов не должно заставить читателя потерять из виду тот факт, что убеждения о нас, наших политических лидерах и даже наших научных теориях действительно меняются. Частично такое изменение может просто быть результатом грубой

силы. Даже если логические или эмпирические опровержения имеют меньше воздействия, чем могли бы гарантировать нормативные стандарты (см. Ross и Lepper, 1980), они все-таки могут подействовать. Частично, такое изменение может отражать факт, что формальные методы тестирования гипотезы, иногда преднамеренно используются, чтобы защитить нас от опасностей неофициальных методов. Но мы подозреваем, многое еще предстоит исследовать, так как существует свидетельство того, что предшествующие теории могут иногда изменяться без огромного количества опровержений или убедительных экспериментов. Таким образом, изменения в мировоззрении и убеждениях, которые могут быть вызваны ярким, конкретным, непосредственным опытом (см. Nisbett и Ross, 1980) и эффективность групп и лидеров, которые выполняют серьезные политические или религиозные преобразования, предлагают цели для будущего исследования.

10. Очевидное воздействие базового значения*

Амос Тверски и Даниель Канеман

Во многих ситуациях людям необходимо оценить вероятность некоторого целевого события (например, диагноз пациента или продажи учебника) на основе (а) частоты базового значения целевого исхода в некоторой совокупности (например, частота различных диагнозов или распределения продаж учебника), (б) определенного доказательства данного случая (например, реакция пациента на диагностические осмотры или оглавление рассматриваемого текста).

Важность данных базового значения в интуитивных прогнозах относительно частных случаев была изучена Милом и Роузенем (Meehl и Rosen, 1955), которые, используя правило Байеса утверждали, что прогнозирование редкого исхода (например, самоубийства) на основе ненадежных данных – это главный источник ошибок в клиническом прогнозировании. Мил и Роузен не проводили экспериментальных исследований, но они привели примеры из литературы по клиническому диагнозу, в котором информация базового значения не была принята во внимание.

Чтобы получить экспериментальное воздействие данных базового значения, мы предоставили испытуемым описание аспиранта или специалиста и попросили их спрогнозировать его область специализации или его профессию (Kahneman и Tversky, 1973, 4). Эти исследования показали, что последующие оценки вероятности были определены прежде всего степенью, в которой описание было подобно или репрезентативно соответствующему стереотипу специалиста (например, библиотекарям или адвокатам). Частотами базового значения этих категорий, которые были либо известны испытуемым из их ежедневного опыта, либо явно указаны в вопросе, в значительной степени пренебрегали. (Мы используем термин *пренебрегать*, чтобы описать ситуации, в которых базовое значение либо игнорируется, либо крайне недооценивается.)

* Эта работа была поддержана Службой военно-морских исследований согласно Контракту N00014-79-C-0077 со Стэнфордским университетом.

Прогнозы на основе репрезентативности или подобия не зависят от частот базового значения. Однако явление пренебрежения базовым значением гораздо более распространено, так как оно также наблюдается в оценках, которые не могут с легкостью интерпретироваться в терминах репрезентативности (Hammerton, 1973). Для примера, Касселлс, Шоенберг и Грейбойс (Casscells, Schoenberger, и Grayboys, 1978) задавали 60 студентам и персоналу Медицинского факультета в Гарварде следующий вопрос:

Если анализ, для обнаружения болезни, распространенность которой = $1/1.000$, имеет ложную положительную пропорцию 5%*, какова вероятность, того, что человек, у которого обнаружили, положительный результат, фактически болен, если предположить, что вы ничего не знаете о симптомах этого человека? (с.999)

Наиболее обычный ответ, который дали почти половина участников, был 95%. Средний ответ был 56% и только 11 участников дали правильный ответ – 2%, предполагая, что анализ правильно выявляет каждого больного человека. Очевидно даже высоко образованные респонденты часто не в состоянии оценить важность базового значения исхода в относительно простых формальных задачах (см., например, Bar-Hillel, 1980a; Lyon и Slovic, 1976). Осуждение Милом и Роузенем (1955) неспособности оценивать базовое значение не ограничивается клиническими психологами; это также свойственно и врачам, и вообще обывателям.

Условия, при которых данные базового значения используются или ими пренебрегают, широко исследовались социальными психологами и учеными, изучающими производство суждений (см. литературные обзоры Borgida и Brekke (1981) и Kassin (1979b)). Независимые переменные, рассмотренные в этих исследованиях, могут быть разделены на два типа: процедурные и доказательные. Процедурные переменные относятся к свойствам проекта, задачи и демонстрации, в то время как доказательные переменные касаются характера источника и интерпретации сведений.

Например, процедурная переменная большой важности – обращается ли испытуемый с каждой задачей как с отдельным случаем или участвует в задаче с многократным прогнозированием. Значительные сведения из исследований вероятности и похожих задач показывают, что люди имеют тенденцию уравнивать распределение критерия при многократных предсказаниях, особенно в присутствии обратной связи исхода. Поскольку люди пытаются создавать образец прогнозов, который является репрезентативным распределению исхода, эксперименты, использующие повторные оценки с тем же базовым значением производят большие эффекты базового значения, чем эксперименты, в которых с каждой оценкой обращаются как с отдельной задачей. (См. Bar-Hillel и Fischhoff, 1981; Manis и другие, 1980).

Другая интересующая нас процедурная переменная – это различие между внутригрупповым и межгрупповыми планами. Например, Фишхтоф,

* Т. е. при положительном диагнозе болезнь отсутствует у 5% людей.

Словик и Лихтенштейн (Fischhoff, Slovic и Lichtenstein (1979)) показали, что данные базового значения имеют больше воздействия, когда они различны в разных задачах, предоставленных каждому испытуемому, чем когда различные базовые пропорции предоставлены различным испытуемым. Внутригрупповой план, однако, стимулирует общую тенденцию приписывать больший вес варьирующемуся признаку, даже, когда он не соответствует нормативно (Fischhoff и Bar-Hillel, 1980). Дальнейшее обсуждение различий между внутри- и межгрупповыми планами и их приложение для исследования статистических убеждений смотри в гл. 34.

Хотя процедурные переменные имеют значительное влияние, данная глава ограничивается обсуждением доказательных переменных, которые управляют интерпретацией и воздействием данных базового значения. Определенно, мы обращаем внимание на различия между двумя типами базовых значений, которые мы назовем причинными и случайными.

Причинные и случайные базовые значения

Базовое значение называется причинным, если оно предполагает существование причинного фактора, который объясняет, почему отдельный случай приведет скорее к этому исходу, чем другой. Базовое значение называется случайным, если оно не ведет к такому выводу.

Яркая демонстрация контраста между причинными и случайными базовыми значениями была представлена Айзенем (Ajzen, 1977). В одном эксперименте, респонденты оценили вероятность того, что студент, академическая успеваемость которого была кратко описана, успешно сдал отдельно взятый экзамен. Причинное базовое значение было представлено следующим образом:

Два года назад сдавался заключительный экзамен в Йельский университет. Приблизительно 75% студентов потерпели неудачу (успешно сдали) экзамен.

Это базовое значение является причинным, потому что оно подразумевает, что экзамен был исключительно труден (если 75% студентов потерпели неудачу) или относительно легкий (если 75% студентов сдали его). Выведенная причина (то есть, трудность экзамена) “объясняет” базовое значение и делает менее (более) вероятным то, что каждый студент сдаст экзамен.

Случайное базовое значение было представлено следующим образом:

Два года назад, в Йельском университете сдавался заключительный экзамен. Психолог-педагог, интересующийся успеваемостью, опросил большое количество студентов, которые прошли курс обучения. Так как он был, прежде всего, заинтересован в реакциях на успех (неудачу), он выбирал главным образом студентов, которые успешно сдали (потерпели неудачу) экзамен. В частности, приблизительно 75% студентов в его выборке успешно сдали (провалили) экзамен.

Это базовое значение является случайным, или не-причинным, потому что пропорция студентов, сдавших экзамен или нет, в выборке была наугад

отобрана исследователем. В отличие от базового значения, она не позволяет сделать вывода относительно трудности экзамена.

Исследование Айзена (1977) показало, что причинное базовое значение обладало большим весом, чем случайное, хотя оба типа базового значения оказали существенное влияние. Для причинного базового значения, оцененная вероятность успеха (средний по всем описаниям) была выше 0.34, при большом значении базового значения успеха, чем при низком. Для случайного базового значения соответствующее различие было только 0.12. В терминах данного анализа, легкость или трудность экзамена – одна из способствующих причин, которые затрагивают поведение студента, и поэтому связана с другими причинами, такими как уровень интеллекта и мотивированность рассматриваемого студента.

Базовое значение успеха использовалось в предшествующем исследовании, чтобы определить экзамен как легкий или трудный. Во втором исследовании, базовое значение предпочтения использовалось, чтобы определить выборы как более или менее привлекательные (Ajzen, 1977). Испытуемых попросили оценить вероятность, что студенты, чьи краткие описания личности прилагались, выберут в качестве факультативного курса или историю, или экономику. Причинное базовое значение, которое служило для определения относительной привлекательности одного из этих двух выбора, состояло из пропорции студентов, зарегистрированных на этих двух курсах (0.70 и 0.30). Случайное базовое значение было представлено следующим образом:

Чтобы узнать мнение студентов, профессор истории (экономики) недавно опросил 70 студентов, которые проходили его общий курс по истории (экономике), для сравнения, он также опросил 30 студентов, которые прошли курс экономики (истории).

Обратите внимание, что, в отличие от причинного базового значения, случайное не дает информации относительно популярности этих двух курсов. Влияние случайного базового значения не было существенным в этом исследовании, хотя имелось различие вероятности 0.025 в ожидаемом направлении. Напротив, причинное базовое значение имело сильное влияние: средняя оцененная вероятность выбора была 0.65 для популярного курса (высокое базовое значение) и 0.36 для непопулярного курса (низкое базовое значение). Очевидно, привлекательность курсов выведена из базовых значений выборов и объединена с личными характеристиками при оценке вероятности того, что отдельный студент выберет скорее один курс, чем другой. С нормативной точки зрения, однако, и причинное, и случайное базовое значение в этих примерах должны иметь едва сопоставимые влияния.

Наш следующий пример иллюстрирует другой тип причинного базового значения; оно также позволяет вычислить правильную апостериорную вероятность согласно некоторым разумным предположениям. Рассмотрим следующую измененную версию задачи про такси, первоначально представленную Канеманом и Тверским (1972a) и позже исследованную Бар-Хиллел (1980a), и Тверским и Канеманом (1980, 8).

Такси сбило человека и скрылось с места происшествия ночью. Две компании такси, Зеленая и Синяя, работают в городе. Вам дают следующие данные:

(А) 85% такси в городе – Зеленые и 15% – Синие.

(В) свидетель идентифицировал такси как Синее. Суд проверил надежность свидетеля при обстоятельствах, которые существовали в ночь несчастного случая, и заключил, что свидетель правильно идентифицировал каждый из этих двух цветов в 80% случаев и неправильно в 20% случаев.

Какова вероятность того, что такси, сбившее человека было скорее Синим, чем Зеленым?

Чтобы получить правильный ответ, пусть В и G обозначают соответственно гипотезы, что такси, сбившее человека, Синее или Зеленое, и пусть W – сообщение свидетеля. В соответствии с правилом Байеса в вероятностной форме, с предшествующей вероятностью 15/85 и отношением вероятности 80/20,

$$\begin{aligned} P(B/W) / P(G/W) &= P(W/B) P(B) / P(W/G) P(G) = \\ &= (0.8) (0.15) / (0.2) (0.85) = 12 / 17 \end{aligned}$$

и, следовательно

$$P(B/W) = 12 / (12 + 17) = 0.41$$

Поэтому, несмотря на сообщение свидетеля, такси, сбившее человека, будет скорее Зеленым, чем Синим, потому что базовое значение является более крайним, чем свидетельское показание вероятным.

Большому количеству испытуемых предоставили несколько отличающиеся версии этой задачи с очень последовательными результатами. Средний и модальный ответ – 0.80, значение, которое совпадает с достоверностью свидетельского показания и, очевидно, не зависит от относительной частоты Синих и Зеленых такси.

Информация базового значения, однако, использовалась в отсутствии целевых данных. Когда пункт (b) был опущен из вопроса, почти все тестируемые дали основное значение (0.15) как ответ. Кроме того, базовое значение управляло ожиданием испытуемых относительно свидетельских данных. Другой группе испытуемых была предоставлена вышеупомянутая задача за исключением того, что предложение “свидетель определил такси, как Синее” было заменено на “свидетель определил цвет такси”. Затем этих испытуемых спросили, “какова вероятность, что свидетель определил такси как Синее?” Средний и модальный ответ на этот вопрос был 0.15. Обратите внимание, что правильный ответ – $0.2 \times 0.85 + 0.8 \times 0.15 = 0.29$. В отсутствии других данных, знание базового значения использовалось должным образом, чтобы спрогнозировать целевой исход и ненадлежащим образом, чтобы спрогнозировать сообщение свидетеля.

Другой вид суждений наблюдался, когда случайное базовое значение (такси) было заменено причинным базовым значением (несчастных случаев). Это было сделано путем замены пункта (а) на:

(а') Хотя эти две компании приблизительно равны в размере, 85% несчастных случаев с такси в городе произошло с Зелеными такси и 15% — с Синими.

Ответы на эту задачу были различными, но базовое значение больше не игнорировалось. Средний ответ был 0.60, что находится между достоверностью свидетельского показания (0.80) и правильным ответом (0.41). Базовое значение в случае (а) является причинным, потому что различие в пропорциях несчастных случаев между компаниями равного размера легко выявляет вывод, что водители Зеленых такси менее осторожны или менее компетентны, чем водители Синих такси. Этот вывод объясняет дифференциальные базовые значения несчастных случаев и подразумевает, что любое Зеленое такси, более вероятно, будет вовлечено в несчастный случай, чем любое Синее такси. Напротив, основное значение в (а) случайно, потому что различие между количеством Синих и Зеленых такси в городе не делает возможным причинный вывод, который делает более вероятным, что любое Зеленое такси, а не любое Синее такси попадет в несчастный случай.

Обратите внимание, что согласно данному анализу, последующая вероятность, что такси, которое принимало участие в несчастном случае, является, скорее Синим, чем Зеленым – та же самая и в (а), и в (а'). Однако связь между цветом такси и причастностью к несчастным случаям *равняется* 0 для случайного базового значения и 0.7 – для причинного! Этот статистический факт отражает различие между двумя базовыми значениями и помогает объяснить, почему причинное базовое значение используется, в то время как случайное базовое значение игнорируется.

Другие доказательные переменные

Причинный или случайный характер данных базового значения не единственная доказательная переменная, которая влияет на интуитивные оценки. Даже в отсутствии причинной интерпретации, данные базового значения не заменяются неопределенными, обобщенными или непоследовательными данными случая. Например, Бар-Хиллел (1980а) изучала версию первоначальной задачи с такси, в которой информация относительно свидетеля (пункт b) была заменена сообщением, что сбившее человека такси было оборудовано селекторной связью (переговорными устройствами) и что селекторная связь установлена в 80% Зеленых такси и в 20% Синих такси. В этой задаче, (случайное) базовое значение не было отвергнуто, и средний ответ был 0.48. Бар-Хиллел предположила, что сведения относительно селекторной связи не заменили базовое значение, потому что они менее определенные, чем идентификация свидетелем. Таким образом, данные базового значения объединяются с другими сведениями в двух случаях: если пер-

вые имеют причинную интерпретацию, или если последние не являются более определенными, чем базовое значение (Bar-Hillel, 1980a).

И специфика, и причинная связь могут помочь объяснить различие между результатами Канемана и Тверского (1973, 4), которые обнаружили существенное пренебрежение базовым значением в прогнозировании области специализации студента на основе краткого описания личности и результатами МакКоли и Ститта (McCaulley и Stitt, 1978). Когда они обнаружили существенную корреляцию между оцененным базовым значением черт характера и оцененными вероятностями этих черт при указании национальности, например, вероятность, что человек является практичным, если он немец. Кроме нескольких процедурных различий, последнее исследование отличается от первого в трех важных аспектах. Во-первых, испытуемых просили спрогнозировать скорее относительную частоту (например, пропорция немцев, которые являются практичными), чем вероятность для индивидуального случая. Во-вторых, сведения состояли скорее из принадлежности к классу, например, немец, чем детальных описаний определенного человека. В-третьих, частоту базового значения черт характера легче причинно интерпретировать, чем частоту базового значения профессий. Обывательские теории личности предлагают причины, почему большинство людей любят повеселиться, и только некоторые из них мазохисты. Эти причины могут объяснить поведение людей вообще и немцев, в частности, таким образом, обеспечивая причинную интерпретацию базового значения черт характера.

Особенно интересно рассмотреть ситуацию, касающуюся определенных, но недиагностических сведений (например, описание человека, одинаково подобного инженеру и адвокату). Экспериментальные выводы здесь не полностью последовательны. Канеман и Тверский (1973, 4) обнаружили пренебрежение базовым значением, в то время как Джиносар и Троуп (Ginosar и Trope, 1980) обнаружили исключительную уверенность в данных базового значения при подобных экспериментальных условиях. Большинство исследований, однако, получили промежуточные результаты, где базовое значение не было отвергнуто, а скорее ослаблено недиагностическими сведениями относительно данного случая (см. например, Manis и другие, 1980; Wells и Harvey, 1977).

Внутренние или внешние приписывания

Несколько задач с базовым значением, представляющих особый интерес для социальных психологов, возникает, когда сведения и базовое значение относятся соответственно к внутренне-поведенческим и к внешне-ситуативным факторам, которые влияют на исход. Успех студента на экзамене, например, определен как трудностью экзамена, так и талантом студента. Точно так же ответ на просьбу пожертвовать деньги в благотворительное предприятие зависит как от щедрости дающего, так и от характера просьбы.

Внешние факторы, такие как трудность экзамена или эффективность просьбы, естественно, выражены соответствующими базовыми значениями (например, 75% студентов провалили экзамен; большинство людей внесло вклад в благотворительное предприятие). Вопрос об относительном воздействии ситуативных и поведенческих факторов в социальном приписывании, таким образом, может быть повторно сформулирован в терминах значения, которое придается соответствующим базовым значениям.

Нисбетт и Боргида первыми исследовали связь между использованием информации базового значения в исследовании оценок и относительным весом ситуативных факторов в изучении атрибуции поведения. Они показали, что в исследовании Дарли-Латана (Darley-Latane, 1968) знание того, что испытуемые редко оказывали помощь, не затрагивало прогнозов испытуемых относительно поведения отдельного участника исследования, который был показан в кратком интервью, снятом на камеру. Исследование Нисбетта и Боргиды (1975) внесло вклад в пересекаемость когнитивных и социально-психологических подходов к изучению оценивания. Это также вызвало противоречие (Borgida, 1978; Wells и Harvey, 1977, 1978) и ряд исследований роли согласованности информации в прогнозировании поведения (Borgida и Brekke, 1981; Kassin, 1979b; Nisbett и Ross, 1980; Ross, 1977).

В отличие от примеров с экзаменами и такси, в которых причинные и случайные базовые значения ясно различимы, базовые значения во многих исследованиях консенсуса подвержены альтернативным интерпретациям. Чтобы это проиллюстрировать, давайте сравним исследование Нисбетта и Боргиды (1975) с причинным условием базового значения в эксперименте Айзена (Ajzen, 1977), где испытуемые оценили вероятность того, что отдельно взятый студент сдаст экзамен, который провалили 75% группы. Формальная структура двух задач одна и та же, но базовым значением в значительной степени пренебрегали в первом исследовании и использовали в последнем. Кажется, что необычному базовому значению давали ситуативную интерпретацию в исследовании Айзена, но оно интерпретировалось как случайное формирование выборки в исследовании Нисбетта-Боргиды.

Оценки испытуемых Айзена указывают, что они вывели из низкого базового значения успеха, что экзамен был труден, хотя они могли использовать те же самые сведения, чтобы прийти к выводу, что студенты, которые проходили тест, не были достаточно умными. Напротив, испытуемые Нисбетта и Боргиды, очевидно пришли к выводу, что участники исследования об оказании помощи были главным образом бесчувственными и жестокими людьми (Wells и Harvey, 1977). Они не пришли к правильному заключению, что ситуация исследования Дарли-Латана не располагала к проявлению помощи.

Приписано ли экстремальное базовое значение случайному формированию выборки или ситуативным факторам, зависит от контекста проблемы: более вероятно, что необычное распределение результатов тестов возникло из-за сложности (или легкости) экзамена, чем из-за необычного состава групп.

пы. С другой стороны тяжелее пересмотреть концепцию относительно условий, при которых люди помогают незнакомому человеку, которого хватил удар, чем предположить, что участники исследования об оказании помощи были исключительно бессердечными.

Очевидное пренебрежение данными базового значения в прогнозах отдельных случаев связано с выводом относительно необычных характеристик членов группы. Причинная интерпретация базового значения становится более вероятной, если этот вывод блокирован. Эта гипотеза была поддержана несколькими исследованиями, которые восстановили эффект влияния базового значения, подчеркивая репрезентативность выборки, в которой наблюдалось необычное поведение (Hansen и Donoghue, 1977; Hansen и Lowe, 1976; Wells и Harvey, 1978). Воздействие данных базового значения было даже увеличено в одном исследовании, сообщая испытуемым, что выборка, для которой давались базовые значения, была большой и поэтому надежной (Kassin, 1979a). Главное заключение этого исследования – то, что использование или пренебрежение согласованностью информации в отдельном прогнозе критически зависит от интерпретации этой информации.

Часть IV
Доступность

11. Доступность: эвристика оценки частоты и вероятности*

Амос Тверски и Даниель Канеман

Введение

Большинство недавних исследований были посвящены проблемам валидности и последовательности оценок вероятности и частоты. Однако известно мало о психологических механизмах, с помощью которых люди оценивают частоту классов или вероятность событий.

Мы предполагаем, что при столкновении с трудной задачей оценки вероятности или частоты, люди используют ограниченное количество эвристик, которые сокращают эти оценки до более простых. Ранее мы проанализировали подробно одну из таких эвристик – репрезентативность. С помощью этой эвристики, событие оценивается как вероятное настолько, насколько оно представляет существенные особенности своей родительской совокупности или процесса, породившего его...

При оценке вероятности события с помощью репрезентативности, человек сравнивает существенные особенности события со структурой, из которой оно исходит. Таким образом, человек оценивает вероятности, оценивая подобие или коннотативное расстояние. С другой стороны, можно оценить вероятность, используя доступность, или ассоциативное расстояние. Из жизненного опыта мы знаем, что частные случаи больших классов вспоминаются лучше и быстрее, чем случаи менее частых классов, что вероятные события легче вообразить, чем маловероятные, и что ассоциативные связи усиливаются, когда два события часто происходят одновременно. Таким образом, человек мог оценить численность класса, вероятность события или частоту взаимосвязанных появлений событий, оценивая легкость, с которой может быть выполнено соответствующее умственное действие формирования выборки, построения или ассоциации.

* Эта глава – сокращенная версия работы, которая появилась в *Cognitive Psychology*, 1973, 4, 207-232. Авторское право © 1972 Academic Press, Inc. Переиздано в соответствии с разрешением

Например, можно оценивать пропорцию разводов в данном обществе, вспоминая разводы среди знакомых; можно оценивать вероятность того, что политический деятель проиграет на выборах, рассматривая различные способы, с помощью которых он может потерять поддержку; и можно оценивать вероятность того, что жестокий человек “увидит” хищных зверей в карточках Роршаха, оценивая силу ассоциации между насилием и хищниками. Во всех этих случаях, оценка частоты класса или вероятности случая опосредована оценкой доступности¹. Считается, что человек использует эвристику доступности всякий раз, когда он оценивает частоту или вероятность за счет легкости, с которой события или ассоциации могут прийти ему в голову. Чтобы оценить доступность, совсем не обязательно выполнять фактические действия формирования выборки или построения. Достаточно оценить легкость, с которой эти действия могли быть выполнены, подобно тому, как трудность загадки или математической задачи может быть оценена без рассмотрения определенных решений.

То, что ассоциативные связи усиливаются при повторении – возможно, самый старый закон запоминания, известный человеку. Эвристика доступности использует обратную форму этого закона, то есть она использует силу ассоциации как основание для оценки частоты. В этой теории, доступность является, скорее, переменной-посредником, чем зависимой переменной как обычно бывает в исследовании памяти. Доступность – надежный ключ для оценки частоты, потому что, частые события легче вспомнить или вообразить, чем редкие. Однако на доступность также воздействуют различные факторы, которые не связаны с фактической частотой. Если применяется эвристика доступности, то такие факторы повлияют на воспринятую частоту классов и субъективную вероятность событий. Следовательно, использование эвристики доступности ведет к систематическим предубеждениям.

Эта часть книги исследует эвристику доступности в десяти последовательных исследованиях.² Сначала мы показываем, что люди могут оценивать доступность с разумной скоростью и точностью. Затем, мы показываем, что оцененная частота классов смещается за счет доступности их частных случаев для построения и процесса формирования воспроизведения. Экспери-

¹ Настоящее использование термина “доступность” не совпадает с некоторыми случаями употребления этого термина в литературе по исследованию речи (см., например, Horowitz, Norman и Day, 1966; Tulving и Pearlstone, 1966).

² Приблизительно 1500 испытуемых участвовали в этих исследованиях. При отсутствии дополнительной информации, исследования проводились в группах из 20-40 испытуемых. Испытуемые в исследованиях 1, 2, 3, 9 и 10 были привлечены рекламными объявлениями в студенческой газете в Университете штата Орегон. Испытуемые в исследовании 8 были подобным образом привлечены в Университете Стэнфорд. Испытуемые в исследованиях 5, 6 и 7 были учащимися в 10-ых и 11-ых классов нескольких среднеобразовательных школ-колледжей в Израиле.

ментальные исследования этой части книги касаются оценок частот или вероятностей, которые могут легко уменьшены до относительных частот. Влияние доступности на оцененные вероятности по существу уникальных событий (которые не могут быть уменьшены до относительных частот) обсуждены в пятой и заключительной главе.

Оценки доступности

Исследование 1: Построение

Испытуемым ($N = 42$) предоставлен ряд задач на построение слов. Каждая задача состояла из матрицы 3×3 , содержащих девять букв для составления слов из трех букв или больше. В обучающей стадии исследования, всем испытуемым предоставили шесть задач. На каждую задачу, им давали 7 секунд, чтобы оценить количество слов, которые они могли бы составить за 2 минуты. После каждой оценки, им давали две минуты, чтобы записать (в пронумерованных строках) столько слов, сколько они могли составить из букв в матрице. Данные, полученные в обучающей стадии, были исключены. В стадии испытания, построение и задачи оценивания были отделены. Каждый испытуемый, оценивал для восьми задач количество слов, которое он мог бы придумать за 2 минуты. Для восьми других задач, он составлял слова без предшествующей оценки. Оценивание и задачи построения чередовались. Использовались два параллельных стимульных проспекта, чтобы для каждой задачи половина испытуемых оценивала и половина составляла слова.

Результаты. Среднее количество составленных слов варьировалось от 1.3 (для XUZONLCJM) до 22.4 (для TAPCERHOB), с общим средним 11.9. Среднее оцененное число варьировалось от 4.9 до 16.0 (для тех же самых двух задач), с общим средним 10.3. Корреляция между оценкой и продуктивностью по этим шестнадцати задачам составила 0.96.

Исследование 2: Воспроизведение

Структура и процедура были идентичны исследованию 1, за исключением характера задачи. Здесь, каждая проблема состояла из категории, например, *цветы* или *русские писатели*, частные случаи которых необходимо было вспомнить. Испытуемым ($N = 28$) давали 7 секунд на то, чтобы оценить количество частных случаев, которые они могли вспомнить за 2 минуты, или 2 минуты на то, чтобы фактически вспомнить. Как в исследовании 1, задачи построения и оценивания были объединены в обучающейся стадии и чередовались в стадии испытаний.

Результаты. Среднее количество случаев варьировалось от 4.1 (названия городов начинающиеся на букву Ф) до 23.7 (четвероногие животные), с общим средним 11.7. Среднее оцененное число варьировалось от 6.7 до 18.7 (для тех же самых двух категорий), с общим средним 10.8. Корреляция между продуктивностью и оценением по этим 16 категориям была 0.93.

Обсуждение

В вышеупомянутых исследованиях, доступность событий могла быть измерена общим количеством выбранных или созданных случаев в любой задаче.³ Исследования показывают, что люди могут быстро и точно оценивать доступность. Как такие оценки осуществляются? Один вероятный механизм предложен в работе Боусфилда и Седжевика (Bousfield & Sedgewick, 1944), которые показали, что совокупное воспроизведение случаев является отрицательно ускоренной показательной функцией от времени. Испытуемый мог использовать количество случаев, выбранных в короткий промежуток времени, чтобы оценить количество случаев, которые он мог выбрать в длительный отрезок времени. С другой стороны, испытуемый может оценивать доступность без явного воспроизведения или построения случаев вообще. Харт (Hart, 1967), например, показал, что люди могут точно оценивать свою способность узнавать предметы, которые они не могут вспомнить в тесте на запоминание парных ассоциаций.

Доступность для построения

Мы обратимся теперь к ряду проблем, в которых испытуемому дают правило для построения случаев и просят оценить их общую (или относительную) частоту. В этих задачах – как в большинстве задач на оценивание, испытуемый не может создать и перечислить все случаи. Вместо этого, мы предполагаем, он пытается создавать некоторые случаи и оценивает полную частоту степени доступности, то есть в соответствии с оценкой легкости, с которой случаи могут прийти в голову. Как следствие, классы явлений, случаи которых легко создать или вообразить, будут восприниматься как более частые, чем классы того же самого размера, которые менее доступны. Этот прогноз проверен на оценке частоты слова и на оценке нескольких комбинаторных выражений.

³ Проблемы построения слова могут также рассматриваться как проблемы воспроизведения, потому что слова-ответы существуют в памяти. В этой части книги мы говорим о воспроизведении, когда тестируемый вспоминает случаи естественной категории, как в исследованиях 2 и 8. Мы говорим о построении, когда испытуемый производит образцы согласно указанному правилу, как в исследованиях 1 и 4.

Исследование 3: Оценка частоты слова

Предположим, Вы выбрали наугад слово из английского текста. Что более вероятно, что слово начинается на К, или что К – его третья буква? Согласно нашему предположению, люди отвечают на такой вопрос, сравнивая доступность этих двух категорий, то есть, оценивая легкость, с которой случаи этих двух категорий приходят на ум. Конечно, легче придумать слова, начинающиеся с К, чем слова, где К является третьей буквой. Если оценка частоты опосредована оцененной доступностью, то слова, начинающиеся на К, должны быть оценены как более частые. На самом деле, обычный текст содержит вдвое больше слов, в которых К находится в третьей позиции, чем слов, начинающихся на К.

Согласно обширному подсчету слов Майцнера и Тресселта (Maizner и Tresselt, 1965), существует восемь согласных, которые встречаются чаще в третьей, чем в первой позиции. Из них, две согласные (Х и Z) относительно редки, а другая (D) часто встречается в третьей позиции только в словах из трех букв. Оставшиеся пять согласных (K, L, N, R, V) были отобраны для исследования.

Испытуемым дали следующие инструкции:

Изучалась частота появления букв в английском языке. Был отобран обычный текст, была зарегистрирована относительная частота появления различных букв алфавита в первой и третьей позициях слов. Слова, состоящие менее, чем из трех букв, были исключены из подсчета.

Вам предложат несколько букв алфавита и попросят оценить, появляются ли эти буквы чаще в первой или в третьей позиции, и оценить отношение частоты, с которой они появляются в этих позициях.

Типичная задача, была следующей:

Рассмотрите букву Р. Что более вероятно, что Р появится в

- Первой позиции?

- Третьей позиции?

Моя оценка соотношения этих двух величин ____ : 1.

Испытуемых проинструктировали, чтобы они оценили отношение большего класса к меньшему. Для половины испытуемых, упорядочение этих двух позиций в вопросе было изменено. Кроме того, использовались три различных порядка этих пяти букв.

Результаты. Среди 152 испытуемых, 105 оценили, что первая позиция более вероятна для большинства букв и 47 оценили, что – третья. Предубеждение в пользу первой позиции является существенным ($p < 0.001$, тест знаков). Кроме того, по оценке большинства испытуемых, каждая из пяти букв появлялась чаще в первой, чем в третьей позиции. Отношение средней оценки было 2:1 для каждой из этих пяти букв. Эти результаты были получены, несмотря на тот факт, что все буквы встречались чаще в третьей позиции.

В других исследованиях мы обнаружили то же предубеждение в пользу первой позиции при внутригрупповом плане исследования, где каждый испытуемый оценивал отдельную букву, и при межгрупповом плане, где частоты букв в первой и в третьей позициях были оценены различными испытуемыми. Мы также наблюдали, что вознаграждение за точность в первом проекте не имело никакого влияния. Так как во всех методах был получен сходный характер результатов, мы приводим только результаты, полученные в соответствии с самой простой процедурой.

О подобном результате было сообщено Филлипсом (Phillips, 1966) в исследовании вывода Байеса. Шесть редакторов студенческих газет оценивали вероятности, что различные двухбуквенные сочетания (биграммы), отобранные из собственных записей, были взяты с начала или с конца слов. Дополнительный эффект, наблюдаемый в этом исследовании, показал, что все редакторы были предубеждены в пользу гипотезы, согласно которой двухбуквенные сочетания были взяты с начала слов. Например, редакторы ошибочно оценили слова, начинающиеся с “re”, более частыми, чем слова, заканчивающиеся на “re”. Первые, конечно, являются более доступными, чем последние.

Исследование 4: Перестановки

Рассмотрим две структуры, А и В, которые показаны ниже.

(А)	(В)
x x x x x x x	x x
x x x x x x x	x x
x x x x x x x	x x
	x x
	x x
	x x
	x x
	x x

Дорожка в структуре — это линия, которая соединяет элемент в верхней строке с элементом в нижней строке, и пересекает только один элемент в каждой строке.

В какой из двух структур больше дорожек?

Сколько дорожек, по вашему мнению, имеется в каждой структуре?

Большинство читателей, вероятно, разделят непосредственное впечатление, что в А больше дорожек, чем в В. Наши испытуемые пришли к такому же выводу: 46 из 54 испытуемых увидели большее количество дорожек в А, чем в В ($p < 0.001$, тест знаков). Средние оценки были 40 дорожек в А и 18 в В. Фактически, количество дорожек — одинаковое в обеих структурах, так как $8^3 = 2^9 = 512$.

Почему люди видят больше дорожек в А, чем в В? Мы предполагаем, что этот результат отражает различную доступность дорожек в двух структурах. Существуют несколько факторов, которые делают дорожки в А более доступными, чем в В. Во-первых, наиболее доступные дорожки – столбцы структур. В А есть 8 столбцов, а в В только 2. Во-вторых, дорожки, пересекающие столбцы в А вообще более различимы и меньше запутаны, чем в В. Две дорожки в А, в среднем, пересекают 1/8 их элементов, в то время как две дорожки в В, в среднем, пересекают половину их элементов. Наконец, дорожки в А короче и, следовательно, легче воспринимаются на глаз, чем дорожки в В.

Исследование 5: Комбинации

Рассмотрим группу из десяти человек, которые должны сформировать комитеты из r членов, где r - некоторое число между 2 и 8. Сколько различных комитетов из r членов они могут сформировать? Правильный ответ на эту задачу задается биномиальным коэффициентом (r^{10}), который достигает максимума 252 для $r = 5$. Ясно, что количество комитетов из r членов равняется числу комитетов из $10 - r$ членов, потому что любая выбранная группа, скажем, из двух членов определяет единственно возможную невыбранную группу из восьми членов.

Согласно нашему анализу интуитивной оценки, однако, комитеты из двух членов более доступны, чем комитеты из восьми. Во-первых, самая простая схема построения комитетов – разделение группы на непересекающиеся подмножества. Таким образом, каждый с легкостью видит, что существуют целых пять непересекающихся комитетов из двух членов, но не видит и двух непересекающихся комитетов из восьми. Во-вторых, комитеты из восьми членов намного менее различимы, так как их элементы пересекаются; любые два комитета из восьми имеют, по крайней мере, шесть общих членов. Этот анализ предполагает, что маленькие комитеты более доступны, чем большие комитеты. В соответствии с гипотезой доступности, поэтому, маленькие комитеты должны казаться более многочисленными.

Четыре группы испытуемых (общее количество $N = 118$) оценили количество возможных комитетов из r членов, которые могут быть сформированы из десяти человек. Различные группы, соответственно, оценили следующие значения r : 1 и 6; 3 и 8; 4 и 7; 5.

Средние оценки количества комитетов показаны на Рисунке 1, с правильными значениями. Как и было спрогнозировано, оцененная численность комитетов уменьшается с их размером.

Следующая альтернативная формулировка той же самой задачи была изобретена, чтобы проверить общность результатов:

В рисунке, приведенном ниже, имеются десять станций по маршруту между Стартом и Финишем. Рассмотрим автобус, который следует по этому маршруту, останавливаясь точно на r станциях.

СТАРТ

--	--	--	--	--	--	--	--	--	--

 ФИНИШ

Сколько различных вариантов r остановок автобус может делать?

Количество различных образцов r остановок снова (r^{10}). Здесь также количество вариантов двух остановок такое же, что и количество вариантов восьми остановок, потому что для любого варианта остановок имеется единственно возможный дополнительный вариант не-остановок. Все же, человек оказывается более свободен в построении образцов двух остановок, где “существует много станций, из которых можно выбирать”, чем в построении образцов восьми остановок, где “нужно останавливаться почти на каждой остановке.” Наш предыдущий анализ предполагает, что первые образцы более доступны: большее количество таких образцов замечается с первого взгляда, они более отличительны и их легче визуализировать.

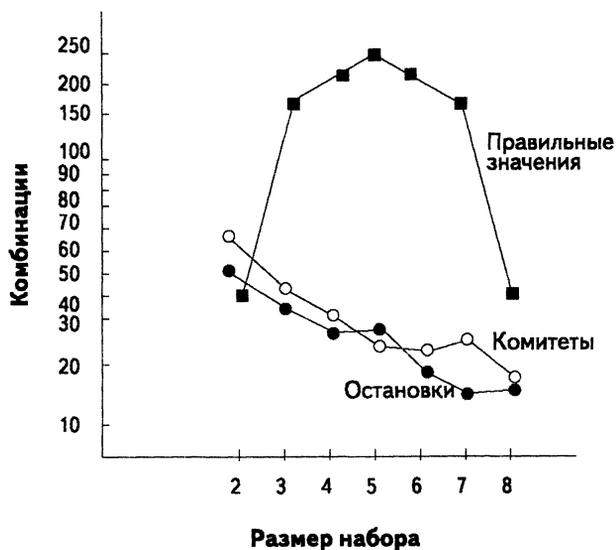


Рис. 1. Правильные величины и средние оценки (на логарифмической шкале) для задачи про комитеты и остановки

Четыре новых группы испытуемых (общее количество $N = 178$) ответили на этот вопрос, для $r = 2, \dots, 8$ по тому же образцу, как указано выше. Средние оценки количества остановок показаны на Рисунке 1. Как и в задаче про комитеты, очевидное количество комбинаций вообще уменьшается с r , в соответствии с прогнозом гипотезы доступности, и в заметном несоответствии с правильными величинами. Оценки количества комбинаций подобны в двух задачах. Как в других комбинаторных задачах, существует явная

недооценка всех правильных величин, с единственным исключением в наиболее доступном случае, где $r = 2$.

Недооценка, наблюдаемая в экспериментах 4 и 5 происходит потому, что люди оценивают комбинаторные величины, экстраполируя от начального впечатления. То, что человек видит сразу или в нескольких шагах вычисления, дает ему неправильную идею относительно неограниченно возрастающей пропорции роста многих комбинаторных выражений. В таких ситуациях, экстраполирование от начального впечатления ведет к явной недооценке. Дело в том, является ли основание для экстраполяции изначальной доступностью случаев, как в предшествующих двух исследованиях или результатом изначального вычисления, как в следующем исследовании.

Исследование 6: Экстраполяция

Мы попросили, испытуемых оценить, в пределах 5 секунд, числовое выражение, которое было написано на доске. Одна группа испытуемых ($N = 87$) оценивала выражение $8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1$, в то время как другая группа ($N = 114$) оценила выражение $1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8$. Средняя оценка для убывающей последовательности была 2. 250. Средняя оценка для возрастающей последовательности была 512. Различие между оценками высоко существенно ($p < 0.001$, тест медианы). Обе оценки далеки от правильного ответа 40. 320.

И недооценка правильной величины, и различия между двумя оценками поддерживает гипотезу, согласно которой, люди оценивают $8!$, экстраполируя от частичного вычисления. Факториал, подобно другим комбинаторным выражениям, характеризуется постоянно увеличивающейся пропорцией роста. Следовательно, человек, который экстраполирует от частичного вычисления, чрезвычайно недооценит факториалы. Поскольку результаты первых шагов умножения (выполненных слева направо) больше в нисходящей последовательности, чем в возрастающей последовательности, первое выражение оценено большим, чем последнее. Оценка нисходящей последовательности может происходить следующим образом: “ $8 \times 7 = 56$ умножить на 6 – уже больше 300, так что, мы имеем дело с достаточно большим числом.” В оценке возрастающей последовательности, с другой стороны, можно рассуждать: “ 1×2 равняется 2, умножить на 3 равняется 6, умножить на 4 – 24, это выражение не очень большое ...”

Исследование 7: Бином – доступность или репрезентативность

Заключительное исследование этой главы исследует роль доступности в оценке биномиальных распределений и иллюстрирует, как формулировка задачи управляет выбором эвристики, которую люди принимают при интуитивной оценке.

Испытуемым (N - 73) предоставили следующие инструкции:

Рассмотрите следующую диаграмму:

```

X X 0 X X X
X X X X 0 X
X 0 X X X X
X X X 0 X X
X X X X X 0
0 X X X X X
  
```

Дорожка в этой диаграмме – любая нисходящая линия, которая начинается в верхней строке, заканчивается в нижней строке и проходит точно через один символ (X или O) в каждой строке.

Как Вы думаете, какой, процент дорожек содержит

6 - X и без - O ___ %

5 - X и 1-O _____ %

.

.

.

без - X и 6-O _____ %

Обратите внимание, что они включают все возможные типы дорожек и, следовательно, Ваши оценки должны в сумме составить 100%.

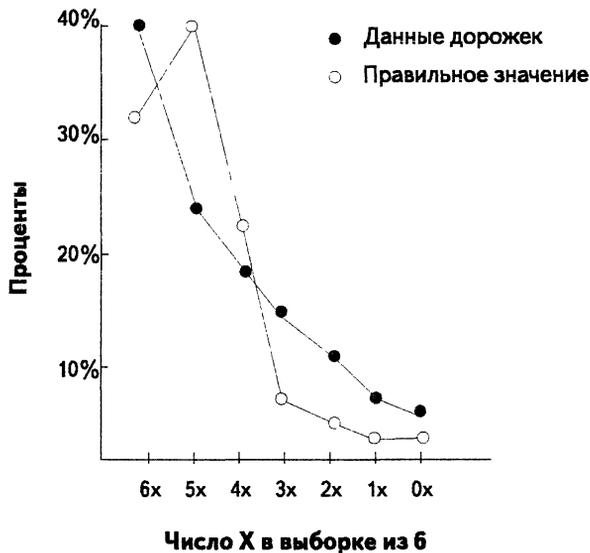


Рис. 2. Правильные значения и средние оценки: задача про дорожки

Фактическое распределение типа дорожки биномиально с $p = 5/6$ и $n = 6$. Люди, конечно, не могут ни угадать правильные ответы, ни перечислить все соответствующие случаи. Вместо этого, по нашему предположению, испытуемые посмотрят на диаграмму и оценят относительную частоту каждого типа дорожки легкостью, с которой отдельные дорожки этого типа могли быть построены. Так как, на каждой стадии построения дорожки (то есть, в каждой строке диаграммы) имеется намного больше X , чем O , легче построить дорожки, состоящие из шести X , чем дорожки, состоящие, скажем, из пяти X и одного O , хотя последние, фактически, более многочисленны. Соответственно, мы спрогнозировали, что испытуемые будут ошибочно оценивать дорожки из 6 X и без O , как наиболее многочисленные.

Средние оценки относительной частоты всех типов дорожек представлены на Рисунке 2, наряду с правильными биномиальными величинами. Результаты подтверждают гипотезу. Из 73 испытуемых, 54 ошибочно оценили, что существует больше дорожек, состоящих из шести X и O , чем дорожек, состоящих из пяти X и одного O , и только 13 оценили последние, как более многочисленные, чем вышеупомянутые ($p < 0.001$, тест знаков). Монотонность субъективного распределения типов дорожек – общее явление. Мы получили тот же самый результат с различными величинами p ($4/5$ и $5/6$) и n (5, 6 и 10), и различными представлениями пропорций совокупности (например, четыре X и один O или восемь X и два O в каждой строке диаграммы дорожек).

Чтобы исследовать устойчивость этого эффекта, мы провели дополнительное испытание. Пятидесяти студентам старших курсов Стэнфордского университета, не имеющим опыта в комбинаторике, предоставили задачу про дорожки. Здесь, испытуемых не просили оценить относительную частоту, а просто решить “имеется ли больше дорожек, содержащих шесть X и O или дорожек, содержащих пять X и O .” Испытуемых опрашивали индивидуально и им обещали поощрение в \$ 1 за правильные оценки. Абсолютное большинство испытуемых (38 из 50, $p < 0.001$, тест знаков) снова выбрало первый упомянутый результат как более частый. Ошибочные интуиции, очевидно, трудно исправляются за счет предоставления денежных вознаграждений.

Мы предположили, что, когда биномиальное распределение представлено как диаграмма дорожек, люди оценивают относительную частоту различных исходов, определяя доступность отдельных дорожек каждого типа. Этот способ оценки предложен последовательным характером определения дорожки и иллюстрированным представлением задачи. Рассмотрим альтернативную формулировку этой же задачи.

Шесть игроков участвуют в карточной игре. В каждом туре игры, каждый игрок получает одну карту, взятую наугад из хорошо перетасованной колоды. В колоде, 5/6 карт помечены X и оставшиеся 1/6 помечены O . При большом количестве туров, какой процент туров, в которых

6 игроков получают X и ни один игрок не получает O ___ %
 5 игроков получают X, и 1 игрок получает O ___ %
 Ни один игрок не получает X и 6 игроков получают O ___ %

Обратите внимание, что здесь включены все возможные исходы и, следовательно, ваши оценки должны в сумме составить 100 %.

Задача про карты формально идентична задаче про дорожки, но мы намеревались выявить другой способ оценивания. В задаче про дорожки, отдельные случаи были выделены путем демонстрации и пропорции совокупности (то есть, пропорция X в каждой строке) не была определена. В задаче про карты, с другой стороны, пропорция совокупности явно задана, и отсутствует упоминание относительно отдельных случаев. Следовательно, мы выдвигаем гипотезу, что исходы в задаче про карты будут определены степенью, в которой они репрезентативны по отношению к составу колоды, чем доступностью отдельных случаев. В задаче про карты, исход “пять X и один O” является наиболее репрезентативным, потому что он соответствует пропорции совокупности (см. Kahneman и Tversky, 1972b, 3). По эвристике репрезентативности, этот исход должен быть оценен как более частый, чем исход “шесть X и без O,” вопреки наблюдаемому образцу оценок в задаче про дорожки. Оценки 71 из 82 испытуемых, которые отвечали на задачу про карты, подтвердили этот прогноз. В задаче про дорожки, только 13 из 73 испытуемых оценили эти исходы таким же образом; различие между этими двумя версиями значимо ($p < 0.001$, тест χ^2).

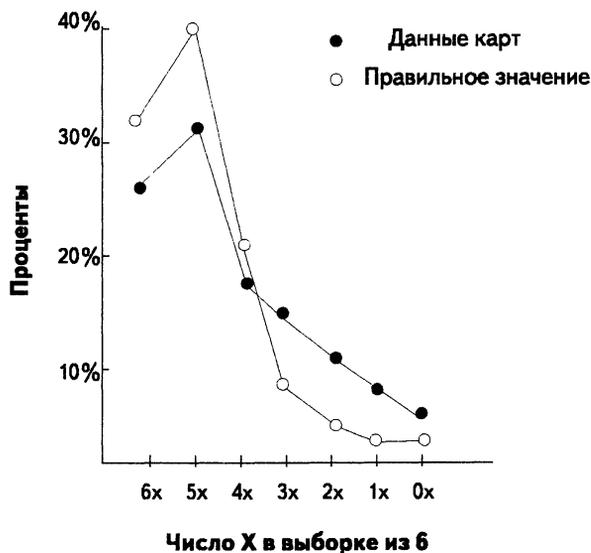


Рис. 3. Правильные значения и средние оценки: задача про карты

Средние оценки для задачи про карты представлены на Рисунке 3. Различие между рисунками 2 и 3 поддерживает гипотезу, что различные представления одной и той же задачи выявляют различную эвристику. В частности, частота класса, вероятно, будет оценена степенью доступности, если выделены отдельные случаи, и репрезентативностью, если существенными являются признаки категории.

Доступность воспроизведения

В этой главе мы обсуждаем несколько исследований, в которых испытуемому сначала предоставляют информацию (например, список имен) и потом просят, оценить частоту объектов данного типа, которые были включены в сообщение. Как и в задачах, исследуемых в предыдущей главе, испытуемый не может вспомнить и сосчитать все случаи. Мы предполагаем, что вместо этого он пытается вспомнить некоторые случаи и оценивает полную частоту степенью доступности, то есть легкостью с которой случаи приходят на ум. Как следствие, классы, чьи случаи с легкостью вспоминаются, будут оценены как более многочисленные, чем классы того же размера, чьи случаи менее доступны. Этот прогноз сначала был проверен в исследовании оцененной частоты категорий...

Исследование 8: Известность, частота и вспоминание

Испытуемым предоставили магнитофонную запись списка имен известных лиц обоих полов. Прослушав список, некоторые испытуемые оценивали, содержал ли он большее количество имен мужчин или женщин, другие попытались вспомнить имена в списке. Некоторые из имен в списке были очень известны (например, Ричард Никсон, Элизабет Тейлор), другие были менее известны (например, Уильям Фулбрайт, Лана Турнер). Известные имена легче вспомнить. Следовательно, если оценки частоты опосредованы оцененной доступностью, то класс, состоящий из известных имен, должен быть оценен как более многочисленный, чем сравнимый класс, состоящий из менее известных имен.

Были подготовлены четыре списка имен: два списка артистов и два списка общественных деятелей. Каждый список включал 39 имен и был записан со скоростью одно имя в 2 секунды. Два списка (один общественных деятелей и один артистов) состояли из 19 имен известных женщин и 20 имен менее известных мужчин. Два других списка состояли из 19 имен известных мужчин и 20 имен менее известных женщин. Следовательно, известность и частота были обратно связаны во всех списках. Имена всех лиц всегда позволяли однозначную идентификацию пола.

Испытуемых попросили слушать внимательно записанное сообщение. Каждый из четырех списков был представлен двум группам. После прослушивания записи, испытуемых в одной группе попросили записать столько

имен, сколько они могли вспомнить из списка. Испытуемых в другой группе попросили оценить, содержал ли список больше имен мужчин или женщин.

Результаты. (а) Вспоминание. В среднем, испытуемые вспомнили 12.3 из 19 известных имен и 8.4 из 20 менее известных имен. Из 86 испытуемых в четырех группах вспоминания, 57 воспроизвели больше известных, чем неизвестных имен, и только 13 вспомнили меньше известных, чем не очень известных имен ($p < 0.001$, тест знаков).

(б) Частота. Среди 99 испытуемых, сравнивавших частоту мужчин и женщин в списках, 80 ошибочно оценили класс, состоящий из более известных имен, как более частый ($p < 0.001$, тест знаков)....

Воспроизведение и построение сценариев

Во всех эмпирических исследованиях, которые были обсуждены в этой части книги, существовала объективная процедура для перечисления случаев (например, слова, которые начинаются на К или дорожки в диаграмме), и, следовательно, каждая из задач имела объективно правильный ответ. Другое дело в случаях из реальной жизни, в которых оцениваются вероятности. Каждый случай экономического спада, успешного медицинского воздействия или развод, является по существу уникальным, и его вероятность не может быть оценена простым числом случаев. Как бы там ни было, эвристика доступности может применяться к оценке вероятности таких событий.

В оценке вероятности, что отдельно взятая пара разведется, например, можно полагаться на свою память в поисках пары, которая приходит в голову при упоминании этого вопроса. Развод будет казаться возможным, если разводы распространены среди случаев, которые выбраны таким образом. В другом случае, можно оценивать вероятность, пытаясь создать истории, или сценарии, которые ведут к разводу. Правдоподобие таких сценариев или легкость, с которой они приходят на ум, могут обеспечивать основу для оценки вероятности. В данной главе, мы обсуждаем роль доступности в таких оценках, размышляем относительно ожидаемых источников предубеждения и предлагаем некоторые направления дальнейших исследований.

Мы иллюстрируем предубеждения доступности, рассматривая воображаемую ситуацию.⁴ Клинический психолог, услышавший жалобу пациента, что тот утомлен жизнью, задается вопросом, вероятно ли, что этот пациент совершит самоубийство, и может вспомнить подобных пациентов, которых он знал. Иногда только один случай приходит на ум, возможно, потому что он самый

⁴ Этот пример был выбран из-за его доступности. У нас нет достаточной причины полагать, что интуитивные прогнозы биржевых маклеров, спортивных комментаторов, политических аналитиков или психологов-исследователей менее подвержены предубеждениям.

незабываемый. Здесь, субъективная вероятность может зависеть, прежде всего, от подобия между тем событием и рассматриваемым. Если два события похожи, то человек ожидает повторение прошлого опыта. Когда несколько случаев приходят на ум, им придается столько значения, насколько они похожи, в существенных особенностях, исследуемой проблеме.

Насколько уместны отобранные случаи? Анализируя свой прошлый опыт, клинический психолог вспоминает пациентов, которые подходят данному случаю, – тех, кто предпринимал попытки самоубийства, или тех, кто похож на данный случай и совершал попытки самоубийства? С актуальной точки зрения, конечно, соответствующий класс – такой класс пациентов, которые подобны, в некотором отношении, данному случаю, и соответствующая статистика – это частота попыток самоубийства в классе.

Поиск в памяти может осуществляться по другим правилам. Так как попытки самоубийства – драматическое и чрезвычайное событие, пациенты, склонные к самоубийству, вероятно, будут легче запоминаться и вспоминаться, чем депрессивные пациенты, которые не предпринимали подобных попыток. Как следствие, клинический психолог может вспомнить пациентов, склонных к самоубийству, с которыми он сталкивался, и оценить вероятность попытки самоубийства степенью подобия между этими случаями и данным пациентом. Такой подход приводит к серьезным предубеждениям. Клинический психолог, замечая, что почти все пациенты, склонные к самоубийству, которых он может вспомнить, были угнетены, может прийти к выводу, что пациент, вероятно, совершит самоубийство, если он показывает признаки серьезной депрессии. С другой стороны, клинический психолог может сделать вывод, что самоубийство маловероятно, если «этот пациент не напоминает ему ни одного случая самоубийства, с которым он когда-либо сталкивался». Такое рассуждение игнорирует факт, что только меньшинство депрессивных пациентов совершает попытку самоубийства, и возможность, что данный пациент весьма отличен от любого, с которым он когда-либо сталкивался.

Наконец, клинический психолог мог бы подумать только о пациентах, которые были как депрессивны, так и склонны к самоубийству. Он тогда оценил бы вероятность самоубийства легкостью, с которой такие случаи, приходят на ум или степенью, в которой существующий пациент является репрезентативным этому классу. Это рассуждение, также, обладает серьезным недостатком. Тот факт, что существует много депрессивных пациентов, которые предприняли попытку самоубийства, не свидетельствует относительно вероятности, что депрессивный пациент предпримет попытку самоубийства, и все же этот способ оценивания обычен. Несколько исследований (Jenkins & Ward 1963; Smedslund, 1963; Ward & Jenkins, 1965) показали, что случайность между двумя бинарными переменными такими, как признак и болезнь оценивается частотой, с которой они происходят взаимосвязано, практически не принимая во внимание случаи, где либо признак, либо болезнь отсутствовали.

Некоторые события воспринимаются настолько уникально, что прошлый опыт не кажется уместным при оценке их вероятности. В размышлении о таких событиях мы часто строим *сценарии*, то есть, истории, которые ведут от существующей ситуации к целевому событию. Правдоподобие сценариев, которые приходят на ум, или трудность их создания, служит ключом к оценке вероятности события. Если обоснованный сценарий не приходит на ум, случай считают невозможным или маловероятным. Если на ум приходит много сценариев, или если один придуманный сценарий особенно убедителен, рассматриваемый случай кажется вероятным.

Многие из событий, вероятность которых люди желают оценить, зависят от нескольких взаимосвязанных факторов. Все же чрезвычайно трудно для человеческого ума прочувствовать последовательности изменений нескольких взаимодействующих факторов. Мы предполагаем, что в оценке вероятности сложных событий рассматриваются только самые простые и наиболее доступные сценарии. В частности, люди будут склонны создавать сценарии, по которым большинство факторов вообще не изменяется, присутствуют только наиболее очевидные изменения, и взаимодействующие изменения редки. Из-за упрощенного характера предполагаемых сценариев, исходы компьютерного моделирования процессов взаимодействия часто противоречат интуиции (Forrester, 1971). Тенденция рассматривать только относительно простые сценарии может иметь особенно существенные эффекты в ситуациях конфликта. В этом случае, собственное настроение и планы более доступны, чем планы другого человека. Не легко принять мнение противника по шахматной доске или полю битвы, вот почему посредственный игрок обнаруживает так много новых возможностей, когда он меняет сторону в игре. Следовательно, игрок может иметь тенденцию расценивать стратегию его противника как относительно постоянную и не зависящую от его собственных шагов. Эти соображения говорят, что игрок восприимчив к *ошибке инициативы* – тенденции приписывать меньше инициативы и воображения своему противнику, чем себе. Эта гипотеза согласуется с выводом поиска приписывания (Jones & Nisbett, 1971), согласно которому люди склонны воспринимать свое поведение как отражение изменяющихся потребностей окружающей среды, а поведение других – как отражение черт характера.

Создание сценария накладывает отпечаток на дальнейшее мышление. Существует много сведений, показывающих, что как только некоторая ситуация воспринимается или интерпретируется определенным образом, становится чрезвычайно трудно воспринимать ее по-другому (см, например, Bruner & Potter). Таким образом, составление определенного сценария может препятствовать появлению других сценариев, особенно тех, которые ведут к другим исходам...

Возможно, наиболее очевидное проявление доступности в реальной жизни – это влияние случайной доступности случаев или сценариев. Многие читатели испытали временное увеличение субъективной вероятности несча-

стного случая, увидев на обочине перевернутую машину. Подобным образом, многие замечают усиление субъективной вероятности того, что несчастный случай или неполадки оборудования спровоцируют начало термоядерной войны после просмотра фильма, ярко описывающего такую возможность. Сосредоточение внимания на исходе, может увеличить его доступность, и, следовательно, его кажущуюся вероятность. Люди увлекаются желаемыми исходами, такими как выигрыш в тотализаторе или крайне нежелательными исходами, как крушение самолета. Следовательно, доступность обеспечивает механизм, согласно которому явления крайней полезности (или бесполезности) могут происходить более вероятно, чем на самом деле...

12. Эгоцентрические преубеждения в доступности и атрибуции*

Михаэль Росс и Фьер Сиколи

Один пример явления, изученного в настоящем эксперименте, знаком практически каждому, кто проводил исследование совместно с другими. Рассмотрим следующее: Вы работали над совместным проектом с другим человеком и поднимается вопрос, кто является “главным автором” (то есть, кто внес больший вклад в окончательный проект?). Часто кажется, что вы оба можете претендовать на это звание. И так как Вы убеждены, что Ваш коллега разделяет ваше видение реальности (где существует только одна реальность), Вы предполагаете, что другой человек пытается использовать Вас. Иногда такие разногласия улаживаются или их появление предотвращается использованием правил условного решения, например, правила “алфавитного порядка” – любимая уступка тех, чьи фамилии начинаются с букв, расположенных в верхней части алфавита.

Мы предполагаем, что отдельные участники совместного проекта склонны принимать на себя больше ответственности за результат, чем другие участники приписывают им. Далее мы предположили, что распределение участниками ответственности за совместный проект является распространенным явлением. Однако, люди, занимающиеся общим делом, не всегда знают о расхождении их взглядов, так как нет необходимости приписывать “авторство”, следовательно, универсальность явления не является очевидной. Цель настоящего исследования – оценить, действительно ли эти эгоцентрические восприятия происходят в различных условиях и исследовать соответствующие психологические процессы.

Изучая основы различного восприятия, мы допускаем, что иногда люди намеренно себя возвеличивают. Однако вероятно, что восприятие может не совпадать при отсутствии намеренного обмана. Именно с этой точки зрения мы и исследуем проблему.

* Выдержки из работы, появившейся в “The Journal of Personality and Social Psychology”, 1979, 37, 32-336. Авторское право © 1979 American Psychological Association. Переиздано в соответствии с разрешением.

Полные благих намерений участники, для распределения ответственности за объединенные усилия, пытаются вспомнить вклад, который каждый из них внес в конечный продукт. Однако некоторые аспекты взаимодействия можно вспомнить легче, или они являются более доступными, чем другие. Кроме того, особенности, которые легко вспоминаются, могут не быть случайным подмножеством целого. В частности, человек может вспомнить большую долю собственного вклада, чем смогут другие участники.

Эгоцентрическое предубеждение в доступности информации в памяти, в свою очередь, может служить причиной необъективного приписывания ответственности за совместный результат. Как продемонстрировали Тверски и Канеман (Tversky и Kahneman, 1973, 11), люди используют *доступность* (“легкость, с которой соответствующие случаи приходят на ум”) как основание для оценки частоты (1973, стр. 209). Таким образом, если входные данные, воспроизведенные участником самостоятельно, были бы доступны, люди тогда принимали больше ответственности за совместный результат, чем другие участники приписывали им.

Существует, по крайней мере, четыре процесса, направленных на увеличение доступности собственных вкладов: (а) выборочное кодирование и хранение информации, (б) дифференциальная выборка, (с) информационные различия и (д) мотивационные влияния.

Выборочное кодирование и хранение

По ряду причин, доступность собственных входных данных человека может быть облегчена дифференциальным кодированием и хранением самопроизведенных реакций. Во-первых, собственные мысли людей (относительно того, что они собираются сказать, фантазии т.д.) или действия могут отвлекать их внимание от вкладов других. Во-вторых, люди могут имитировать или повторять свои идеи или действия; например, они могут обдумывать свою позицию перед тем, как ее высказать и защитить. Следовательно, их собственные входные данные получают больше “времени на изучение”, а степень сохранения тесно связана со временем на изучение (Carver, 1972). В-третьих, вклады людей легче встраиваются в их когнитивную схему, то есть в уникальное понимание проблемы, основанное на прошлом опыте, ценностях и так далее. Вклады, которые вписываются в такие существовавшие ранее схемы, сохраняются с большей вероятностью (Bartlett, 1932; Bruner, 1961).

Дифференциальное воспроизведение

Предубеждение доступности может также являться следствием дифференциального воспроизведения информации из памяти. При распределении ответственности за исход совместной деятельности важный вопрос для каждого участника может звучать так: “Каков *мой* вклад?” Участники могут

попытаться вспомнить собственные вклады и неуместно использовать информацию, полученную таким образом, чтобы оценить их *относительные* вклады, однако подобная оценка не может быть сделана должным образом без рассмотрения входных данных других людей.

Информационные различия

Вероятно, существуют различия в доступной для участников информации, что может спровоцировать эгоцентрическое воспоминание. Люди имеют больший доступ к их собственным внутренним состояниям, мыслям и стратегиям, чем наблюдатели. Кроме того, участники общего дела могут отличаться по знанию частоты и значения независимых вкладов друг друга. Например, супервизоры факультета могут знать меньше о количестве времени, усилиях или изобретательности, которую студенты вкладывают в выполнение заданий, проведение анализов данных и написание предварительных проектов работы, чем сами студенты. С другой стороны, супервизоры более осведомлены о количестве и о важности размышлений, чтения и так далее, которые они включают в исследование прежде, чем студенты начнут его выполнять.

Мотивационные влияния

Мотивационные факторы могут также привести к эгоцентрическому предубеждению о доступности. Чувство собственного достоинства может увеличиться, если сосредотачиваться или придавать больше значения, собственным вкладам. Точно так же озабоченность собственной оперативностью или контролем может привести людей к тому, чтобы уделить больше внимания собственным вкладам в совместное дело (см. deCharms, 1968; White, 1959).

Предшествующее обсуждение выделяет множество процессов, которые могут работать на то, чтобы представить собственные вклады более доступными (и легкими для воспоминания), чем вклады других. Следовательно, трудно опровергнуть гипотезу, согласно которой воспоминания и приписывания являются эгоцентрическими. Как заметил Гринвалд (Greenwald, 1978), эгоцентрический характер памяти “не является неминуемой правдой. Можно предпринять попытку упорядочивания прошлого опыта, например, провести сопоставительную работу, такую как написание текста по истории или индексирование тезауруса” (с. 4). Кроме того, нам не удалось найти опубликованные данные, прямо подтверждающие предполагаемое предубеждение доступности. Наконец, недавние явления в литературе, посвященной участникам и наблюдателям, кажется, противоречат гипотезе, согласно которой воспоминания и приписывания являются эгоцентрическими. Джонс и Нисбетт (Jones и Nisbett, 1971) считали, что участники склонны видеть причину своего поведения в окружающей среде, в то время как наблюдатели приписывают то же самое поведение устойчивым чертам ха-

рактера, которыми обладают участники. Хотя разнообразие объяснений этого эффекта было предложено к рассмотрению (Jones и Nisbett, 1971), недавний акцент был поставлен на перцептивной обработке информации (Storms, 1973; Taylor и Fiske, 1975). Визуальные рецепторы участника нацелены на окружающую среду; наблюдатель может сосредотачиваться непосредственно на нем. Таким образом, различные аспекты ситуации существенны для участников и для наблюдателей, неравенство, которое отражено в их причинных приписываниях. Это предположение, кажется, противоречит положению, что участники во взаимодействии с другими людьми в значительной степени погружены в себя.

Два исследования предлагают подтверждение существующей гипотезы. Роджерс, Куипер и Киркер (Rogers, Kuiper, Kirker, 1977) показали, что черты характера, выраженные прилагательными, вспоминались легче, если испытуемых просили соотнести их с собой (т.е. описывала ли каждая черта их), чем когда их просили о множестве других оценок (например, оценки синонимичности). Эти данные подразумевают, что возможность соотнесения с собой увеличивает доступность. Однако, Роджерс и др. не противопоставляли вспоминание прилагательных, относящихся к себе с вспоминанием прилагательных, относящихся к другим – сравнение, которое было бы более подходящим для текущего обсуждения. Гринвальд и Альберт (Greenwald, Albert, 1968) обнаружили, что люди вспоминали, свои собственные высказывания по определенному вопросу более точно, чем письменные высказывания других испытуемых. Так как высказывания самого человека и других людей отражали всегда противоположные точки зрения, Гринвальд и Альберт подтвердили большую осведомленность и лучшее запоминание аргументов, совпадающих с собственной позицией, чем запоминание само-произведенных утверждений, хотя свидетельство для изучения предубеждениям сомнительно, например, Greenwald и Sakumura, 1967; Malpass, 1969.

Мы провели экспериментальное исследование, чтобы определить, можно ли получить подтверждение предполагаемого предубеждения доступности. Студентов старших курсов, посещающих семинар, попросили оценить, сколько минут каждый участник семинара говорил в течение прошедшего занятия. Дополнительно были отобраны 26 испытуемых из групп по два человека, к которым подходили в кафетериях и залах. Участников в этих группах попросили оценить процент полного времени, в течение которого каждый человек говорил в течение текущего взаимодействия.

Предполагалось, что испытуемые будут основывать свои оценки времени на частях беседы, которые они могли с легкостью вспомнить. Таким образом, если имеется предубеждение в пользу вспоминания собственных утверждений, оценки человеком количества времени, в течение которого он сам говорил, должны превзойти среднее время говорения, приписанное ему другими членами группы.

Результаты подтвердили предположения. Для семи из восьми студентов на семинаре старшекурсников, оценки их собственного времени говорения превысили среднюю оценку времени, приписанную им другими участниками ($p < 0.05$, тест знаков). Точно так же оценки собственного времени разговора у 10 из 13 пар превысили оценки времени, данные другими участниками ($p < 0.05$, тест знаков). Величина предубеждения была существенной для 13 пар, $F(1, 12) = 14.85$, $p < 0.005$; в среднем участники оценили, что они говорили 59% времени. Эти данные обеспечивают предварительные, хотя и косвенные, подтверждения предполагаемого предубеждения доступности в ежедневных ситуациях...

Эксперимент 1

В этом эксперименте, мы хотели исследовать эгоцентрические предубеждения в естественно складывающихся, длительных отношениях. Женатые пары представляли идеальную целевую группу. Супруги участвуют во многих совместных делах различной важности. Это обстоятельство должно предоставлять возможности для эгоцентрических предубеждений.

Соответственно, первый эксперимент проводился (а) чтобы определить, происходят ли эгоцентрические предубеждения в распределениях ответственности в брачных отношениях; (б) чтобы воспроизвести, используя различную меру зависимости, эгоцентрическое предубеждение доступности, полученное в предварительных испытаниях; и (с) чтобы соотнести предубеждение доступности с предубеждением в ответственности. Если предубеждение в ответственности вызвано предубеждением доступности, два набора данных должны быть связаны.

Метод

Испытуемые: 37 супружеских пар, живущих в студенческих общежитиях. Двадцать пар имели детей. Испытуемые были привлечены двумя научными сотрудницами, которые обходили комнаты общежитий и кратко описывали эксперимент. Если пара хотела участвовать, назначалась встреча. Исследование проводилось в квартире пары; каждой паре было оплачено 5\$ за участие.

Процедура. Анкетный опрос был создан на основе обширных предварительных интервью с шестью супружескими парами. В эксперименте, анкетный опрос был выполнен индивидуально мужем и женой; анонимность была гарантирована. На первых страницах анкетного опроса испытуемых просили оценить степень своей ответственности за каждое из 20 действий, соответствующих супружеским парам, проводя косую линию через 150-миллиметровую прямую линию, концы которой были помечены “прежде всего жена” и “прежде всего муж”.¹ Эти двадцать действий были: приготовление завтра-

¹ В предварительных интервью, мы использовали процентные оценки. Мы обнаружили, что тестируемые были способны помнить проценты, которые они записали и что сравнения постав-

рака, мытье посуды, уборка, покупки продовольствия, присмотр за детьми, планирование совместного отдыха, решение, на что должны быть потрачены деньги, решение, где жить, выбор друзей, принятие важных решений, которые затрагивают Вас обоих, инициатор споров между Вами, разрешение конфликтов, причина беспорядка в доме, стирка, поддержание контакта с родственниками, проявление привязанности к супругу (супруге), вынос мусора, раздражение супруга (супруги), ожидание супруга (супруги), принятие решения, иметь ли детей.

Испытуемых затем просили сделать кратко запись примеров своих вкладов или вкладов супруга (супруги), сделанные в каждый вид деятельности. Их письменные отчеты были впоследствии исследованы, чтобы оценить, были ли собственные входные данные человека более “доступны”. То есть, имеют ли тенденцию примеры, о которых сообщили испытуемые, сосредотачиваться больше на их собственном поведении, чем на поведении их супруги (супруга)? Наблюдатель, не принимающий во внимание экспериментальную гипотезу, сделал запись количества различных примеров, в которых испытуемые сообщили о своих собственных вкладах и вкладах супруга (супруги). Другой наблюдатель закодировал одну треть данных; надежность (коэффициент корреляции Пирсона) составила 0.81.

Результаты

Ответы обоих супругов на каждый из вопросов об ответственности были суммированы так, чтобы общее количество включило то значение, которое жена рассматривала как ее вклад, и то количество, которое муж рассматривал как его вклад. Так как шкала ответа была 150 миллиметров длиной, было размещено 150 “единиц ответственности”. Сумма больше, чем 150 указывала на эгоцентрическое предубеждение в кажущемся вкладе, — то, что, по крайней мере, один из супругов переоценивал свою ответственность за эту деятельность. Чтобы оценить степень пере- или недооценки, которую этот супруг показал для каждого вида деятельности, 150 вычиталось от общей суммы каждой пары. Общий счет был получен для пары в среднем по 20 действиям (или 19, когда у пары не было детей).

Дисперсионный анализ, использующий пару как единицу анализа, показал, что общий счет был значительно больше нуля, $M = 4.67$, $F(1, 35) = 12.89$, $p < 0.001$, указывая эгоцентрическое предубеждение в воспринятых вкладах. Двадцать семь из 37 пар показали некоторую степень переоценки ($p < 0.025$, тест знаков). Более того, в среднем, переоценка произошла в 16 из этих 20 пунктов в анкетном опросе, включая отрицательные пункты — например, причина споров между вами, $F(1, 32) = 20.38$, $p < 0.001$. Хотя величина переоценки была относительно маленькой, в среднем, испытуемые имели тенденцию использовать ограниченный диапазон шкалы. Большинство ответов были слегка выше или слегка ниже середины шкалы. Ни один пункт не показал существенный эффект недооценки.

кетных процентов явились источником конфликтов между супругами. Использование 150-миллиметровой шкалы обошло эти трудности; тестируемые не были склонны, преобразовывать свои косые линии в точные проценты, которые могли обсуждаться в дальнейшем.

Второй набор пунктов в анкетном опросе требовал от испытуемых записать примеры их собственных вкладов и вкладов их супруга (супруги) в каждый вид деятельности. Средний счет различия был получен по этим 20 видам деятельности (в среднем муж и жена), с количеством примеров вкладов супругов вычтенных из количества примеров собственных вкладов. Тест генеральной средней был существенен, $F(1, 35) = 36.0$, $p < 0.001$; как и ожидалось, испытуемые привели большее количество примеров своих собственных вкладов ($M = 10.9$), чем вкладов супруга(супруги) ($M = 8.1$). Отношение между разницей счета сам-другой и изначальной мерой воспринятой ответственности была определена. Как было предположено, чем больше тенденция вспоминать свое релевантное поведение, тем больше была переоценка воспринятой ответственности, $r(35) = 0.50$, $p < 0.01$.

Количество слов, содержащихся в каждом поведенческом примере, о котором сообщили испытуемые, было также оценено, чтобы обеспечить степень проработки или полноту вспоминания. Среднее количество слов в примере не выделяется как функция того, породилось ли данное поведение самим собой ($M = 10.0$) или супругом(супругой) ($M = 10.1$), $F < 1$. Далее, эта мера была некоррелирована с мерой воспринятой ответственности, $r(35) = -0.15$, не значимо.

Итак, как мера ответственности, так и мера, отражающая доступность соответствующего поведения, показала предполагаемые эгоцентрические предубеждения. Кроме того, имеется существенная корреляция между величиной предубеждения в доступности и величиной предубеждения в ответственности. Этот вывод согласуется с гипотезой, согласно которой эгоцентрические предубеждения в приписываниях ответственности опосредованы предубеждениями доступности. Наконец, количество воспроизведенных актов поведения было важным фактором, более чем богатство вспоминания....

Эксперимент 2

В эксперименте 2 мы попросили игроков из 12 университетских сборных баскетбольных команд индивидуально заполнить анкетный опрос, в котором было необходимо вспомнить важный поворотный момент в их последней игре и оценить, почему их команда выиграла или проиграла.

Это возможность уйти от сравнений сам-другой, которые мы рассмотрели в предыдущих исследованиях, к сравнению команда - другая команда. Существует, однако, множество причин ожидать, что действия собственной команды должны быть более доступны наблюдателям, чем действия другой команды: я знаю имена моих товарищей по команде, и поэтому, у меня есть готовые средства для организации хранения и выборки данных, соответствующих им; наш успех в будущих играх с другими противниками зависит больше от наших собственных наступательных и защитных способностей, чем от способностей противостоящей команды. Следовательно, я

могу внимательнее следить за действиями моих товарищей по команде, что облегчает кодирование и хранение. Также, имеются информационные различия: стратегии собственной команды более существенны, чем стратегии противостоящей команды (Tversky и Kahneman, 1973, 11). Если инициативы собственной команды дифференциально доступны, игроки должны вспомнить поворотный момент в терминах действий своей команды и приписать ответственность за исход игры своей команде....

Метод

Испытуемые. В исследовании участвовали игроки университетских сборных команд по баскетболу: 74 женского пола и 84 мужского пола. С менеджерами команды связались по телефону; после обсуждений с игроками, они разрешили командам участвовать в исследовании.

Процедура. Анкетные опросы проводились после шести игр, в которых команды, участвующие в исследовании, играли друг с другом. Таким образом, для трех выбранных игр с мужскими командами, три из шести мужских команд в исследовании играли против других трех мужских команд. Точно так же три выбранные игры между женщинами, включали все шесть женских команд. Анкетные опросы проводились при первой практике команды, последовавшей за целевой игрой (1 или 2 дня после игры), кроме одного случая, где из-за расписания игр команд, было необходимо собрать данные немедленно после игры (две женские команды). Анкетные опросы были проведены индивидуально, и анонимность испытуемых была гарантирована. Следующие вопросы были заданы:

1. Пожалуйста, кратко опишите один важный поворотный момент в прошлой игре и обозначьте, в каком периоде игры он произошел.

2. Наша команда выиграла / проиграла прошлую игру потому что...

Ответы на первый вопрос были исследованы, чтобы определить, был ли поворотный момент описан как созданный действиями собственной команды, обеих команд, или другой команды. Ответы на второй вопрос были исследованы, чтобы оценить количество причин победы или поражения, которые имели отношение либо к собственным действиям, либо к действиям другой команды. Данные были закодированы человеком, не знавшим экспериментальной гипотезы. Второй наблюдатель независимо закодировал ответы 50% испытуемых. Имелось соответствие 100 % для обоих вопросов.

Результаты

Не было обнаружено никаких существенных половых различий в двух зависимых измерениях, поэтому о результатах сообщается независимо от пола. Так как ответы членов команды не могут рассматриваться как независимые, было подсчитано среднее значение ответов, и команда, служила как единица анализа.

Предварительная проверка данных “поворотного момента” показала, что даже в пределах команды игроки вспоминали весьма различные события. Однако 119 игроков вспомнили поворотный момент, который они описали как созданный действиями их собственной команды; 13 игроков вспомнили поворотный момент, который они рассмотрели, как вызванный обеими командами; 16 игроков вспомнили, поворотный момент, который, как им казалось, инициирован действиями противоположной команды (оставшиеся 10 игроков не ответили на вопрос). Испытуемые описали такие события как сильную защиту в течение последних 2 минут игры, изменение в наступательных стратегиях, и так далее.

Процент игроков, которые вспомнили поворотный момент, как вызванный членами их команды, был получен для каждой команды. Эти 12 множеств были подвергнуты анализу, в котором они сравнивались со случайным ожиданием 50%. Полученное распределение сильно отличалось от случайного, $F(1, 11) = 30.25$, $p < 0.001$, со средним 80.25%. Как и было предположено, большинство сообщений выделяли действия собственной команды.

Процент игроков, которые вспомнили поворотный момент, как вызванный членами их команды, был исследован относительно игры команды. Средний процент был выше у проигравшей команды, чем у победившей в пяти из шести игр ($p < 0.11$, тест знаков). Средняя разница между процентами проигравших ($M = 88.5$) и выигравших ($M = 72$) команд была незначимой ($F < 1$).

Были также исследованы объяснения игроков относительно победы или проигрыша их команды. Из 158 участников, только 14 указали причины, которые включали действия противостоящей команды. В среднем, испытуемые сообщили о 1.79 причинах, касающихся собственной команды и 0.09 причин, касающихся противоположной команды, $F(1, 11) = 272.91$, $p < 0.001$. Наконец, тенденция приписывать больше причин для собственной команды была незначительно большая после проигрыша ($M = 1.73$), чем после победы ($M = 1.65$), $F < 1$.

Обсуждение

Ответы на вопрос о поворотном моменте указывают, что действия товарищей по команде испытуемых были более доступны, чем действия членов противостоящей команды. Далее, испытуемые приписывали ответственность за исход игры действиям или бездействию скорее своих товарищей по команде, чем членов противоположной команды. Таким образом, предубеждения в доступности и оценки ответственности могут происходить на уровне групп. Ратер и Хесковиц (Rather, Heskowitz, 1977) приводят другой пример группового эгоцентризма: “CBS (новости) стали лидером после запуска на Луну Апполона в 1968. Если Вы – сотрудник CBS, Вы склонны считать, что ваше описание приземления на Луну перевернуло наше восприятие. Если

Вы – сотрудник NBC, Вы склонны рассматривать раскол команды Хантли-Бринкли(Huntley-Brinkley) как ключевой фактор “ (с. 307)....

Эксперимент 3

В эксперименте 3, мы попытались изменить центр внимания человека, чтобы затронуть доступность. Мы использовали манипуляцию, предназначенную для активизации выборочного исправления информации, непосредственно соответствующей приписываниям ответственности.

В нашем первоначальном анализе, мы предположили, что причиной эгоцентрических приписываний ответственности могло быть выборочное воспроизведение информации из памяти и что на воспроизведение можно повлиять с помощью вопросов, которые люди задают самим себе. Эксперимент 3 проводился для проверки этой гипотезы. У испытуемых вызывались разные виды воспроизведения с помощью изменения формулировок вопросов. Выпускников попросили подумать либо об их собственных вкладах в их дипломы на звание бакалавра гуманитарных наук, либо о вкладах их научных руководителей. Оценивать степень ответственности за научную работу, которую испытуемые приписали либо себе, либо научным руководителям, было оценено. Предположили, что испытуемые примут меньше ответственности за исследуемую проблему при условии направленности на научного руководителя, чем при условии направленности на себя.

Метод

Испытуемые. 17 выпускников факультета психологии женского пола и 12 мужского. Большинство закончили 1 или 2 года аспирантуры. Все студенты провели эксперименты, ставшие основанием для диплома на звание бакалавра гуманитарных наук на последнем году обучения.

Процедура. С испытуемыми общались индивидуально в их офисах и просили заполнить краткий анкетный опрос на отношение студент-научный руководитель. Ни один не отказался участвовать. Две формы анкетного опроса были наугад распределены среди испытуемых; анонимность и конфиденциальность были гарантированы.

В одной из форм анкетного опроса испытуемых просили указать свой собственный вклад в каждое из множества действий, связанных с их бакалаврскими дипломами. Вопросы были следующие: (а) “Я предложил __ процентов методологии, которая в конечном итоге использовалась в исследовании.” (В) “Я обеспечил _ процентов интерпретации результатов.” (С) “Я был инициатором __ процентов обсуждений, касающихся диплома, с моим руководителем.” (D) “В течение связанных с дипломом обсуждений я был склонен управлять курсом и содержанием обсуждения __ процентов времени.” (Е) “Учитывая все, я был ответствен за ___ процентов всех усилий, приложенных к исследованию.” (F) “Как бы Вы оценили Ваш диплом относительно других дипломов, сделанных на факультете? “

Вторая форма анкетного опроса была идентична вышеупомянутой, за исключением

того, что слово *Я* (условие направленности на себя) было заменено на *мой руководитель* (условие направленности на руководителя) в вопросах 1-5. Испытуемых попросили заполнить бланки с ответами на первые пять вопросов и провести косую черту через 150-миллиметровую линию, с концами, помеченными “хуже” и “лучше”, в ответ на вопрос 6.

Результаты и обсуждения

С целью анализа, было принято, что вклад руководителя и студента в каждый пункт составляет в целом 100%. Хотя эксперимент был представлен как изучение отношения студент-руководитель, возможно, что студенты рассматривали в своих оценках входные данные других людей (например, других студентов). Однако данная процедура обеспечивает умеренное испытание экспериментальной гипотезы. Например, если тестируемый отвечал 20% на пункт в “Я”-версии анкетного опроса, считалось, что его или ее руководитель внес вклад 80%. Руководитель, возможно, внес только 60%, а неизвестный человек остальные 20%. Возможно переоценивая вклад руководителя, однако, мы смещаем данные против экспериментальной гипотезы: “Я”-версия, как ожидалось, уменьшит процент ответственности, приписываемой руководителю.

Ответы испытуемых на первые пять вопросов в “Я”-версии анкетного опроса вычитались из 100 так, чтобы более высокие числа отражали большие вклады руководителя в обоих условиях. Вопрос 5 имел дело с полной ответственностью за исследование. Как и ожидалось, испытуемые приписали больше ответственности руководителю при условии направленности на руководителя ($M = 33.3\%$), чем при условии направленности на себя ($M = 16.5\%$), $F(1, 27) = 9.05$, $p < 0.01$. Первые четыре вопроса касались различных аспектов диплома, и средний ответ показал похожие результаты: направленность на руководителя $M = 33.34$; направленность на себя $M = 21.82$; $F(1, 27) = 5.34$, $p < 0.05$. Наконец, испытуемые имели тенденцию оценивать свой диплом более положительно при условии направленности на себя, чем при условии направленности на руководителя: 112.6 против 94.6, $F(1, 27) = 3.59$, $p < 0.10$.

Контрастирующая формулировка вопросов имела ожидаемое воздействие на распределения ответственности. Версия анкетного опроса, касающаяся руководителя, возможно, заставила испытуемых вспомнить, большую часть вкладов их руководителей, чем “Я”-версия анкетного опроса. Эта дифференциальная доступность была отражена в распределениях ответственности. Обратите внимание, что вопросы не поддавались контролю испытуемых за счет исправлений. Научному руководителю приписывалась только третья часть ответственности за диплом в условии направленности на руководителя.

В свете существующих данных, приписывания баскетболистов ответственности за исход игры в эксперименте 2 должны были быть заново исследованы. Вспомните, что игроков просили закончить предложение, “Наша

команда выиграла / проиграла последнюю игру потому что...” Этот вопрос выдал высоко значимое эгоцентрическое предубеждение. Оценивая прошлое испытание, очевидно, что форма вопроса - “*Наша* команда ... *нашу* прошлую игру” — , возможно, побудила испытуемых сосредоточиться на действиях их собственных команд, даже при том, что формулировка не устраняет ссылки на противоположную команду. Вопрос про “поворотный момент” в эксперименте 2 был более нейтрально сформулирован и не может интерпретироваться двояко.

Главные вопросы в этих исследованиях исходят из внешнего источника; многие из наших вопросов исправления само-инициированы, однако, и наше вспоминание может вполне быть смещено формой, в которой мы самим себе излагаем вопросы исправления. Например, баскетболисты, скорее всего, будут думать в терминах “Почему *мы* выигрывали или проигрывали?” “Чем в терминах, сформулированных нейтрально “Какая команда была ответственной за исход игры?”...

Настоящее исследование демонстрирует распространенность центрированных на себе (self-centered) предубеждений доступности и оценок ответственности. В повседневной жизни, эти эгоцентрические тенденции могут не замечаться, когда объединенные усилия не требуют явных распределений ответственности. Если распределения обозначены отчетливо, однако, появляется повод для разногласия, и люди, вряд ли, поймут, что различия в оценках могут исходить из оценок информации, которая является дифференциально доступной.

13. Предубеждения доступности в социальном восприятии и взаимодействии*

Шелли И. Тейлор

Каждый день человек делает многочисленные, сложные социальные оценки – прогнозирование поведения других людей, приписывание ответственности, отнесение человека к категории, оценка других людей, оценка власти, влияния человека или причинную атрибуцию. Главная задача социальной психологии состоит в том, чтобы определить, как субъект делает эти оценки. До недавнего времени, исследование этой темы было отмечено рационалистичным предубеждением, предположением о том, что оценки делаются, используя полные, оптимальные стратегии (см., например, Fischhoff, 1976, для обсуждения этой темы). Ошибки в оценках были приписаны двум источникам: (а) случайные ошибки из-за проблем с информацией, о которой человек не знал; и (б) ошибки, которые являлись следствием иррациональных мотивов и потребностей человека.

В пределах социальной психологии эта перспектива представлена исследованием причинной атрибуции. В ранних формулировках атрибуции (например, Jones & Davis, 1965; Kelley, 1967) человек был охарактеризован как неопытный ученый, который собирал информацию из множества источников окружающей среды, чтобы делать приписывания относительно причинно-следственных отношений. Когда наблюдались отклонения от нормативных моделей, считалось, что они происходят от предубеждений, таких как гедоническая релевантность (Jones & Davis, 1965) или других эгоцентрических потребностей (см. Miller & Ross, 1975).

Однако в течение многих лет, все большее количество данных доказывало не только то, что оценки и решения людей менее закончены и рациональны, чем считалось, но что не все ошибки могут исходить из мотивационных факторов. Даже при отсутствии мотивов, оценки часто делались на основании скудных данных, которые, по-видимому, были случайно объединены и

* Подготовка части книги была поддержана Исследовательским Грантом Национального Научного Фонда № BNS 77-09922

находились под влиянием предвзятых мнений (см., например, Dawes, 1976). Эти выводы привели к пересмотру когнитивной системы. Люди стали рассматриваться как обладающие ограниченными возможностями, способными к оперированию только небольшим количеством данных одновременно. В отличие от “наивного” ученого, который оптимизирует, человек, как считается, “сатисфицирует” (Simon, 1957) и использует сокращения, которые приведут к решению и оценке эффективно и точно.

Один из наиболее провокационных и влиятельных вкладов в пересмотр процесса оценивания – это работа Канемана и Тверского о когнитивной эвристике (Kahneman и Tversky, 1973, 4; Tversky и Kahneman, 1974, 1). Согласно Тверскому и Канеману (1974, 1), эвристика используется в условиях неопределенности, или недоступности, или неоднозначности важной информации. В не-социальных оценках неопределенность вытекает, прежде всего, из факта, что информация, соответствующая отдельно взятой оценке, является почти всегда неполной. Соответствующий фактический материал может быть недоступен, он не может быть собран вовремя, чтобы влиять на решение, или может быть слишком объемным, чтобы быть должным образом организованным и использоваться в задаче оценки.

Эвристика – это вероятные стратегии для создания социальных, а также несоциальных оценок по нескольким причинам. Во-первых, различие между социальными и несоциальными оценками условно, фактически любая существенная оценка имеет социальные последствия. Во-вторых, социальные оценки включают те же виды неопределенности, которые характеризуют несоциальные оценки. В-третьих, социальные оценки включают новые источники неопределенности. Информация относительно людей более неоднозначна, менее надежна и более непостоянна, чем информация относительно предметов или несоциальных событий, так как личные признаки не написаны на лицах людей, как у предметов обозначен их цвет, форма или размер. Таким образом, личные признаки скорее должны быть выведены, чем наблюдаемы непосредственно. У людей есть намерения, не все из которых непосредственно обозначены. Учитывая, что наиболее существенные социальные действия могут совершаться по различным причинам и приведут к различным последствиям, значение социального действия по существу неоднозначно. Хотя предметы сохраняют свои признаки независимо от ситуации и времени, мотивы людей меняются от ситуации к ситуации, и цели меняются как с минуты на минуту, так и в течение жизни; таким образом, даже правильный вывод в одной ситуации может быть бесполезен с точки зрения прогнозов. Отсутствие полной, надежной, прогнозирующей информации относительно людей и социальных взаимодействий предполагает, что люди используют эвристики, позволяющие им делать выводы и предсказания, отталкиваясь от тех скудных и ненадежных данных, которые являются доступными.

Эвристика доступности

Одна из таких эвристик – доступность. “Считается, что человек, использует эвристику доступности всякий раз, когда он оценивает частоту или вероятность посредством легкости, с которой случаи или ассоциации приходят на ум” (Tversky и Kahneman, 1973, с. 208, 11). Человек предполагает, что, если примеры приходят на ум быстро, то их должно быть много, или если ассоциация проделывается легко, тогда она должна быть точной, так как ассоциативные связи приходят с опытом. Кроме того, именно *легкость* воспоминания, построения и ассоциации, обеспечивает оценку частоты или вероятности, а не общая сумма примеров или ассоциаций, которые приходят на ум. Таким образом, важная разница между использованием эвристики доступности и использованием более сложного процесса – то, что производится мало поправок; оценка легкости, с которой этот процесс может быть выполнен, является достаточной как основание для вывода.

Тверски и Канеман (1973, 11) предлагают два общих класса задач, в которых предубеждение доступности могло бы заметно фигурировать: построение случаев и ассоциаций и выборка ассоциаций и случаев. Эти две общие задачи – также задачи субъекта. При некоторых обстоятельствах нас могут попросить воспроизвести социальное поведение, как в попытке угадать, как друг поведет себя, узнав, что жена от него уходит. В таких случаях каждый конструирует социальную действительность, с которой данный случай может быть сравнен. В других случаях, человек может основываться на прошлых примерах поведения, чтобы сделать выводы, например, он может вспомнить случаи, как друг справлялся с кризисными ситуациями в прошлом, чтобы спрогнозировать, как он справится теперь.

В некоторой степени предположения относительно отношения между легкостью воспроизведения и количеством примеров или ассоциаций, точны, и настолько, насколько они точны, человек, использующий эвристику доступности, придет к правильным выводам или, по крайней мере, к тем, которые будут соответствовать выводам, достигнутым, при использовании более сложных процедур. При других обстоятельствах, однако, эти выводы могут быть неточными, потому что существуют предубеждения в доступных данных, которые имеют отношение к задаче. Существует, по крайней мере, три способа, с помощью которых предубеждения в доступных данных могут смещать следующие друг за другом социальные процессы. Во-первых, только существенные данные могут быть более доступны и, следовательно, оказывать непропорциональное влияние на процесс суждения. Во-вторых, предубеждения в самом процессе произхождения могут выдавать нерепрезентативную основу данных. В-третьих, стойкие когнитивные структуры человека, такие как убеждения и ценности, способствуют появлению предубеждений, усиливающих доступность определенных свидетельств, таким образом, смещая процесс оценки. Присутствие любого из таких предубеждений может привести к необъективным выводам.

Предубеждения доступности и заметности: пример

Предубеждения заметности проявляются в том, что красочные, динамические или другие отличающиеся стимулы непропорционально привлекают внимание, и соответственно влияют на оценки. Один из примеров такого предубеждения был назван *фундаментальной ошибкой атрибуции* (см. Ross, 1977), и он относится к распространяющемуся предубеждению видеть людей как причинных агентов в окружающей среде. То есть в социальном окружении, где либо человек, либо некоторая ситуативная переменная является вероятным причинным фактором для исхода, существует общее предубеждение, чтобы видеть людей как причинных агентов, особенно их устойчивые поведенческие признаки.

Некоторые люди более заметны, чем другие, и эта дифференциальная заметность в пределах социального окружения может также смещать процесс оценивания. Исследования, в которых применили гештальт-принципы фигурального акцента к социальному миру (см., например, McArthur и Post, 1977; Taylor и Fiske, 1975; Fiske и другие, 1979), показывают, что человек, который ярко освещен, перемещается (как в кресле-качалке), контрастен (с помощью таких тривиальных манипуляций, как цвет рубашки), или новичок привлекает к себе непропорциональное количество внимания.

Социальные последствия предубеждения заметности проиллюстрированы исследованиями, изучающими воздействие статуса соло или символической интеграции на впечатления людей о других людях. Когда компания собирается включить членов группы меньшинств, таких, как чернокожих, женщин или инвалидов, часто полной интеграции предшествует промежуточный шаг. На этом этапе, один или два члена этой группы, могут быть приняты в то, что иначе было бы группой белых рабочих мужского пола, таким образом, создав примеры статуса соло. Статус соло может возникать по нескольким причинам. Отсутствие квалифицированных претендентов от отдельно взятой группы меньшинств; организация желает избежать угрозы большого притока членов группы меньшинства; или статус соло может использоваться, чтобы отразить силы конструктивного действия. Независимо от причин для символической интеграции, к символу или человеку-соло часто относятся как к представителю его социальной группы. Соответственно, оценки его поведения часто используются, чтобы предсказать, как другие члены этой группы вели бы себя, если бы они также были приняты в организацию. Значение статуса соло – это его новизна. Женщина выделяется среди мужчин, также как чернокожий выделяется среди белых людей. Такая отчетливость способствует предубеждению заметности.

В экспериментальном аналоге этой ситуации, испытуемые просмотрели видеофильм о маленькой группе (шесть человек), беседующих в неформальной обстановке. Некоторым испытуемым показали группу, в которой был один афро-американец, один мужчина или одна женщина в среди белокожих, в женской, или мужской компании, соответственно (Taylor и другие, 1976;

Taylor, Fiske, Etcoff и Ruderman, 1978). Другие испытуемые наблюдали группы сравнений, где содержание обсуждения в группе было так же, как и в условии соло, но пол, или расовый состав группы был уравнен (например, три мужчины, три женщины). Испытуемые наблюдали обсуждение в группе и затем делали запись своих впечатлений о людях в группах.

Согласно положению о том, что существует предубеждение значимости, созданное статусом соло или символическим статусом, поведение чернокожего-соло вспоминалось несколько лучше, чем поведение этого же человека в сопоставимой смешанной группе, и было также оценено, что человек-соло говорил больше по сравнению с соответствующим человеком в смешанной группе.¹ Это предубеждение значимости, в свою очередь, приводит к крайности оценок. Когда человек был соло в группе, его поведение было оценено более экстремально в любом положительном или отрицательном направлении, по сравнению с тем же самым поведением в смешанной группе. Неприятный человек был воспринят как еще более неприятный в ситуации соло; хороший человек был воспринят как более хороший в ситуации соло.

Используя эвристику доступности как объяснительную структуру, можно вывести, что, когда испытуемых просили оценить человека в группе, они пытались приблизиться к примерам соответствующего поведения или ассоциациям с людьми стимула, и легкость, с которой такие случаи или ассоциации приходили в голову, приводили их к крайним суждениям. В случае с соло, доступно больше примеров соответствующего поведения, потому что имеется большая база данных, ведущая к крайности оценок. Эти результаты предполагают, что выделяющийся человек, будь то соло, неполноценный человек, или просто человек, отличающийся от людей, с которыми он взаимодействует, вызовет крайние оценки у окружающих. Значение этих результатов для социального мира, конечно, огромно. Например, если люди-соло используются как основание для включения в группу представителей меньшинств, и восприятие соло необъективно, то могут быть приняты неправильные решения. Преувеличенно отрицательные оценки человека-соло могут привести к необоснованному завершению десегрегации. Преувеличенно положительные оценки могут стать основанием для ложных ожиданий относительно поведения других членов группы меньшинства, ожидания, которые могут не оправдаться.

Мотивация не может служить объяснением предубеждения в восприятии значимых людей. Хотя некоторые мотивационные процессы могут определять отношение к инвалидам или людям-соло, трудно найти их в оценке человека, который отличается цветом рубашки или движениями. Соответственно, нужно обратиться к когнитивным факторам, а эвристика доступности обеспечивает возможное объяснение этих эффектов.

¹ Не было собрано данных о воспоминании в исследованиях по мужчинам- и женщинам-соло.

Доступность и предубеждения воспроизведения: Два примера

Другой источник предубеждений доступности происходит от того, как информация хранится или восстанавливается. То есть память имеет специфическую организацию, которая может облегчить процесс выбора или воспроизведения некоторых примеров или выводов и помешать появлению других. Одним из способов вмешательства является простое ограничение количества информации, которая может храниться в памяти, что может привести к беспорядку в ассоциативных связях или примерах, которые хранятся в памяти.

Ротбарт и его коллеги (Rothbart и другие, 1978) продемонстрировали эту проблему в исследовании признаков социальных групп. Испытуемым предложили информацию о характерных чертах членов гипотетической группы (например, Фил – ленивый) при одном из двух условий. Испытуемые либо видели имена некоторых членов группы (Эд, Фил, Фред, Джо) со специфической чертой (ленивый), либо они видели, что ту же самую пару: характер-имя (Фил ленивый) одинаковое число раз. Позже испытуемых попросили охарактеризовать группу в целом. Если испытуемые способны точно вспомнить, какие имена были соединены с какими чертами характера, то выводы относительно группы в целом должны быть более сильными, если несколько членов группы имеют некоторую черту, чем если только один член группы имеет эту черту. Если общее количество пар черт характер-имя, которые были предоставлены испытуемым, было низко, испытуемые показали это предостережение в своих выводах относительно группы. Однако когда общее количество пар черт характер-имя было высоко, испытуемым было трудно удержать в голове, сколько человек имели определенную черту характера, и они начинали вести себя, как будто многократные упоминания одних и тех же пар черт характер-имя были столь же информативны, как несколько различных имен, соединенные с этой чертой характера. Группа характеризовалась как ленивая даже, когда только несколько членов фактически были ленивы.

Социальный мир активен и часто очень информативен, и будучи таковым, он обычно отражает состояние большой загруженности памяти (Rothbart и др., 1978) больше, чем состояние малой загруженности памяти. Эти состояния облегчили бы получение информации о стереотипах группы, отталкиваясь от поведения только нескольких человек, чье поведение можно наблюдать большое количество раз. Как отмечают Ротбарт и другие, мы получаем больше информации от отрицательных, чем положительных событий, и соответственно все эти предубеждения могут поддерживать формирование отрицательных стереотипов группы, особенно, если членство группы существенно при упоминании в средствах информации.

Второй пример воздействия предубеждений выборки на социальные оценки – это эгоцентрическая атрибуция (Ross & Sicoly, 1979, 12). Во многих контекстах человек должен делать оценки относительно того, кто является ответственным за результат. Авторство книги должно быть определено, пла-

та за консультацию должна быть разделена среди сотрудников, или ответственность на победу группы, или поражение должна быть разделена. Кроме того, что каждый человек хранит список своих собственных вкладов и вкладов другого человека (стратегия некоторую отчаявшиеся пострадавшие были вынуждены принять), все стороны должны прийти к компромиссу относительно того кто что делал в совместном предприятии. Эвристика доступности обеспечивает потенциальную стратегию для этого. Можно перебирать примеры собственных вкладов и вкладов других людей и на основе того, сколько примеров приходит на ум, решать, кто сделал больше. Однако могут присутствовать предубеждения, мешающие точно оценивать. Одним из предполагаемых предубеждений является эгоцентрическое вспоминание, способность вспоминать собственные вклады несколько лучше, чем вклады другого человека. Это предубеждение вспоминания может, в свою очередь, производить предубеждения кажущейся ответственности. Например, если я могу вспомнить шесть случаев, когда я выносила мусор, и только три раза, как это делал мой муж, я могу сделать вывод, что на мне лежит ответственность за вынос мусора. В недавнем исследовании, распределение обязанностей в семье, оказалось особенно полезным, также как и накаленный контекст, в котором необходимо исследовать предубеждение доступности (Ross & Socoly, 1979, 12). Тридцать семь супружеских пар были опрошены относительно их вкладов в различные действия, которые женатые люди должны исполнять, такие как приготовление завтрака или мытье посуды. Каждого члена пары просили указать, был ли каждый вид деятельности выполнен прежде всего мужем или прежде всего женой. Каждый человек также указывал определенные примеры того, что он или супруга вложили в деятельность, измерение, предназначенное, чтобы выявить относительную доступность собственных вкладов против вкладов супруга(супруги).

Результаты ясно показали, что каждый супруг (супруга) думал что он вложил больше в совместные действия, чем другой супруг (супруга). Когда счет ответственности двух партнеров был сложен, он превысил уровень полной ответственности, которую можно было принять. Количество определенных примеров вкладов, которые каждый супруг (супруга) сделал, было также подсчитано, и результаты показали, что количество вкладов приписанных себе, намного превышает количество вкладов, приписанных супругу (супруге). Корреляция между запомненными примерами и приписыванием ответственности была высока (+ 0.51), предполагая, что предубеждение вспоминания могло уменьшить предубеждение ответственности.

Одно из возможных объяснений этих эффектов основывается на мотивационных принципах и предполагает, что люди приписывают себе больше, чем им полагается, чтобы сохранить или увеличить положительную самооценку. Если это правильно, то мы должны обнаружить, что при неудаче совместного проекта, люди будут отрицать свой вклад в нее и приписывать ответственность другим. Например, если супружеская пара делает покупку, которая, оказывается ненужной, каждый может винить другого за решение. Росс

и Сиколи (1979, 12) исследовали эту возможность и обнаружили, что это только слабо способствовало эгоцентрическим приписываниям. Люди так же вероятно принимали непропорциональную ответственность за неудачный совместный проект, как и за хороший.

Когнитивное объяснение этих эффектов основано на том, как воспроизведение или хранение информации порождают предубеждения кажущейся ответственности. Предубеждение близости может присутствовать, то есть каждый наблюдает собственный вклад более близко, чем вклады других, и, решая, кто за что отвечает, большее количество собственных вкладов приходит на ум. Например, если оба супруга работают в одно и то же время, каждый может отвлечься от вкладов другого и более тщательно наблюдать свои собственные вклады. С другой стороны, можно знать меньше о вкладах других людей, потому что каждый физически не присутствует, когда один из супругов делает свою долю работы; соответственно, можно недооценить количество времени и усилия, которые приложил супруг или супруга. К тому же, предубеждение может присутствовать, если один человек думал о своих вкладах больше, чем о вкладах другого. Это особенно вероятно, когда совместный проект включает много размышлений, письма и других видов длительной по времени работы, таких, как планирование капитального ремонта в доме или организация вечеринки. Третья возможность состоит в том, что собственные вклады подходят личным конструкциям или схемам, то есть способам обработки и кодирования информации. Информация, которая удовлетворяет существующей ранее схеме, вспоминается более легко, и таким образом, увеличенная доступность собственных вкладов может отражать предубеждение воспроизведения (см. Taylor & Crocker, 1979b). В итоге, предубеждения относительно того, как информация хранится или восстанавливается, могут привести к необъективным социальным оценкам. Один из возможных посредников – использование эвристики доступности.

Доступность и предубеждения, возникающие благодаря когнитивным структурам

Третий способ, за счет которого доступность может смещать социальные оценки – использование субъектом испробованных правил, схем или других когнитивных структур. Люди обладают определенными структурами для обработки поступающей информации, которые они часто используют и которые, постепенно, превращаются в познавательные привычки. Мы предполагаем, например, что люди будут использовать некоторые черты характера людей как способы организации информации о людях по другим чертам. Например, ученые часто различают людей по уровню интеллекта, принимая во внимание, что для спортсменов, скорее спортивный навык, чем интеллект является отличительным фактором. Этот вид предубеждения в тенденции использовать специфические схемы или конструкции, может также исказить оценки. Например, если кого-нибудь просят описать впе-

чатления об отдельно взятом человеке, можно использовать предпочитаемые черты (например, интеллект) как способ поиска в памяти и описывать человека не столько в терминах его фактического поведения, а скорее, в терминах предпочитаемых черт характера (см., например, D'Andrade, 1965). Подобное предубеждение может происходить в воспроизведении социального поведения, так как люди также используют конструкторы и схемы при столкновении с новыми ситуациями или когда их просят сделать прогноз относительно будущего. Например, ученый может прогнозировать способность друга выйти из тяжелой ситуации на основе того, насколько он умен, в то время как, спортсмен может прогнозировать способность того же самого друга выйти из ситуации на основе того, как быстро друг бегаёт. Соответственно, использование правил, схем и персональных конструкторов, может привести к выводам, что человек, который не разделяет те же самые когнитивные структуры, не справится.

Возможно, наиболее интригующий пример влияния долгосрочных схематических структур на восприятие данных обеспечивается явлением стереотипов. Ожидание, что человек поведет себя определенным образом, может вести к выводам, что человек уже повел себя так. Стереотипы – это особые виды ожиданий, которые могут работать таким образом, чтобы формировать действительность, и они могут делать это, по крайней мере, частично, с помощью предубеждения доступности. Гамильтон и Роуз (Hamilton and Rose, 1978), проверяли эту возможность в исследовании стереотипов. В одном исследовании, испытуемым предложили списки предложений, например, “Кэрл, библиотекарь, привлекательна и серьезна.” В каждом предложении представитель профессии был описан как обладающий двумя чертами. Некоторые черты носили стереотипную ассоциацию с профессией как, например, в случае черты “серьезный”, с профессией “библиотекарь”; в других случаях, черты не были стереотипными для этого занятия (как, например, “привлекательный” для библиотекаря), но они были стереотипными для другой профессии (как, например “привлекательный” для стюардессы). Во всех предложениях каждая черта была соединена с каждой профессией одинаковое число раз. Каждая черта была соединена со стереотипной профессией в одной трети случаев и с не-стереотипными профессиями в двух третьих случаев. Однако когда испытуемых просили оценить количество раз, когда каждая черта описывала представителя каждой профессии, они не верно вспоминали пары черта характер-профессия, отдавая предпочтение стереотипными ассоциациями. Например, они с большей вероятностью вспоминали, что библиотекари серьезны чем, что официантки серьезны.

Хотя мотивационные факторы могут предложить объяснение этих состояний, они, вряд ли, это сделают. Некоторые стереотипы могут заметно фигурировать в потребностях человека и целях, но стереотипы профессий, таких как официантка или библиотекарь, довольно банальны и вероятно, не фигурируют. Соответственно, эти результаты лучше понимаются как когнитивное явление. Эвристика доступности обеспечивает одно из возможных

объяснений, предполагая, что, когда испытуемых спрашивают, как часто черта характера и профессия соединены, они оценивают частоту, используя силу ассоциации между профессией и чертой; во многих обстоятельствах ассоциации сильнее после многих соединений в пары. Однако, в этом случае, имеется предубеждение в силе ассоциативной связи на основании стереотипной ассоциации между некоторыми из пар черта характера - профессия. Соответственно, так как эти ассоциации более сильные, воспринимающий оценивает, что эти черты характера и профессия более часто объединены чем, нестереотипные пары черта характера-профессия. Опять же, социальные значения этого вида предубеждения огромны. Стереотипы, когда-то сформированные, могут исказить сбор и хранение информационных и последующих впечатлений. Последствием является то, что необоснованные выводы относительно социальных групп или людей могут навсегда сохраниться без какого-либо эмпирического основания.

Выводы

Несколько прошедших десятилетий изменили точку зрения на оценки как результаты рационального, логического принятия решения, искаженного случайным присутствием иррациональных потребностей и мотивов к рассмотрению человека как пользователя эвристики. Эмпирическая работа над не-социальными оценками показывает, что субъект использует сокращения или эвристику, чтобы освободить место и передавать информацию настолько быстро, насколько возможно, и недавние исследования в области социальной психологии предполагают, что эти процессы также обращаются к формированию и использованию социальных оценок. Легкость, с которой примеры или ассоциации приходят в голову, обеспечивает оценку подобия (то есть, частоту или вероятность), которая в свою очередь обеспечивает основание для создания других социальных оценок, таких как оценки другого человека, причинная атрибуция или приписывание ответственности, описание признаков другого человека, отнесение других людей к категории или описание себя.

Наблюдательный критик обратит внимание, что в представленных исследованиях свидетельство для использования эвристики доступности является скорее логически выведенным, чем прямым, и можно требовать более ясного свидетельства, что именно легкость, с которой примеры или ассоциации приходят в голову, фактически опосредует оценки. Имеются несколько причин для неопределенности. Во-первых, в отличие от когнитивного исследования доступности, ни одно из этих социальных исследований не было предназначено для исследования доступности как таковой; скорее, каждое было предназначено, чтобы объяснить некоторое интересное социальное явление, и эвристика доступности применялась как одно из нескольких возможных объяснений явления. Во-вторых, не существует пока еще согласованного измерения доступности. Некоторые исследования исполь-

зовали скорость, с которой информация восстанавливается в памяти как меру доступности (см. Pryor & Kriss, 1977), в то время как другие исследования рассматривали объем вспомненной информации как коэффициент доступности информации (см. Ross & Sicoly, 1979, 12). Так как эти две меры не всегда хорошо коррелируют (см. Pryor и Kriss, 1977), двусмысленность измерения препятствовала эмпирическому прогрессу.

Часть этой проблемы измерения, однако, является спорной из-за третьей проблемы, а именно, концептуальной двусмысленности, касающейся использования термина *доступность*. Имеется тривиальный смысл, в котором весь социальный вывод обуславливается доступностью; оценки всегда основаны на том, что приходит на ум. Это использование термина *доступность* должно отличаться от использования эвристики доступности с помощью ясных критериев для определения, действительно ли эвристика доступности была задействована и как она была задействована, и действительно ли доступность примеров или ассоциации является посредником последующих оценок. Социальные психологи имели тенденцию сосредотачиваться на том, *какая* информация является доступной и *почему*; увеличение масштабов социального объяснения требует перенаправления усилий на то, *как* эвристика доступности связывает содержание того, что является доступным когнитивной обработке.

Влияние концепции доступности на социальную психологию, тем не менее, было большим по нескольким причинам. Во-первых, она выдвигает на первый план ошибки в обработке, которые могут быть поняты, не обращаясь к мотивационным конструкциям. Это не говорит о том, что воздействие мотивов на оценки незначительно, но просто указывает на то, что главные ошибки в познании также существуют. Во-вторых, хотя рациональные теории обеспечили полезные нормативные модели, с которыми фактические суждения могут быть сравнены, отклонения от этих моделей настолько обычны и явны, что необходимы описательные модели. В примерах, приведенных здесь, также как во многих других, эвристика доступности обеспечила одно из возможных описаний. При некоторых обстоятельствах, использование эвристики доступности приводит к уместным заключениям; однако, при тех обстоятельствах, где имеется предубеждение относительно того, какая информация является доступной, следуют ложные выводы. Определенно, предубеждения заметности, предубеждения воспроизведения и предубеждения, возникающие из-за познавательных структур, таких как схемы, убеждения и ценности, могут приводить к повышенной доступности неправильной или вводящей в заблуждение информации в задачах социальных оценок.

Учитывая, что предубеждения доступности могут быть весьма распространены, почему же используется эвристика доступности? Одна из очевидных причин – то, что она производит больше правильных, чем неправильных ответов, и в смысле затрат/выгоды, она имеет преимущество, экономя время и энергию. Второй ответ – то, что многие ошибки в выводах, произведе-

денных эвристикой доступности, не будут иметь значение. Например, если искаженные впечатления не повлияют на будущее функционирование, как при формировании неправильного впечатления о человеке, который встречается только один раз, предубеждение будет иметь мало значения. Предубеждение доступности может иметь мало значения, если оно постоянно в течение некоторого времени. Например, если Вы расцениваете руководителя как грубияна, не имеет значения то, что он грубый только когда, находится в этой роли, если это единственное обстоятельство, при котором человек взаимодействует с ним. Третий ответ – то, что многие ошибки могут быть исправлены. Если предубеждение доступности не исчезает само со временем, процесс начнет самоисправляться с каждым новым случаем. Например, если несколько друзей одного человека недавно развелись, оценка пропорции развода может быть временно преувеличена, но если друзья не продолжают разводиться, оцененная пропорция развода должна, в конечном итоге, совпасть с объективными данными. Наконец, в некоторых случаях, ошибка будет обнаружена по средством общения. Например, если утверждение, что человек сделал некоторую часть работы по дому, сталкивается с серьезными возражениями со стороны супруга (супруги), то позиция, вероятно, будет изменена. Короче говоря, обычное социальное общение обеспечивает основание для испытания выводов, и очевидно ложные заключения с далеко идущими значениями, вероятно, будут исправлены.

Но ошибочное восприятие с серьезными последствиями, при некоторых обстоятельствах могут сохраняться. Например, как исследования стереотипов иллюстрируют, если имеется предубеждение в формировании суждения, это может, в свою очередь, привести к предубеждениям в сохранении этой оценки с помощью сформировавшейся познавательной структуры (в этом данном случае, стереотипа). Эти “сопутствующие” или дополняющие друг друга предубеждения могут быть высоко устойчивыми к опровержению. Существовая, они могут иметь разрушительное влияние как на индивидуальные восприятия, так и на социальную политику. Короче говоря, нельзя считать предубеждения доступности несущественными, и соответственно, необходимы стратегии для обнаружения и исправления искаженных выводов.

14. Эвристика моделирования*

Даниель Канеман и Амос Тверски

Наше первоначальное трактование эвристики доступности (Tversky и Kahneman, 1973, 11) описывает два класса умственных действий, относящихся к процессу “активизации в памяти” (bring things to mind): припоминания случаев и конструирования примеров или сценариев. *Вспоминание* и *конструирование* – весьма различные способы “активизации в памяти”; они используются в качестве ответа на различные вопросы и подчиняются различным правилам. Предыдущее исследование было сосредоточено главным образом на припоминании, а процессом умственного конструирования пренебрегли.

Чтобы способствовать процессу изучения доступности для конструирования, мы в общих чертах опишем умственное действие, которое обозначим как *эвристику моделирования*. Мы начнем с обычного самоанализа: кажется, существует много ситуаций, в которых на вопросы о событиях можно ответить с помощью мыслительной операции, которая имеет сходство с имитационной моделью. Моделирование может ограничиваться и управляться несколькими способами: стартовые состояния для “функционирования” могут оставаться при их реальных изначальных значениях или изменяться так, чтобы принять некоторое непредвиденное значение; исходы могут не указываться, или же может быть установлено целевое состояние, с условием нахождения способа приближения к этому состоянию от первоначальных условий. Моделирование не обязательно подразумевает наличие единственной истории, которая стартует с начала и заканчивается определенным исходом. Скорее, мы рассматриваем результат моделирования как оценку легкости, с которой модель может производить различные

* Эта глава взята из лекции Катца и Ньюкомба по социальной психологии “О психологии возможных миров”, Энн Арбор, Мичиган, апрель, 1979. Работа была поддержана Службой Военно-морских исследований согласно Контракту N00014-79-C-0077 со Стэндфордским университетом.

исходы, учитывая ее начальные состояния и операционные параметры. Таким образом, мы предполагаем, что ментальное моделирование определяет меру, в которой чья-либо модель ситуации может генерировать различные исходы, также как возможности статистической модели могут быть оценены методами Монте-Карло. Легкость, с которой модель системы достигает определенного состояния, в конечном счете, используется для оценки способности (реальной) системы произвести это состояние.

Мы докажем, что оценки способности и вероятности, выведенные из ментального моделирования, используются в нескольких задачах оценивания, а также что они играют существенную роль в некоторых эмоциональных состояниях. Сначала мы перечислим некоторые оценочные действия, в которых, по-видимому, используется ментальное моделирование. Затем мы опишем исследование когнитивных правил, которые управляют умственным аннулированием прошлых событий. И мы кратко обсудим значения этих правил для эмоций, которые возникают, когда реальность сравнивается с альтернативой, которая пользуется одобрением, но которую человек не смог достигнуть, хотя мог легко вообразить, как он это сделает. Мы завершаем краткий набросок по эвристике моделирования замечаниями по сценариям и предубеждениям, которые, вероятно, возникнут, при использовании этой эвристики.

1. *Прогнозирование.* Вообразите первую встречу двух человек, которых вы хорошо знаете, но которые никогда раньше не встречались. Как Вы делаете такие прогнозы, как “Они хорошо поладят” или “Они будут раздражать друг друга”?

2. *Оценка вероятности определенного события.* Как Вы оцениваете вероятность вооруженного вмешательства американцев для обеспечения безопасности на месторождениях нефти в Саудовской Аравии в следующем десятилетии? Обратите внимание на различие между этой задачей и предыдущей. Моделирование в данном случае имеет указанное целевое состояние, и ее цель состоит в том, чтобы получить меру “легкости”, с которой это целевое состояние может быть произведено в пределах ограничений реалистичной модели международной системы.

3. *Оценка условных вероятностей.* Если гражданская война начнется в Саудовской Аравии, какие будут вероятные последствия? Обратите внимание, что это упражнение моделирования отличается от простого прогнозирования, потому что оно включает указанное начальное состояние, которое может отклоняться в большей или меньшей степени от существующей действительности. Оценка отдаленных последствий, в частности, вызывает интересную двусмысленность: какие изменения должны быть сделаны в текущей модели до “функционирования”? Нужно ли делать только минимальные изменения, которые включают указанное непредвиденное обстоятельство (например, гражданская война в Саудовской Аравии), подчинясь элементарным требованиям последовательности? Или нужно представлять все изменения, которые становятся вероятными за счет этого условия? В этом случае, напри-

мер, модель политической системы была бы сначала установлена таким образом, чтобы гражданская война в Саудовской Аравии стала само собой разумеющейся, и моделирование будет использовать параметры пересмотренной модели.

4. *Оценки, основанные на противоположных фактах.* Насколько ученые Гитлера приблизились к созданию атомной бомбы во время Второй мировой войны? Если бы они создали ее в феврале 1945, изменился бы результат войны? Оценки, основанные на противоположных фактах, также используются во многих повседневных рассуждениях, как, например, в предложении “она могла бы решить проблему с работой, если бы ее ребенок не был болен”.

5. *Оценки причинной связи.* Чтобы проверить породило ли событие А событие В, мы можем мысленно отменить событие А, и посмотреть, происходит ли В все еще в модели. Моделирование может также использоваться для проверки, увеличило ли событие А вероятность В, возможно даже сделало В неизбежным. Мы предполагаем, что испытание причинной связи моделированием включено в такие примеры, как “Вы очень хорошо знаете что, они поссорились бы, даже если она не упомянула бы его мать”.

Исследования отмены действий

Наши первоначальные исследования эвристики моделирования сосредотачивались на оценках, основанных на противоположных фактах. В частности, нас интересовал процесс, на основании которого люди говорят, что событие “почти произошло”. Пространственная метафора убедительна и была принята во многих философских исследованиях: кажется разумным говорить об отклонении действительности от некоторых некогда возможных, но нереализованных событиях. Психологическое значение этой оценки отклонения между тем, что случилось и что могло случиться, проиллюстрировано в следующем примере:

Г-н Крейн и г-н Тис должны покинуть аэропорт по расписанию на различных рейсах в одно и то же время. Они выехали из города в одном и том же лимузине, попали в пробку, и прибыли в аэропорт через 30 минут после намеченного времени вылета их самолетов.

Г-ну Крейну сказали, что его самолет улетел вовремя.

Г-ну Тису сказали, что его рейс был задержан и вылетел 5 минут назад.

Кто больше расстроен?

Г-н Крейн Г-н Тисс

Неудивительно, что 96% выборки студентов, которые отвечали на этот вопрос, заявили, что г-н Тис будет больше расстроен. Что же делает стереотип столь очевидным? Обратите внимание, что объективная ситуация этих двух господ абсолютно идентична, поскольку оба опоздали на свои самолеты. Кроме того, так как оба ожидали опоздания, различие между ними не может быть приписано разочарованию. Вообще, различие между Тисом и Крейн-

ном несущественно. Единственная причина для большего расстройства г-на Тиса – то, что у него была большая “возможность” успеть на свой самолет. Мы предполагаем, что стандартный эмоциональный сценарий для этой ситуации – это привлечение двух путешественников в упражнение моделирования, в котором они проверяют, насколько они были близки к тому, чтобы успеть на самолет. Конструирование, основанное на противоположных фактах, действует таким же образом, как и ожидание. Хотя из истории ясно, что ожидания г-на Тиса и г-на Крейна не могли быть различными, г-н Тис теперь больше разочарован, потому что ему легче вообразить, как он мог бы прибыть на 5 минут раньше, чем г-ну Крейну вообразить, как этой задержки на 30 минут можно было бы избежать.

Такие примеры напоминают Алису в Стране чудес, с их странной смесью фантазии и действительности. Если г-н Крейн способен вообразить единорогов, а мы думаем, что это так, почему он считает относительно трудным представить, что ему удалось избежать 30-минутной задержки? Очевидно, имеются ограничения свободы фантазии, и психологический анализ ментального моделирования состоит, прежде всего, в исследовании этих ограничений.

Наше знание правил ментального моделирования все еще находится в зачаточной стадии, и мы можем представлять только ранние результаты и предварительные предположения в области, которая кажется исключительно богатой и обещающей. Мы получили предварительные наблюдения относительно правил, которые управляют специальным классом деятельности моделирования – отмена прошлого. Наши исследования отмены событий сосредоточились на ситуации, в которой эта деятельность является особенно обычной – реакция выживших родственников на несчастный случай со смертельным исходом. И опять же, как в случае г-на Тиса и г-на Крейна, мы хотели исследовать то, что мы называем эмоциональным сценарием ситуации. Для примера, рассмотрим следующую историю:

Г-ну Джонсу было 47 лет, он отец троих детей и успешный банковский администратор. Его жена оставалась дома из-за болезни в течение нескольких месяцев.

В день несчастного случая г-н Джонс уехал из своего офиса в обычное время. Он иногда уезжал рано, чтобы заниматься домашними делами по просьбе его жены, но в этом не было необходимости в тот день. Г-н Джонс не поехал домой своим обычным маршрутом. День был исключительно ясным, и г-н Джонс сказал своим друзьям в офисе, что он поедет вдоль берега и будет наслаждаться видом.

Несчастный случай произошел на главном перекрестке. Когда подъехал г-н Джонс, загорелся желтый свет. Свидетели отметили, что он резко затормозил, чтобы остановиться на перекрестке, хотя мог легко проехать. Его семья признала, что это было привычной манерой вождения г-на Джонса. Когда он начал пересекать перекресток после изменения света, малотоннажный грузовик влетел на перекресток на высокой скорости и протаранил автомобиль г-на Джонса слева. Последний умер мгновенно.

Позже было установлено, что грузовик вел мальчик-подросток, находящийся в состоянии наркотического опьянения.

Как обычно случается в таких ситуациях, семья г-на Джонса и их друзья часто думали и говорили: “Если бы только...”, в течение нескольких дней, после несчастного

случая. Как они продолжали эту мысль? Пожалуйста, предложите одно или более предложений предложения.

Эта версия (обозначенная как версия “маршрута”) была дана 62 студентам Колумбийского университета. Другая группа из 61 студента получила версию “времени”, в которой второй абзац был следующим:

В день несчастного случая, Г. Джонс уехал из офиса ранее, чем обычно, чтобы выполнить работу по дому по просьбе его жены. Он ехал домой по своему обычному маршруту. Г-н Джонс иногда ездил вдоль берега, чтобы наслаждаться видом в особо ясные дни, но тот день был совершенно обычным.

Анализ данных, полученных при первом заполнении образца “Если бы только”, дается в табл. 1. Были получены четыре категории ответа: (1) отмена маршрута; (2) отмена времени отъезда из офиса; (3) пересечение перекрестка на желтый свет; (4) удаление из сценария мальчика в состоянии наркотического опьянения.

Таблица 1. Анализ первого заполнения образца “если бы только”

<i>Категории ответа</i>	<i>Версия "времени"</i>	<i>Версия "маршрута"</i>
(1) маршрут	8	33
(2) время	16	2
(3) пересечение перекрестка	19	14
(4) мальчик	18	13
(5) другое	1	3

Особенно внушительная сторона результатов, показанных в табл. 1 – это событие, которое не произойдет: ни один испытуемый не упомянул, что, если бы г-н Джонс приехал на перекресток на две или три секунды ранее, он проехал бы благополучно. Типичный вывод: события мысленно не отменяются с помощью произвольных изменений в значениях непрерывных переменных. Очевидно, испытуемые не выполняют задачу отмены событий, устраняя необходимое условие критического события, которое имеет самую низкую предполагаемую вероятность – процедура, которая привела бы их к тому, чтобы сосредоточиться на экстраординарном совпадении того, что эти два автомобиля встретились на перекрестке. Что бы тогда люди не делали, это не связано с априорной вероятностью.

Изменения, которые люди включают в истории, могут классифицироваться как нисходящие, восходящие или горизонтальные. Нисходящие изменения – те, которые отменяют неожиданный аспект истории или иначе увеличивают ее внутреннюю последовательность. Восходящие изменения – те, которые вносят маловероятные случайности. Горизонтальные изменения – те, в которых произвольное значение переменной заменено дру-

гим произвольным значением, которое не является ни более, ни менее вероятным, чем первое. Экспериментальная манипуляция заставила изменение маршрута быть нисходящим в одной версии, восходящим в другой, с соответствующим разнообразием в характере изменения времени фатальной поездки г-на Джонса. Манипуляция была успешна: испытуемые прогнозировали, что несчастный случай не произойдет, восстанавливая нормальное значение переменной, а не проводя исключение. Вообще, восходящие изменения относительно редки в ответах испытуемых, а горизонтальных изменений не было.

Понятие восходящих и нисходящих изменений заимствовано из опыта лыжника, едущего по пересеченной местности, и предназначено для иллюстрации особого характера расстояния, которое может быть определено для возможных состояний системы. Существенная особенность этого отношения – то, что оно не симметрично. Для лыжника, едущего по пересеченной местности, краткий спуск от А до В часто сопровождается длительным и трудоемким подъемом от В до А. В этой метафоре, исключительные состояния или события – горы, нормальные состояния или события – равнины. Таким образом, мы предполагаем, что психологическое расстояние от исключения до нормы, которую оно нарушает, меньше, чем расстояние от нормы до того же исключения. Предпочтение нисходящих изменений – это возможно, главное правило, которому подчиняется ментальное моделирование; оно олицетворяет существенные ограничения, которые придают реализм фантазиям, основанным на противоположных фактах.

Известный аспект результатов, показанных в табл. 1 – относительно низкая пропорция ответов, в которых несчастный случай отменяется за счет устранения случая, который естественно рассматривается как его причина, например, безумное поведение принявшего наркотики мальчика-подростка на перекрестке. Этот вывод иллюстрирует другую особенность ментального моделирования, которое мы назовем *правилом фокуса*: истории обычно перерабатываются за счет изменения некоторого свойства объекта, находящегося в центре внимания. В данном случае, конечно, центром внимания был г-н Джонс, так как испытуемых попросили сочувствовать его семье. Чтобы проверить правило фокуса, новая версия истории несчастного случая была построена таким образом, что последний абзац был заменен следующей информацией:

Как был позже установлено, грузовик вел мальчик-подросток, по имени Том Сирлер. Отец Тома только что обнаружил его дома в состоянии наркотического опьянения. Это было обычное явление, поскольку Том давно и постоянно принимал наркотики. Произошла ссора, в результате которой Том схватил ключи, которые находились на столе в гостиной, и поехал на грузовике в бессознательном состоянии. Он был серьезно ранен в результате несчастного случая.

Испытуемых, которым предоставили эту версию истории, попросили закончить образец “Если бы только...” либо от имени родственников г-на Джон-

са, либо от имени родственников Тома. Тут тоже, мы рассматриваем первый ответ, данный испытуемыми. Большинство испытуемых, которые приняли роль родственников Тома (68%), изменили историю, исключив Тома из сцены несчастного случая – наиболее часто, не позволяя фатальным ключам лежать на столе. Напротив, меньшинство (28%) испытуемых, принявших роль родственников г-на Джонса, упомянули Тома в своих ответах.

Мы описали это исследование отмены событий достаточно подробно, несмотря на его предварительный характер, чтобы проиллюстрировать удивительную точность правил, которые управляют ментальным моделированием, и продемонстрировать существование распространенных норм фантазий, основанных на противоположных факторах, которые являются уместными в некоторых ситуациях. Мы полагаем, что когнитивные правила, которые управляют легкостью умственной отмены событий, будут полезны в изучении ряда эмоций, которые могут называться эмоциями, основанными на противоположных фактах, из-за их зависимости от сравнения действительности с тем, что могло бы или должно было быть: расстройство, сожаление и некоторые случаи негодования, печали и зависти. Общей чертой этих эмоциональных состояний является то, что у человека гедонический уровень адаптации выше, чем его реальность, как будто неосуществленные возможности придали значимость уровню адаптации, с помощью веса, соответствующего легкости, с которой эти возможности достигаются при ментальном моделировании.

Заметки по сценариям

В контексте прогнозирования и планирования в условиях неопределенности, намеренная манипуляция умственными моделями является достаточно важной, чтобы заслужить название эвристики характерного моделирования. Самый яркий пример таких действий – явное воспроизведение сценариев как процедуры оценки вероятностей.

Что делает сценарий хорошим? В терминах, уже представленных, хороший сценарий – тот, который соединяет промежуток между начальным состоянием и целевым событием с помощью ряда промежуточных событий, с общей нисходящей тенденцией и без существенных подъемов. Неофициальные наблюдения свидетельствуют, что правдоподобие сценария зависит намного больше от правдоподобия его самой слабой связи, чем от числа связей. Сценарий особенно хорош, когда путь, который ведет от начального к последующему состоянию, не является очевидным, так, чтобы введение промежуточных стадий фактически усилило субъективную вероятность целевого события.

Любой сценарий – обязательно схематичный и неполный. Поэтому интересно обнаружить правила, управляющие выбором событий, которые явно определены в сценарии. Мы выдвигаем гипотезу, что “узлы” сценария – это события, у которых низкая избыточность и высокое причинное значение. Не

избыточный случай представляет локальный минимум в предсказуемости последовательности, точку, в которой существенные альтернативы могли бы возникнуть. Причинно значимый случай – это тот, возникновение которого изменяет величины, которые считаются нормальными для других событий в цепочке, которая, в конечном счете, ведет к цели сценария.

Разработка единственного вероятного сценария, который ведет от реалистических начальных состояний к указанному конечному состоянию, часто используется для поддержания оценки, что вероятность конечного состояния является высокой. С другой стороны, мы склонны делать вывод, что исход является маловероятным, если он может быть достигнут только с помощью восходящих предположений о редких событиях и странных совпадениях. Таким образом, оценка “совершенства” сценариев может быть эвристикой для оценки вероятности событий. В контексте планирования, в частности, сценарии часто используются для оценки вероятности, что план будет удачным и риска возникновения различных причин неудачи.

Мы предположили, что воспроизведение сценариев используется в качестве эвристики для определения вероятности событий посредством оценки способности некоторой причинной системы породить эти события. Подобно любой другой эвристике, эвристика моделирования должна быть подвержена характерным ошибкам и предубеждениям. Проведено мало исследований в этой области, но появляются следующие гипотезы, убеждая, что: (i) поиск не избыточных и причинно значимых “узлов” в построении сценария приведет к предубеждениям для него (и конечных состояний), в котором драматические события обозначают причинные переходы; будет наблюдаться тенденция недооценивать вероятность событий, произведенных медленными и возрастающими изменениями; (ii) использование сценариев для оценки вероятности связано с предубеждением в пользу событий, для которых вероятный сценарий может быть найден, с соответствующим предубеждением против событий, которые могут быть порождены множеством маловероятных способов. Такое предубеждение может иметь особенно пагубные последствия в контексте планирования, потому что оно производит чрезмерно оптимистические оценки вероятности того, что план будет осуществлен. По своему характеру, план состоит из цепи вероятных связей. В любой точке цепи, имеет смысл ожидать, что события развернутся, как было запланировано. Однако, совокупная вероятность по крайней мере одной фатальной неудачи может быть чрезвычайно высока даже, когда вероятность каждой индивидуальной причины неудачи ничтожно мала. Планы рушатся из-за сюрпризов, случаев, в которых происходит неожиданное восходящее изменение. Эвристика моделирования, которая искажается в пользу нисходящих изменений, связана с риском больших и систематических ошибок.

Часть V
Ковариация и контроль

15. Субъективная оценка ковариации: суждения, основанные на данных против суждений, основанных на теориях*

Денис Л. Дженнингс, Тереза М. Эмабайл и Ли Росс

Протекание социального опыта часто ставит перед нами задачу узнавания эмпирических ковариаций. Иногда, эти ковариации – просто испытание наших способностей к наблюдению и не несут никакого непосредственного практического интереса для нас. С другой стороны, например, когда эти ковариации включают первые признаки проблем и более поздние проявления, или используемые поведенческие стратегии и полученные исходы, или относительно открытые характеристики людей или ситуаций, а также относительно скрытые – такие способности обнаружения помогают определять наш успех в приспособливании к требованиям повседневной социальной жизни. Вообще, обнаружение ковариации будет играть важную роль в нашей продолжительной борьбе как “интуитивных ученых” (см. Nisbett и Ross, 1980; Ross, 1977, 1978), чтобы оценить и усовершенствовать имеющиеся гипотезы относительно нас самих, других людей и нашего общества. Очевидный вопрос, поэтому, звучит так: насколько опытны мы как обыватели, при оценке эмпирических ковариаций, представленных в нашем опыте?

* Мы хотим поблагодарить профессоров Даниела Канемана, Амоса Тверского и Пола Словика за их полезные комментарии и предложения. Мы также хотели бы отметить большие усилия Риты Френч и Джулии Стейнмец, которые принимали участие в исследовании, о котором сообщается здесь. Исследование было поддержано частично стипендией Национальной научной организации первому автору, грантом Национального института психического здоровья МН-26736 и грантом Национальной научной организации BNS-78-01211 Ли Россу и Марку Лепперу.

Организация и содержание этой главы появились благодаря более раннему исследованию этой же темы Л. Россом и Р. Х. Нисбеттом в книге “Предположения людей: Стратегии и недостатки социальных оценок” (Human Inference: Strategies and Shortcomings of Social Judgment), Prentice-Hall, Inc., 1980.

* Ковариация – (от англ. covariation) совместная изменчивость, взаимосвязь переменных величин (*прим. редактора*).

Перед тем, как начать обсуждение прошлого или настоящего исследования, мы должны обратить внимание, что каждодневное наблюдение обеспечивает нас большим количеством данных; и это наводит на мысль, что ответ на вопрос об опыте может быть далеко не простым. С одной стороны, и адаптивный характер социального поведения, и гармоничное качество социального взаимодействия не оставляет сомнений в том, что представители нашей культуры обладают глубокой проницательностью относительно причин и последствий поведения. Фактически, современные теоретики атрибуции (например, Jones и др., 1971; Kelley, 1967, 1971, 1973; Weiner, 1974), которые интересовались основаниями для такой проницательности, рассматривали способность обнаружить ковариацию и перешли к логическому использованию таких воспринятых ковариаций. С другой стороны, каждодневный опыт также предлагает красноречивое доказательство существования необоснованных предубеждений, разрушительных поведенческих стратегий и других проявлений ошибочных причинных или корреляционных теорий. Очевидно, что любое адекватное описание неофициальной оценки ковариации должно затрагивать неоднородность поведения обывателя. Это должно, так или иначе, урегулировать тонкую проницательность относительно некоторых функциональных отношений и невежество или иллюзии о других.

Предшествующее исследование оценки ковариации

Примечательно, что даже самые верные защитники способностей обывателей как интуитивных ученых (например, Peterson & Beach, 1967) могли мало хорошего сказать относительно того, как обыватели справляются с двумерными наблюдениями. Мы рассмотрим две линии исследования, которые поддерживают это нелестное описание, сосредотачиваясь в каждом случае на процедурных аспектах и результатах, которые окажутся подходящими для последующего описания нашего собственного исследования обывательской оценки ковариации.

Толкование таблиц сопряженности: проблема “экологической валидности”

Одно из важных направлений предшествующего исследования посвящено способности людей понимать функциональные связи, представленные в простых 2x2 таблицах сопряженности. Как правило, в этих таблицах суммировалось количество случаев присутствия и отсутствия переменной X (например, какая-либо болезнь), по сути связанной с наличием и отсутствием переменной Y (например, какой-либо симптом). Так как двумерные данные были аккуратно собраны и “рассортированы”, задача испытуемых была очень проста и незамысловата. Однако, экспериментальные данные (например, Jenkins & Ward, 1965; Smedslund, 1963; Ward & Jenkins, 1965) показывают, что необученные люди справлялись с такими задачами весьма плохо.

Оценочные стратегии, игнорирующие одну или более из четырех ячеек, явились причиной большинства трудностей испытуемых. Одна из самых распространенных неудач, например, включает фактически исключительную уверенность о размере ячейки “наличие-наличие” относительно всей совокупности. Таким образом, многие испытуемые, могут сказать, что симптом X связан с болезнью A , просто потому, что в таблице сопряженности показано большое количество случаев, где присутствует и болезнь, и признак. Испытуемые, обращающие внимание только на две из этих четырех ячеек, могут ввестись в заблуждение. Заметив, например, что у большинства больных людей скорее отсутствует симптом A , испытуемые могут заключить, что соответствующие связи – отрицательные; или заметив, большинство людей, у которых наблюдается симптом, больны, они могут сделать вывод, что связи – положительные.

Без формального статистического обучения, однако, очень немного людей оценивают факт, что правильные выводы в таких случаях могут быть сделаны только при рассмотрении *всех* четырех ячеек. Один из удовлетворительных методов, например, мог бы включать сравнение *пропорций* (то есть, сравнение пропорции больных людей, проявляющих специфический признак, с пропорцией здоровых людей, проявляющих этот признак). Формальная стратегия, продиктованная современными работами по статистике, конечно, включает рассмотрение всех четырех ячеек. Определенно, можно было бы потребовать от человека обратить внимание на несоответствия между наблюдаемыми частотами ячейки и теми частотами, которые следует “ожидать” на основе соответствующих минимальных частот или вероятностей (таким образом, обеспечивая информацией, используемой для вычисления статистики хи-квадрат).

Скептики, склонные отстаивать основания обывателей, могли бы почувствовать соблазн отклонить такие проявления как свидетельство только того, что люди плохо “оперируют” таблицами сопряженности и отстаивать то, что ошибки испытуемых в таких задачах являются, прежде всего, следствием новизны задачи и искусственности. Отношения между лабораторными действиями и каждодневной реальностью, то есть проблема “экологической валидности”, являются важной проблемой, которую нельзя избежать в исследовании, о котором говорится в этой главе. Давайте рассмотрим эту проблему в надежде, что это предупредит и вооружит читателя необходимыми сведениями.

Сначала, учитывая, что задача оперирования таблицами бесспорно необычна и экологически нетипична, заметим, что логика (и различные логические недостатки), показанные испытуемыми в этих лабораторных задачах, соответствует логике, наблюдаемой в разнообразии каждодневных выводов. Рассмотрим, например, вопрос, “Отвечает ли Бог на молитвы?” “Да,” может ответить необученный человек, который смотрит только на ячейку “наличие – наличие”, “потому что много раз я просил Бога о чем-то, и Он давал это мне”. Более искушенный и скептически настроенный чело-

век может спросить, “Но как часто Вы просили, Бога о чем-либо и *не* получали этого?” Сравнение двух ячеек, однако, все еще крайне неадекватно для данной задачи на логику. Даже дополнение третьей ячейки, благоприятные исходы, которые произошли в отсутствии молящегося, оставит проблему нерешенной, поскольку необходимы все четыре ячейки прежде, чем основательный вывод может быть сделан. И даже наиболее искушенный из интуитивных психологов, вероятно, оставил бы без внимания предложения, что данные из ячейки “отсутствие - отсутствие” (то есть, благоприятные исходы, о которых не просили и которые не произошли) обязательны для оценки воздействия молитв на мирские исходы.

На обвинение в нерепрезентативности можно ответить даже более убедительно. Если задача с таблицей из четырех ячеек создает некоторые уникальные проблемы, эта же задача также удовлетворяет большинство требований испытуемых, которые усложняют каждодневные оценки ковариации. Эти требования изначально включают припоминание, кодирование, хранение и восстановление соответствующих данных, а затем подготовку их в той форме, которая позволяет оценить ковариацию. Рассмотрим, например, человек, который решается проверить стереотипное понятие, что рыжеволосые люди являются вспыльчивыми (или, более точно, что наличие рыжих волос положительно коррелирует с наличием вспыльчивого характера). Сначала, он должен решить, какие данные должны быть рассмотрены. Попытается ли он сделать выборку или вспомнить случаи вспыльчивых рыжеволосых людей? Или он должен рассмотреть некоторых рыжеволосых людей и обратить внимание, насколько они вспыльчивы, или рассмотреть некоторых вспыльчивых людей и обращать внимание, сколько из них рыжеволосые? Очень вероятно, он примет некоторую стратегию или комбинацию стратегий, не рассматривая возможность, что относительная частота уравновешенных брюнеток могла быть уместной в данной логической задаче. Это положение дел не соответствует задаче с таблицей из четырех ячеек, которая делает все соответствующие данные одинаково доступными для испытуемого, даже если не требуется, чтобы их все надлежащим образом рассмотрели.

Предположим, что рассматриваемый человек, так или иначе признает, что тестирование его стереотипа, требует, чтобы он рассмотрел данные во всех четырех ячейках. Как он должен сформировать соответствующую выборку данных, из которой он взял бы уместные частоты ячейки? Если он просто рассмотрит первое множество людей, которые приходят на ум? Может быть, ему следует рассмотреть всех людей, принадлежащих некоторому ограниченному классу (например, члены его семьи, друзья или арендаторы в его доме)? Или ему следует попытаться использовать некоторый “случайный” метод для построения частных случаев из полной популяции? Потенциальные источники предубеждения в наиболее неофициальном методе формирования выборки очевидны. Выборка из людей, которые первыми приходят на ум, может привести к серьезным предубеждениям “доступности” (Tversky и Kahneman, 1973,11). А выборки друзей или членов семьи или соседей, веро-

ятно, проявят предубеждение, а также нарушение требования независимости (не тривиальная проблема, так как семья, друзья или даже соседи могут разделять поведенческие тенденции, физические характеристики и даже соответствия между ними). “Новая” процедура с таблицей из четырех ячеек помогает человеку, оценивающему ковариацию, решить задачу сопоставления этих тупиковых и критических проблем.

Задачи кодирования данных, их хранения и исправления добавляет множество дополнительных факторов и проблем в ежедневную оценку ковариации. Вероятно, гипотеза, что рыжеволосые люди являются вспыльчивыми, может искажать решения относительно того, кто является или не является вспыльчивым, и чьи волосы являются или не являются рыжими. Эта проблема рассматривается ниже в этой главе. А сейчас давайте просто обратим внимание, что предшествующие ожидания или гипотезы включают потенциально искажающее влияние на каждую стадию обработки информации, влияние, которое устраняется, когда человеку предоставляются данные, для которых процессы кодирования, хранения и восстановления уже закончены. Давайте повторим, что такие жизненные препятствия к точной оценке ковариации способны оказаться гораздо более устрашающими, чем любые препятствия являющиеся результатом искусственности или неопытности с методологией таблицы сопряженности.

Ложная корреляция: воздействие предрассудков на восприятие

Насколько успешно люди обращаются с двумерными распределениями, когда соответствующие наблюдения представлены скорее индивидуально, чем перераспределены в форме таблиц сопряженности? Многие из существующих данных взяты из конструктивной работы Чепманов по ложной корреляции. (Для детального рассмотрения этого исследования см. Chapman & Chapman, 1967, 1969; также см. Главу 17, этого издания). Наш краткий обзор сосредоточивается на специфических проблемах и выводах, которые предоставляют пространство для наших собственных, более современных, исследований.

Возможно, самое простое резюме результатов Чепманов и выводы относительно способности испытуемых обнаружить ковариации между клиническими признаками и симптомами – это то, что ковариации, о которых было сообщено, отражают истинные ковариации гораздо меньше, чем предубеждения относительно характера отношений, которые “должны” существовать, основанные на теории или на семантике. В то время как эти выводы были впечатляющими, спорными и обладали непосредственной уместностью для практика, то общее, что они сделали относительно эффектов предубеждений по поводу наблюдаемых ассоциаций, было по существу знакомым.

Как отмечал Чепмен в своей первоначальной работе (L.J. Chapman, 1967), ложные корреляции не ограничены областью клинических суждений. Большинство предрассудков – это по существу опытным путем необоснованные убеждения относительно ассоциации между специфическими действиями

или событиями и последующими положительными или отрицательными исходами. Расовые, этнические, региональные, религиозные или профессиональные стереотипы – это также убеждения относительно ковариаций, убеждения, которые настоятельно поддерживаются и являются стойкими к воздействию опровергающих данных (см. Adorno и др., 1950; Allport, 1954; P. Goldberg, 1968; J.M. Jones, 1972; Taunor & Deaux, 1973). Исследование относительно “неявной теории личности” может также быть приведено в этом контексте. Более пятидесяти лет назад Торндайк (Thorndike) описал так называемый эффект ореола, при котором воспринимающий ожидает и сообщает об ассоциации всех положительных характеристик личности. Последующие исследователи (например, Koltuv, 1962; Norman & Goldberg, 1966; Passini & Norman, 1966) значительно расширили эту работу и исследовали концептуальные схемы и убеждения относительно личности, которые располагают испытуемых сообщать о последовательных связях между различными чертами характера или различными индикаторами одной и той же черты, чьи эмпирические связи очень слабы или даже не существуют.

Таким образом, существует большое количество эмпирических свидетельств, что, в области оценки ковариации противоборство между ожиданиями и сведениями может быть неравным. Также как в других типах концептуальных и когнитивных оценок (ср. Bruner, 1957a, 1957b; Bruner, Postman, & Rodrigues, 1951), предубеждения интуитивного ученого относительно эмпирических связей определяют, что он обнаруживает, что ему не удастся найти, и что он видит такого, чего не существует. Такие поспешные выводы о превосходстве теории над данными могут привести нас к тому, что мы не примем во внимание значимую неопределенность – или, наконец, важный нерешенный вопрос – в исследовании Чепманов. К примеру, мы можем установить две отдельные проблемы об интуитивной оценке ковариации, которые, в некотором смысле, смешаны в этих и последующих исследованиях (например, Golding & Rorer, 1972, Strarr & Katkin, 1969). Первая проблема касается трудностей, возникающих у людей при обнаружении ковариации в непосредственно доступных данных – то есть их способность узнать и оценить ковариации, когда они “не усложняются” теориями и объяснениями относительно рассматриваемых эмпирических связей. Вторая проблема рассматривает тенденцию интуиции и теорий испытуемых приводить их к тому, что они признают существование устойчивых корреляций, где такие предположения опровергаются лучшими “объективными” сведениями.

Разделение этих проблем предполагает необходимость исследовать проведение испытуемыми оценок ковариации в двух разных заданиях. Первое задание по своей сути “психологическое”, и оно дает сведения об отношении между субъективными оценками и объективными измерениями. В этом задании испытуемым предлагается множества двумерных наблюдений, по поводу которых у них не может быть никакой предвзятой теории, а затем их просят оценить силу связи в каждом множестве. Исследователь затем

должен попытаться измерить сложность каждого задания (в основном, отмечая степень изменчивости, связанную с отдельными оценками) и определить характер “психометрической” функции, соотнеся субъективные оценки и объективные измерения.

Второе задание касается оценок, совершенно противоположных оценкам, которые основаны на данных; оно имеет отношение к оценкам, которые в целом, или почти целиком, основаны на теории. Таким образом, пары переменных или измерения, которые могут быть связаны с интуицией испытуемых или их теориями, должны быть определены, но исследователь не может предоставлять никаких релевантных двумерных данных. Если испытуемым предлагаются какие-либо данные в таких оценках, то они должны быть взяты из ежедневного опыта или воспоминаний испытуемых; и, будучи таковыми, они подвержены множеству “процессуальных” ошибок и предубеждений, о которых мы упоминали ранее в нашем обсуждении. С целью оценки и сравнения экспериментатор должен обладать средствами оценки точности или природы ошибок, связанных с такими интуитивными оценками. То есть, у него должна быть некоторая “объективная” мера корреляции для каждой пары переменных, предоставленной испытуемым.

Экспериментальное сравнение оценок ковариации, основанных на данных и на теориях

Недавно мы провели исследование, преследуя вышеупомянутые цели и используя вышеупомянутую базовую стратегию. Для задач, основанных на данных, были использованы три различных типа двумерного распределения, которые будут подробно описаны в следующей главе. Для оценок, основанных на теории, были обозначены пары связанных переменных, но экспериментатор не предоставил никаких фактических данных. Чтобы облегчить сравнение между оценками, основанными на данных и на теориях, использовался внутригрупповой план, в котором испытуемые использовали одинаковые шкалы субъективных оценок для обоих заданий (сбалансированные по порядку представления).

В каждом случае в задачах релевантной оценки корреляции использовались скорее непрерывные, чем дихотомические переменные (в отличие от работы Charman & Charman, 1967, 1969 и Jenkins & Ward, 1965). Эта особенность была разработана для того, чтобы лучше понять природу большинства ежедневных задач оценивания.¹

¹Наше изучение литературы выявило, что оценка ковариации с непрерывными переменными очень редко использовалась в предыдущих исследованиях. Исключением, обнаруженным в этом поиске, было одно исследование, к которому редко обращались (исследование Erlick, 1966; Erlick & Mills, 1967), которое касалось, в основном, влияния вариаций на воспринятую соотнесенность; лишь немногие данные из этого исследования применимы для настоящего исследования.

Задачи на оценивание

В этом эксперименте, организованном для сравнения оценок ковариации, основанных на данных и на теориях, участвовали в целом 64 студента старших курсов Стэндфордского университета, причем ни один из них не окончил курсов по статистике. Для обоих типов оценок испытуемые использовали простую шкалу для описания своего субъективного впечатления о направлении и силе связи между парами переменных. Используя эту шкалу, испытуемые сначала оценивали, была ли рассматриваемая связь положительной или отрицательной. Затем они оценивали силу связи, помещая “X” на 100-бальной шкале оценок, концы которой были помечены значениями “полная связь” и “отсутствие связи”.

Оценки ковариации, основанные на данных. Одна часть эксперимента касалась оценок ковариации, основанных на множествах двумерных наблюдений, предоставленных экспериментатором. Чтобы создать релевантные двумерные распределения было использовано три варианта стимульного материала. Для каждого типа стимульного материала было создано множество двумерных распределений, чтобы показать различные объективные ковариации между двумя переменными. Испытуемых попросили изучить каждое распределение и оценить связи, которые оно описывает, используя 100-бальную шкалу субъективных оценок.

Мы ожидали, что точность оценок испытуемых в этом задании будет частично зависеть от трудностей, с которыми они столкнулись при кодировании, оперировании и запоминании данных. Чтобы проверить это предположение, три типа стимульного материалов сильно различались в требованиях к обработке информации, которые они предъявляли к испытуемым до оценки ковариации, особенно учитывая необходимость оценки величины стимула и необходимость хранения и выборки информации из памяти. Таким образом, первый тип представлял множества из 10 пар простых чисел. Второй тип состоял из рисунков, каждый из которых изображал группы из 10 человек разного роста, держащих трости разной длины. Заключительный тип распределения был представлен в форме аудиокассет, на которых можно было услышать, как каждый из 10 человек сначала говорил некоторую букву, занимающую некоторую порядковую позицию в алфавите (по сути, первую букву своей фамилии) и затем сразу же пел музыкальную ноту разной продолжительности.

Таблица 1. Пары переменных, описанные испытуемыми и их объективные корреляции

	Приведенные переменные	Объективная корреляция
A	Оценки студентами собственного либерализма Оценки студентами деятельности экономических лидеров в прошлом десятилетии	-0.28 ^a
B	Оценки студентами собственной интеллектуальности Оценки студентами деятельности американских президентов в прошлом десятилетии	-0.19 ^a
C	Оценки студентами собственной застенчивости Количество американских штатов, которые посетили студенты	-0.12 ^a
D	Оценки студентами собственного честолюбия Рост студентов	0.01 ^a
E	Оценки студентами работы ректоров университета в прошлом десятилетии Время, потраченное студентами в неделю на занятия спортом	0.08 ^b
F	Оценки студентами собственной интеллектуальности Семейный доход студентов	0.17 ^a
G	Нечестность детей, измеряемая в соответствии с ложным сообщением о спортивных достижениях Нечестность детей, измеряемая в соответствии со степенью обмана, используемого при решении загадки	0.18 ^a
H	Семейный доход студентов Оценки студентами действий экономических лидеров в прошлом десятилетии	0.28 ^a
I	Способность шестиклассников "препятствовать вознаграждению" Способность шестиклассников сопротивляться искушению обмануть	0.31 ^c
J	Оценки студентами собственной сознательности Оценки сознательности этих студентов, предоставленные их соседями по комнате	0.35 ^d
K	Оценки студентами собственной интеллектуальности Оценки студентами собственного честолюбия	0.37 ^a
L	Оценки студентами деятельности Конгресса в прошлом десятилетии Оценки студентами деятельности трудовых лидеров в прошлом десятилетии	0.40 ^a

Продолжение табл. 1

М	Время, потраченное студентами в неделю на занятия спортом Оценки студентами собственной физической формы	0.52 ^а
Н	Оценки студентами деятельности американских президентов в прошлом десятилетии Оценки студентами деятельности экономических лидеров в прошлом десятилетии	0.55 ^а
О	Оценки собственного политического консерватизма Объединение оценок пунктов шкалы этноцентричности	0.57 ^б
Р	Рост студентов Вес студентов	0.79 ^а

^аКорреляция, полученная из исследования студентов (N=295) Стэнфордского университета.

^бКорреляция, полученная из работ Гартшорна и Мэя (Hartshorne & May, 1928).

^вКорреляция, полученная из работ Мишеля и Гилигана (Mischel & Gilligan, 1964).

^гКорреляция, полученная из работ Бема и Аллена (Bem & Allen, 1974).

^дКорреляция, полученная из работ Адорно и др. (Adorno..., 1950).

Оценки ковариации, основанные на теории. Другая часть эксперимента (порядок представления этих двух частей был сбалансирован) рассматривала оценки ковариации, основанные скорее на априорных ожиданиях испытуемых или теориях, чем на непосредственно доступных двумерных данных.

Пары измерений или переменных были определены, и испытуемых попросили оценить сначала направление, а затем силу связей для каждой пары, всегда используя ту же простую субъективную 100-бальную шкалу оценки, как было использовано для оценок, основанных на данных. Некоторые из этих пар в этой части эксперимента касались расходящихся поведенческих измерений личных черт характера, например, два измерения честности, используемые в классическом исследовании Гаршорна и Мэя (Hartshorne & May 1928) кросс-ситуативной последовательности в моральном поведении; другие касались личных отношений, привычек или предпочтений. В каждом случае, однако, субъективные оценки ковариации, которые мы получили, сравнивались с "объективными" корреляциями, взятыми из предыдущих эмпирических исследований. (Полный список этих пар переменных и соответствующих корреляций представлен в табл.1)

Результаты

Оценки ковариации, основанные на данных

Наше первое исследование включает оценки ковариации, которые испытуемые сделали в ответ на непосредственно доступные двумерные данные. Могут возникнуть два специфических вопроса: во-первых, как легко человек может обнаружить ковариации различной величины при отсутствии любой “теории” о соответствующих распределениях данных? Во-вторых, какова была природа “психофизической функции”, связывающей среднее значение субъективных оценок и объективное измерение ковариации?

Проверка данных предполагала, что задача оценки была очень трудна. Стандартные отклонения и межквартильные широты, связанные с полными оценками группы для каждого набора данных были чрезвычайно высоки (см. Рисунок 1). Что является особенно интересным – это трудность, с которой столкнулись многие испытуемые, просто в признании существования положительных связей, даже связей большой величины. Фактически только, когда объективные корреляции достигают уровня от +0.6 до +0.7, межквартильный интервал последовательно исключает отрицательные оценки связей. Нестабильность оценок ковариации систематически не отличалась для трех типов двумерных распределений данных, несмотря на явные различия в количестве обработанной информации (и, следовательно, возможности для случайной или систематической ошибки), связанной с этими тремя типами. Эта очевидная нечувствительность к увеличению требований к процессу “обработки информации” может предложить решения относительно природы оценочной “стратегии” испытуемых – возможность, которую мы будем исследовать несколько позже в этой главе.

Также на Рисунке 1 представлены данные, соответствующие второму вопросу об оценках, основанных на данных, вопросу о “функциональной связи” между объективными и субъективными показателями ковариации. (Читатель обратит внимание, что средние величины этих трех областей стимула объединены в этом представлении результатов. Ни проверка, ни статистические тесты значимости не показали никаких существенных или последовательных тенденций, отличающие результаты для трех областей стимула, и результат такого объединения – более однородная и более понятная функция, чем функция, полученная для любой из трех индивидуальных областей.) Для удобства читателя мы подогнали кривую к пунктам в нашем графике. Мы должны отметить, что эта кривая, была нарисована позднее; она не была получена из какой-либо “психофизической теории”.

Из проверки ясно, что, несмотря на внутрigrupповую изменчивость для индивидуальных оценок, функция, связывающая средние субъективные оценки с объективными корреляциями (корреляция Пирсона) была довольно однородная. Существует резко возрастающая функция, связывающая эти две переменные. Таким образом, связи в изменениях, с которыми обычно

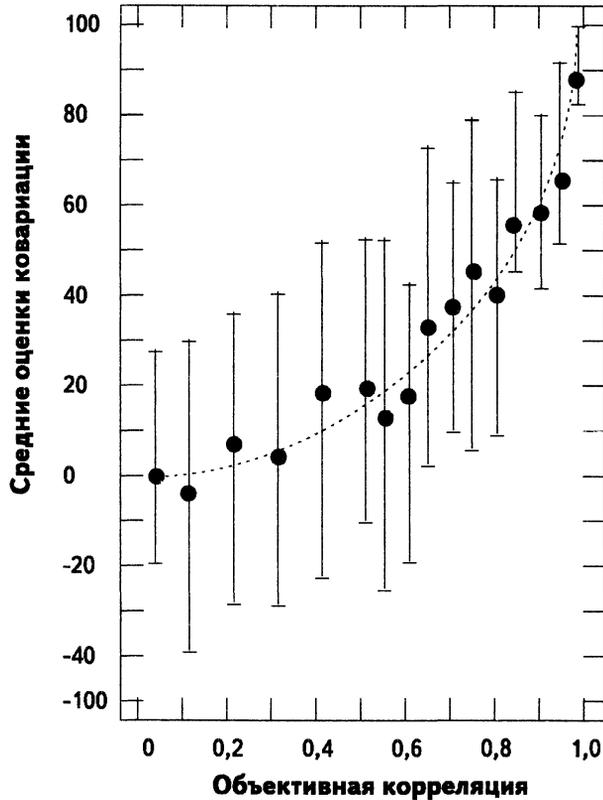


Рис 1. Средние величины и межквартильные изменения для оценок ковариации, основанных на данных, причем результаты для трех типов распределения объединены. “Психофизическая функция” $100(1-\sqrt{1-r^2})$ показана пунктирной линией

сталкиваются психологи, занимающиеся измерениями черт характера личности или другой кросс-ситуативной устойчивости в социальном поведении (то есть, r – от 0.2 до 0.4), можно обнаружить только, выдавая средние оценки в диапазоне от 4 до 8 на 100-бальной шкале. Даже связи, рассматриваемые такими психологами как очень прочные (то есть, r = от 0.6 до 0.8), заканчиваются довольно скромными субъективными оценками ковариации. Объективные корреляции 0.7, например, произвели среднюю субъективную оценку 34 – оценка, находящаяся между пунктами помеченными “довольно слабая” и “умеренная” на субъективной 100-бальной шкале. Только, когда объективные корреляции приблизились к уровню 0.85, среднее значение группы достигло средней точки субъективной шкалы, и только за этой точкой испытуемые последовательно оценили связи как крайне положительные.

Более близкое рассмотрение оценок предполагает, что субъективные оценки связанности или ковариации – не линейная функция r или не r^2 .

Скорее, образец объединенной средней оценки кажется хорошо описанным выражением, $(1 - \sqrt{1 - r^2})$. (Читатель может узнать это выражение как “коэффициент отчуждения”, меру редукции стандартной ошибки, связанной с прогнозированием переменной Y , основанным на знании как переменной X , так и корреляции между X и Y (Huntsberger, 1967)). Действительно, подобие между фактическими средними оценками и оценками, описанными этим выражением, было поразительно – 95% вариации между средними объясняется функцией, основанной на коэффициенте отчуждения. Мы спешим напоминать читателям, однако, что они не должны неверно понять значение этого точного соответствия. Степень соответствия, показанная на Рисунке 1, применяется только к групповым оценкам; индивидуальные оценки свидетельствовали о трудностях испытуемых в различении разных уровней ковариации.

Очевидно, преждевременно делать какой-либо вывод относительно природы “психофизической” функции, связывающей субъективные ответы и объективные измерения ковариации. Детали задачи, контекст, и возможно, даже совокупность оценивающих, несомненно, произвели бы различия в функции, различия, которые могли быть зафиксированы только путем включения соответствующих параметров. Однако, имеющихся результатов достаточно, чтобы предположить, что умеренные корреляции подобного вида, о которых так часто сообщают психологи-эксперты (ср. Mischel, 1968, 1969), вероятно, будут слабо обнаруживаться или произведут только очень слабое впечатление на обывателей, которые сталкиваются с такими ковариациями в отсутствии предубеждений, основанных на теории.

Далее стоит перенести внимание на то, что три различные задачи оценок ковариации, основанных на данных, произвели очень похожие результаты, несмотря на различные требования к обработке информации, которые они предъявили испытуемым. Таким образом, “психофизическая” функция, проиллюстрированная на Рисунке 1, точно отражает связи между субъективными оценками и объективными измерениями для *каждой* из этих трех задач, выполненных испытуемыми. Корреляция между средними значениями фактических оценок испытуемых и оценками, спрогнозированными оценкой функции от $r = 0.91$ для пары нота / буква до $r = 0.98$ для пары чисел.

Оценки, основанные на теориях

Ранее мы представили свидетельство того, что необходимо, чтобы относительно сильная объективная корреляция (r в терминах Пирсона) породила субъективные оценки, даже незначительно отличные от нуля. Рассмотренная отдельно, психофизическая функция для оценок, основанных на данных, может просто быть свидетельством того, что субъективные измерения обывателя осторожны или консервативны по отношению к измерениям, используемым наиболее формальными статистиками. (Конечно, трудно поспорить, что r^2 — более соответствующий индекс соотнесенности, чем r ,

т. к. он может быть легче связан с прогнозирующей полезностью или уменьшением неопределенности относительно значения одной переменной, основанной на знании значения другой, связанной с ней, переменной.) Короче говоря, полученные оценки, основанные на данных, не предполагают какой-либо систематической ошибки со стороны интуитивного психолога, хотя изменчивость, связанная с такими оценками, не оставляет сомнения относительно трудностей и неопределенности в задачах оценки.

Учитывая эти результаты и возможные интерпретации, мы можем теперь перейти к рассмотрению результатов для субъективных оценок ковариации относительно связей, перечисленных в табл. 1. Эти оценки были сделаны при отсутствии непосредственно доступных данных; предположительно, они были осуществлены только на основе неофициальных теорий или интуиции оценивающих, без сохранения этих “обработанных” данных, отмеченных и извлеченных из ежедневного опыта. Рисунок 2 предоставляет первое рассмотрение этих результатов. Ясно, что никакая отдельная линейная или какая-либо другая функция не охватывает связи между субъективными оценками и объективными измерениями. Однако существует бесспорная тенденция, согласно которой оценки испытуемых, основанные на теории, показывают приблизительное соответствие объективным сведениям. То есть положительные эмпирические связи были оценены как положительные; отрицательные связи были оценены как отрицательные; относительно прочные эмпирические связи были оценены как более прочные, чем относительно непрочные, и т.д.

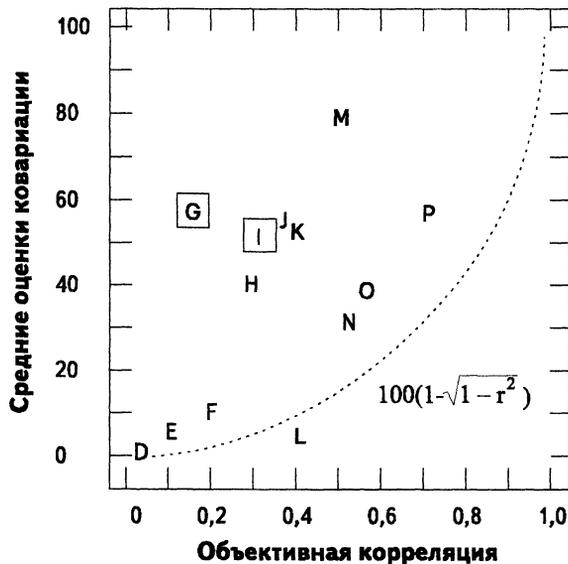


Рис. 2. Средние значения для оценок ковариации, основанных на теории. (См. обозначение переменных в табл. 1)

Возможно, наиболее поразительная особенность оценок, основанных на теории, – это то, что однажды освобожденные от ограничений, наложенных непосредственно доступными данными, испытуемые переставали быть осторожными или консервативными. Они охотно отмечали середину и даже верхние области субъективной 100-бальной шкалы даже, когда сталкивались с парами переменных, которые объективно были незначительно коррелированы. Иллюстрация, приведенная на Рисунке 2, убеждает в этом. Несколько пар переменных, коррелированных на уровнях, которые в случае оценок, основанных на данных, выдали средние субъективные оценки, весьма близкие к 0 (например, $r = 0.3$), теперь выдали оценки 30, 40 или более на субъективной 100-бальной шкале.

Пары переменных G и I заслуживают особенного внимания, так как они касаются основанных на теории оценок кросс-ситуативной поведенческой устойчивости и, таким образом, выражают проблему, представляющую теоретический интерес в настоящее время (см. Alker, 1972; Allport, 1966; Bem и Allen, 1974; Bem и Funder, 1978; Bem и Lord, 1979; Mischel, 1968, 1969). То, о чем говорят субъективные оценки для этих двух связей, очевидно. При наличии непосредственно доступных объективных данных, представляющих корреляции в диапазоне от $r = 0.2$ до $r = 0.3$, испытуемые дали средние оценки между 0 и 10 в оценочной 100-бальной шкале. Теперь, при отсутствии непосредственно доступных данных, но касающихся двумерной области, показывающей тот же диапазон от $r = 0.2$ до $r = 0.3$, подразумеваемые теории индивидуальности подтолкнули испытуемых к тому, чтобы предложить оценки ковариации, которые в среднем находились в диапазоне от 50 до 60. Какая степень ковариации в непосредственно доступных двумерных данных потребовалась бы, чтобы породить подобные оценки при отсутствии “теорий” или предубеждений? Ответ виден на Рисунке 2: только эмпирические корреляции в диапазоне $r = 0.90$, могли породить такие оценки.

Результаты нашего исследования, таким образом, нашли отражение в эмпирических обобщениях.

1. Когда непосредственно доступные двумерные данные исследованы при отсутствии каких-либо теорий или предубеждений относительно их связанности, интуитивному психологу труднее обнаружить ковариации величин, которые могут характеризовать широкий диапазон функциональных связей, представленных в ежедневном социальном опыте. В частности, ковариации в диапазоне величин, с которыми обычно имеют дело психометристы в поиске кросс-ситуативной поведенческой устойчивости, вероятно, останутся необнаруженными или будут восприняты как близкие к нулю.

2. Когда объективные доступные в настоящий момент двумерные данные не могут быть исследованы, а предшествующие теории, или предубеждения существуют, интуитивный психолог может ожидать и прогнозировать ковариации значительной величины – часто большие, чем те, которые представлены прошлым опытом или будут представлены в будущем.

Рассматриваемые вместе, эти обобщения помогают сосредоточить внимание на контрасте между оценками, основанными на данных и оценках, ос-

нованных на теориях. Теории, сторонниками которых мы являемся, заставляют нас ожидать и прогнозировать более прочные эмпирические связи, чем они есть на самом деле; и многие существующие эмпирические связи, даже большой величины, могут быть обнаружены, только если мы уже ожидаем это сделать.

Происхождение и жизнеспособность теорий

Стратегии оценки ковариации

Кроме предложения обобщений относительно исходов различных заданий оценки, важно начать рассматривать процессы оценки и стратегии, которые могли бы лежать в основе таких исходов. Снова различие между оценками, основанными на данных и теориях, оказываются критическими, и снова конструктивные выводы Чепманов обеспечивают удобный отправной пункт.

Данные вместо теории. Из исследований Чепманов стало ясно, что предубеждение одержало победу над информацией, предоставленной непосредственно доступными данными. Менее очевидно, однако, как это происходит. “Видели” ли испытуемые на самом деле те связи, о которых они сообщали? Или они просто сообщали об отношениях, которые, как они ожидали, будут присутствовать в данных, без какого-либо соответствующего субъективного впечатления? Или испытуемые “шли на компромисс” между своим субъективным восприятием и ожиданиями, придавая некоторое значение каждому из них? Невозможно дать однозначный ответ на основе сообщенных результатов, хотя имеется определенное свидетельство, как того, что испытуемые, придавали, по крайней мере, некоторое значение своему непосредственному восприятию данных и что их восприятие полностью не было определено их ожиданиями. В частности, мы обращаем внимание на сообщение Чепманов, что повторное воздействие не-отношений (non-relationships) в конечном счете, уменьшало количество сообщений об иллюзорной корреляции. Однако было также ясно, что данные никогда не могли полностью одержать победу над предубеждениями испытуемых, так как даже при увеличении числа раз, когда испытуемые подвергались воздействию *отрицательных* связей, не могло полностью устранить сообщения об иллюзорных положительных корреляциях.

Вопросы об интуитивных стратегиях для оценки ковариации, таким образом, оказываются в центре внимания. Как испытуемые решают, какая степень связи присутствует в данных? Как они решают, какая степень связи “должна”, присутствовать в репрезентативном наборе двумерных наблюдений, которые предоставлены в соответствии с их интуитивными теориями или предубеждениями? Когда и как предубеждения преобразовывают субъективные опыты ковариации? Начиная размышлять над ответами на

такие вопросы еще раз полезно вначале сосредоточиться на оценках, основанных только на данных и оценках, основанных только (или, по крайней мере, в значительной степени) на теории, и только затем вернуться к случаю, когда сталкиваются непосредственно доступные данные и ожидания, основанные на теории.

Стратегии оценок, основанных на данных. Возможно, наше наиболее поразительное открытие относительно оценок, основанных на данных, включало ответы испытуемых на двумерные распределения, предлагающие ковариации величин, которые встречаются в каждодневном опыте и соответствуют неофициальным социальным теориям, в частности, для испытуемых трудно обнаружить такие связи. И, в среднем, они оценили ковариации, как весьма близкие к нулю. Какие последствия имеют такие результаты на неофициальные стратегии оценки ковариации, которые могли бы использоваться нашими испытуемыми?

Размышляя над этой проблемой, важно помнить, что отдельные испытуемые не считали, что объективно слабые корреляции были близки к нулю. Вместо этого, они предложили довольно широкий диапазон оценок, для которых только соответствующие средние значения были близки к нулевой отметке. Снова, трудно точно различить, что испытуемые фактически “видели” и что они оценили, несмотря на то, что они видели. Поразительно однородная и последовательная “психометрическая” функция, связывающая средние оценки с фактическими корреляциями, не оставляет сомнения, что испытуемые использовали характеристики данных, которые были последовательно связаны с объективной корреляцией. Величина изменчивости в их оценках, однако, указывает, что используемые характеристики данных были только приблизительно связаны с объективными измерениями ковариации.

Мы подозреваем, и самоанализы некоторых неопытных испытуемых, которые выполняли это задание, подтверждают это, что впечатления испытуемых о “связанности” не отражают попытки рассмотреть полную выборку соответствующих двумерных наблюдений. Скорее, испытуемые могут полагаться на несколько специальных, возможно, экстремальных случаев переменной X или переменной Y . Таким образом, они могут просто наблюдать направление и экстремальность величины Y , связанной с экстремальностью величины X , и наоборот. Факторы, влияющие на внимание и память, могли бы также иметь значение; таким образом, наглядность и конкретность, так же как новизна представления, могли бы влиять на то, на какие моменты данных полагаются.

Уверенность в ограниченном количестве специальных или “тестовых” случаев, в частности, уверенность в крайних множествах, произвела бы последствия, сопоставимые с нашими результатами: во-первых, манипуляции со стимулами (то есть, числа – рисунков – аудиозаписи), которые были спланированы, чтобы изменять потребность в оценке величины и запоми-

нения до оценки ковариации, должны были оказать мало воздействия, так как эти дополнительные требования были восприняты тривиальными, если испытуемые просто полагались на очень ограниченное число отдельных случаев. Во-вторых, в зависимости от того, какие отдельные случаи они принимали во внимание, испытуемые, сталкивающиеся с непрочными связями, как ожидалось, могли оценить их, скорее, как умеренно положительные или как умеренно отрицательные, чем как равные нулю. Наоборот, испытуемые, сталкивающиеся с прочными связями, могли бы показать вполне последовательные и крайние оценки, так как положительные связи допускают мало вариативности в значениях одной переменной, которая может быть связана с крайним значением другой переменной. Форма нашей психометрической функции, а также сокращенная вариативность, связанная с оценками самой прочной эмпирической связи, совместимы с теоретически допущенными последствиями того, что люди полагаются на экстремальные случаи, хотя такие сведения не убедительны. Ясно, что более определенные ответы могут быть получены только из исследовательских проектов, в которых значением экстремальных (или непропорционально доступных) случаев сознательно манипулировали.

Стратегия оценок, основанных на теории. Когда мы переключаем наше внимание на оценки, основанные на теории, вопрос, который сразу же возникает – это насколько данные могут быть использованы в добавление к чистой теории. В некоторых случаях, вероятно, что ни к каким данным не будут обращаться. Испытуемые, например, могли полагаться только на интуицию, основанную на теориях или семантике, и ожиданиях; то есть применяли критерий репрезентативности (ср. Kahneman и Tversky, 1973, 4), рассматривая связи между двумя переменными в каждой паре. Мало кто из читателей будет оспаривать утверждение, что люди придерживаются социальных теорий подобного происхождения (или, по крайней мере, могут без труда создать). И действительно, любой обыватель может создать новые теории или дать прогнозы о функциональных связях слишком быстро, чтобы такие теории зависели от создания и анализа фактических случаев.

Однако мы не подвергаем сомнению возможность, что *некоторые* оценки ковариации, основанные на теории, подразумевают обращение к фактическим сведениям. Опять же, однако, мы подозреваем, что обыватель, более вероятно, будет полагаться на специфические тестовые случаи, подверженные многочисленным предубеждениям в кодировании, хранении и вспоминании, чем производить и неформально анализировать некоторую полную выборку двумерных данных.

Некоторые теории или убеждения относительно специфических связей могут, таким образом, частично основываться на данных. Другие могут основываться на выводах из более обширных убеждений о мире, народной мудрости, семантических ассоциациях или убедительности общения с семьей, друзьями или информации, полученной из средств массовой инфор-

мации. Независимо от их происхождения, однако, ясно, что многие убеждения являются как ошибочными, так и способными закрепляться и, возможно, даже усиливаться, несмотря на сведения, которые вызвали бы серьезные сомнения у любого непредубежденного наблюдателя. Конечно, у любого непредубежденного наблюдателя, у которого был бы калькулятор, вводный текст по статистике, и некоторые обычные знания относительно того, как их использовать. Поэтому мы закончим эту главу, кратко задав пару близких вопросов о столкновении между интуитивными теориями и ежедневным опытом с эмпирическими ковариациями, которые “проверяют” эти теории: во-первых, каковы механизмы, с помощью которых ошибочные или высоко преувеличенные убеждения относительно функциональных связей могут закрепляться, несмотря на убедительные логические опровержения? Во-вторых, как может наше описание недостатков обывателя быть приведено в соответствие с очевидной способностью организмов к обучению (в частности, инструментальное обусловливание и реакционное обусловливание и с очевидной точностью и адекватностью многих наших каждодневных убеждений и социальных стратегий?)

Этих вопросов можно касаться только здесь, и мы отсылаем заинтересованного читателя к более всесторонним обсуждениям в Главе 9 этого издания и в других публикациях (например, Nisbett и Ross, 1980).

Механизмы устойчивости теории

Теории об эмпирических связях, таких как впечатления об отдельных людях, часто показывают удивительную способность сохраняться сталкиваясь с эмпирическим опровержением. Одно из множеств механизмов, которые объясняют такую устойчивость, происходит из простого факта, что люди обычно действуют в соответствии со своими убеждениями; эти действия могут варьироваться от простой общественной защиты до вложения времени, энергии, богатства или репутации. Такое поведенческое обязательство мешает субъекту усовершенствовать свои убеждения в свете новых сведений (Abelson и другие, 1968; Festinger, 1957, 1964). Действительно, убеждения или ожидания человека могут превратиться в самосбывающиеся предсказания, посредством чего первоначально необоснованные убеждения заставляют его вести себя так, чтобы произвести исходы или данные, которые полностью подтверждают его убеждения (Merton, 1948; Rosenhan, 1973; Rosenthal & Jacobsen, 1968; Snyder & Swann, 1976; Snyder, Tanke, & Berscheid, 1977).

Недавнее исследование, проведенное Леппером, Россом и их коллегами, проиллюстрировало дополнительные механизмы или процессы, которые могут лежать в основе устойчивости убеждений. Эти механизмы включают наклонности “наивного” ученого как к ассимиляции недавно рассмотренного отрезка информации как функции их соответствия с предшествующими убеждениями, так и к выходу за пределы ассимиляции данных к разра-

ботке причинных объяснений того, *почему* предполагаемое состояние сохраняется (см. Главу 9 этого издания Anderson, Lepper, & Ross, 1980; Jennings, Lepper, & Ross, 1980; Lord, Ross, & Lepper, 1979; Nisbett & Ross, 1980; Ross, 1977; Ross, Lepper, & Hubbard, 1975; Ross, Lepper, Strack & Steinmetz, 1977).

Выявление ковариации, обуславливание и другие успехи реальной жизни

Наше нелестное описание способностей обывателей выявлять и оценивать ковариацию, может с первого взгляда показаться несоответствующим огромному количеству лабораторных сведений и даже еще более обширной базе ежедневного опыта, иллюстрирующего классическое и инструментальное обуславливание. Каждая крыса, которая научилась убегать из мышеловки, каждый ребенок, который оценил преимущество добавления сахара в кашу, каждый возлюбленный, который понял, что честность – не всегда хорошая политика, каждая мать, у которой на плач ее ребенка возникал рефлекс кормления грудью, является доказательством того, что организмы действительно распознают ковариации среди стимулов окружающей среды. И разве работа Чепманов, а также наши работы, касаются некоторого очень узкого и сокращенного класса задач оценки ковариации, сталкиваясь с которыми, обыватель проявляет однозначные недостатки? Нисбетт и Росс (Nisbett & Ross, 1980) обсудили эту проблему и пришли к выводу, что ответ на этот вопрос отрицателен. Они заявляют вместо этого, что именно явления обуславливания, являются “исключением”, и именно ограниченная возможность выявить и оценить ковариацию, описанная в этой главе, иллюстрирует “правило”.

Мы не можем подробно рассмотреть соответствующие аргументы, но сущность исследования Нисбетта и Росса в том, что классическое и инструментальное обуславливание получено при строго ограниченных обстоятельствах, включающих значимость и выраженность стимула, оптимальный интервал между стимулами и моментами применения стимула, или отсутствием события, несоответствующего стимулу или рассеивающего стимул. Два дополнительных фактора, которые они упоминают, заслуживают особого внимания. Во-первых, стоит отметить, что в то время как обуславливание может сохраняться с относительно низкой ковариацией между обусловленным стимулом (ОС) и необусловленным стимулом (НОС) или реакцией и закреплением, оно почти неизбежно может быть получено при условии идеальной ковариации, то есть, за ОС или за реакцией неизбежно следует НОС или закрепление, и последнее никогда не наблюдается в отсутствие первого. Асимметрия между необходимыми условиями для приобретения обусловленной реакции и сохранением или “устойчивостью” такой реакции, таким образом, иллюстрирует, в некоторой степени, такую же асимметрию между оценкой ковариации, основанной на данных, и оценкой, основанной на теории, которая исследовалась в данной главе. То есть во время поддер-

жания состояния организм может продолжать ожидать, и возможно, даже чувствовать большую ковариацию между ОС и НОС или между реакцией и закреплением, чем объективно оправдано. Организм может принимать закрепленный опыт в номинальной стоимости при приписывании не-закрепленного опыта к изменению обстоятельств, влиянию третьих переменных или даже случая.

Возможно, самым важным является веское свидетельство, что как инструментальное, так и классическое обусловливание выигрывает от (а в некоторых случаях даже требует) хорошего соответствия между случайностями, которые необходимо изучить, и предшествующими теориями или ожиданиями, с которыми организм попадает в лабораторию. По крайней мере, ясно, что не все случаи ОС-НОС или реакция-закрепление одинаково усваиваются. Возможно, наиболее внушительные доказательства этого обеспечил Гарсиа и его коллеги (например, Garcia, McGowan и Greene, 1972). Эти исследователи сообщили, что крысы могут научиться на единственном испытании, избегать пробовать пищу, которая вызывает желудочно-кишечное расстройство, даже если интервал между едой и расстройством – целых 12 часов. И, напротив, у животного, которое заболевает через несколько часов после потребления пищи со знакомым вкусом, но новой формы, не наблюдается подобной обучаемости. С другой стороны, когда немедленная боль заменяет отсроченную болезнь как НОС, получается противоположный результат; то есть крыса с готовностью учится избегать новых форм, а не новых вкусов, когда такие стимулы немедленно сопровождаются ударом током. Как заключают Нисбетт и Росс (Nisbett & Ross, 1980), крыса может характеризоваться как обладающая двумя “теориями”, обе из которых хорошо подходят для фактических непредвиденных обстоятельств ее экологии: (а) отличительные вкусовые сигналы, которые сопровождаются более поздним (даже намного более поздним) расстройством, должны рассматриваться как подозрительные; (б) отличительные осязательные или пространственные сигналы, которые сопровождаются немедленной телесной болью, должны рассматриваться как подозрительные.

Таким образом, в более общих терминах (ср. Testa, 1974), организмы, как люди, так и крысы, будут видеть те и только те ковариации, которые их собственная история или история их вида предполагает видеть. Если не руководствоваться “теориями”, обнаружить ковариации становится очень трудно и вероятно, это произойдет только тогда, когда соответствующие корреляции приближаются к единице, и/или когда условия для обучения оптимальны в терминах самих факторов, которые были объяснены в лабораториях с помощью длинной и выдающейся линии между Хола и Скиннера.

Иногда, конечно, каждодневные обстоятельства являются оптимальными для обучения. Таким образом, в нашем каждодневном опыте мы изучаем, что делают бесчисленные выключатели, рычаги, кнопки и другие механизмы управления, и что обозначает изумительное разнообразие знаков, символов и сигналов, потому что соответствующие ковариации так близки к совер-

шенным. Одинаково важным является факт, что обыватель, подобно ученому, часто может “проверить” новые гипотезы, которые он принимает во внимание. Могут преднамеренно быть созданы выборочные данные, которые гораздо лучше подходят для данной логической задачи, чем выборки, предложенные случайным опытом и воспоминанием.

Действительно, наше влияние на окружающую среду все более и более зависело от нашей способности заменять относительно формальные инструменты для заключений на неформальные. Наш успех отражает наследие поколений обычных мужчин и женщин, которые тщательно отметили и записали результаты и, еще раньше, от поколений бесчисленных ученых, достигших мастерства в искусстве формального экспериментирования и статистического анализа.

16. Иллюзия контроля*

Эллен Й. Лангер

В то время как большинство людей согласится, что имеется много точек соприкосновения между навыком и удачей, все же необходимо достичь полного понимания того, насколько неразрывно связаны эти два понятия. В принципе, различие кажется ясным. В ситуациях навыка имеется причинная связь между поведением и исходом. Таким образом, успех в задачах навыка управляется. Удача, с другой стороны, является случайным событием. Успех в действиях удачи или случая, очевидно, не поддается контролю. Интересующая нас проблема – действительно ли это различие признается. Предположение, принятое здесь – то, что нет. В то время как люди признают концепцию случая, они ведут себя, как если бы случайными событиями можно было управлять. Если это правильно, интересно определить переменные, ответственные за эту путаницу....

Некоторые наблюдения говорящие в пользу утверждения, что люди обращаются со случайными событиями как с управляемыми, идут от социологов Гоффмана (Goffman, 1967) и Хэнслина (Henslin, 1967). При изучении методов игры в казино в Лас-Вегасе, Гоффман отметил, что работники казино, раздающие карты, в периоды возникновения неудач, подвергались риску потери своих рабочих мест. Хэнслин изучил игру в кости и отметил, что игроки в кости явно ведут себя так, как будто они управляют исходом броска. Они осторожно бросают кости, если они хотят выпадения низкого числа или бросают их резко для высоких чисел. Они полагают, что усилия и концентрация окупятся. Контроль можно также наблюдать и при заключении пари; например, всегда держите пари с человеком, который выглядит так, как будто у него все под контролем. Это поведение весьма рационально, если человек полагает, что игра требует навыка.

* Выдержки из книги, которая появилась в журнале "Journal of Personality and Social Psychology", 1975, 32, 311-328. Авторское право © 1975 принадлежит Американской Психологической Ассоциации. Переиздано в соответствии с разрешением

Если бы человек попробовал проявить контроль над случайным событием, можно было бы оказать влияние прежде, чем исход случая определен. Стрикленд, Левики и Кац (Strickland, Lewicki, Katz 1966) проверили это предположение. Испытуемые были привлечены в игру в кости, в которой они выбирали одно из множества альтернативных пари либо до того, как кости были брошены, либо только после броска, но прежде, чем исход объявлен. Они обнаружили, что испытуемые шли на больший риск, то есть делали большие ставки, когда держали пари перед броском, чем после броска.

Предыдущее исследование показывает, что люди часто не могут реагировать по-разному на управляемые и не поддающиеся контролю события. Однако, факторы, которые управляют этим иллюзорным процессом контроля, систематически не изучались. Один способ идентифицировать эти факторы состоит в том, чтобы исследовать характеристики ситуаций навыка. В ситуациях навыка люди участвуют в различном открытом и тайном поведении, результатом которого является увеличение вероятности успеха: выбор того, какие материалы являются соответствующими ситуации и какие реакции проявлять, ознакомление с этими материалами и реакциями, трата времени на обдумывание задачи для достижения возможных стратегий, которые могут использоваться, и трата усилий, будучи активно занятыми в задаче, чтобы увеличить шанс успеха. Кроме того, ситуации навыка имеют некоторые характеристики, не обязательно вызванные человеком, чтобы максимизировать вероятность успеха. Конкуренция – один из таких факторов.

Эти связанные с навыком факторы могут отвечать за появление иллюзии контроля. Иллюзия контроля определена как ожидание, что личная вероятность успеха будет выше, чем гарантировала бы объективная вероятность. Следующие исследования были предназначены для оценки эффективности этих относящихся к навыкам факторов в создании иллюзии контроля. В частности, исследование, которое будет описано, было предназначено, чтобы проверить следующую гипотезу: поощряя или позволяя участникам случайного события вести себя так, как они себя вели, если бы участвовали в ситуации, требующей навыка, увеличиваем ли мы вероятность появления ориентации навыка; то есть стимулируем ли мы иллюзию контроля. Таким образом, нужно уметь ввести любой из выше упомянутых аспектов ситуации навыка, – *выбор, стимул или обычность реакции, пассивное или активное вмешательство, конкуренция*, – в ситуацию, определяемую случаем, где участники больше не влияют на исход и случайное поведение, более соответствующие событию, требующему навыка.

Серьезное испытание этой гипотезы – внедрение этих факторов в ситуации такие, как лотереи, где исходы полностью определены случаем. Если эти факторы успешно стимулируют иллюзию контроля в этих механических ситуациях, то их влияние должно быть гораздо большим, когда они представлены в ситуации, где уже имеется элемент контроля....

Эксперимент 1: Влияние конкуренции на иллюзию контроля

Так как люди часто вовлекаются в конкуренцию, когда они оценивают свои навыки, вероятно, что введение этого связанного с навыком фактора в ситуации случая стимулирует иллюзию контроля. Количество контроля, которое человек фактически проявляет в создании успешного исхода в квалифицированном соревновании, изменяется как функция способности противника. Если люди реагируют на случайные события, в которых имеется конкуренция, как будто эти события были определены навыком, тогда иллюзия контроля должна также измениться как функция характеристик противника.

В следующем исследовании испытуемые конкурируют в случайной задаче против либо привлекательного, самоуверенного человека или неуклюжего и нервного человека. Если на задачу реагируют, как будто исходы не поддается контролю, то факторы, отличные от вероятности победы, играют большую роль во влиянии на ставки испытуемых. При таких обстоятельствах испытуемые, ставят на кон много при конкуренции против уверенного человека либо потому что он, как ожидается, поставит много, и испытуемые хотят казаться подобными ему, либо потому что риск ценится в нашем обществе (Wallach и Wing, 1968). Испытуемые могут также поставить много при игре против неуклюжего человека, чтобы казаться отличным от него или, опять же, потому что риск ценится. Однако они могут также поставить меньше, держа пари против неуклюжего человека, потому что он, как ожидается, поставит меньше, и таким образом, испытуемые, рискуют меньше, но все еще считаются азартными. В любом случае, испытуемые не должны ставить больше против неуклюжего человека, чем против уверенного в себе противника. С другой стороны, если, как было спрогнозировано, конкуренция стимулирует ориентацию навыка, то испытуемые будут держать пари на основе вероятности победы. Чем менее компетентен противник, тем более вероятно, что человек победит, следовательно, испытуемые должны ставить больше при конкуренции против неуклюжего человека, чем при конкуренции против уверенного в себе человека.

Метод

Испытуемые: 36 студентов старших курсов мужского пола, зарегистрированных на вводном курсе психологии в Йельском университете. Они были привлечены через рекламное объявление, которое предлагало зачет и шанс выиграть деньги за участие в исследовании, посвященном связям между когнитивными и физиологическими реакциями. Они были наугад распределены в одну из двух экспериментальных групп по 18 испытуемых в каждой.

Процедура. Когда испытуемый входил в комнату, в которой должен был проходить эксперимент, он находил ожидающего его противника, который был другим испытуемым. Противник, студент-старшекурсник мужского пола, не знающий об эксперимен-

тальной гипотезе, играл роль либо уверенного, либо неуверенного в себе человека (изображающий щеголя или растяпу).

Условие – противник щеголь. При этом условии противник был уверенным и общительным человеком, одетым в хорошо сидящую спортивную куртку. Он представлялся испытуемому и указал на записку, находящуюся в комнате. В записке говорилось, что экспериментатор скоро вернется, а также, чтобы испытуемые заполнили за это время краткую анкету. Чтобы показалось, что исследование касается психологических вопросов, в анкетном опросе спрашивалось относительно диеты, болезней в семье и т.п. Испытуемый и противник заполняли анкеты и общались в течение приблизительно 10 минут. Беседа была логически структурирована, но сосредоточена, главным образом, на спортивных событиях. После этого взаимодействия противник беспечно стучал в стену, которая отделила его и испытуемого от экспериментатора, чтоб дать знак возвращаться в комнату.

Условие – противник растяпа. При этом условии противник был довольно застенчивым, вел себя неловко, нервно подергивался, и был одет в спортивное пальто, которое было слишком мало для него. Во всех остальных отношениях события были идентичны условию противник-щеголь.

В обоих условиях противник снимал пальто прежде, чем экспериментатор входила в комнату. После того, как были принесены извинения за опоздание, экспериментатор просила испытуемых сесть и не разговаривать, в то время как она подготавливала материалы для исследования. Используя эти меры, было возможно заставить экспериментатора не принимать во внимание предшествующую экспериментальную манипуляцию. Испытуемый и противник сидели за столом, лицом друг к другу. После того, как экспериментатор положила на стол датчик напряжения, спирт, марлевые губки, электроды, электродный гель и ленту, она дала следующие инструкции:

Мы интересуемся влияниями некоторых моторных и когнитивных реакций на психологические реакции. В частности, мы интересуемся изменениями сопротивления кожи как функцией напряженных и не напряженных задач. Исследование было разработано так, чтобы Вы получили от него удовольствие, в то время как я получаю информацию, в которой нуждаюсь. У вас будет возможность, либо выигрывать, либо проиграть деньги, это должно быть весело, но нет никакой гарантии, что вы выйдете отсюда, получив дополнительные деньги. Хорошо, теперь первое, что я попрошу вас сделать, это чтобы вы примотали лентой электроды к вашим рукам. Я хочу поместить ее на руку, которой вы не пишете. Вы правша или левша? Не волнуйтесь, это не причинит вреда. (Экспериментатор приматывает лентой электроды, включает в розетку датчик, и приносит колоду игральных карт.)

Первое задание – игра в карты. Правила таковы: вы будете выбирать карту из колоды, и кто выберет более высокую, тот побеждает в этом раунде. Будет 4 раунда, и перед каждым вы запишете, сколько вы ставите на кон. Вы можете ставить где-то от 0 до 25 центов в каждом раунде. Вы покажете ваши ставки мне, но не друг другу. Не смотрите на карту, которую вы выбираете. Таким образом, ваши ставки и исходы не будут влиять на ваши психологические реакции в следующей задаче. Позже я переверну карты и подсчитаю, сколько было выиграно или проиграно, с каждым из Вас индивидуально. Пари заключается только между одним из вас и мной непосредственно, так что если Вы побеждаете, я плачу вам, а если Вы проигрываете, вы платите мне либо в деньгах, либо временем. Вы хотите участвовать? (Противник быстро отвечает “конечно!”) Хорошо, теперь, мы можем начинать. Не записывайте, сколько вы поставили на кон, пока я не скажу, чтобы я могла прочитать базовую линию на аппаратуре.

Экспериментатор тогда попросила испытуемых записать ставки и показать их ей.

Ставки были записаны, и затем испытуемые, поочередно тянули карту и, по просьбе экспериментатора, одновременно показывали их ей, она записывала исход и затем клала карты рубашкой вверх на ближайший стол. Перед каждым шагом экспериментатор делала запись колебаний сопротивления кожи. Процедура повторялась для четырех попыток.

Зависимые измерения и проверка манипуляции. В качестве зависимого измерения выступало количество денег, которые испытуемые ставили на кон в каждом раунде.

После того, как игра в карты была закончена, испытуемым сказали, что следующая задача будет проведена индивидуально, поэтому одному из испытуемых необходимо пройти в другую комнату, где другой экспериментатор даст ему инструкции. Им также сказали, что как только этот эксперимент закончится, этот экспериментатор сообщит другому экспериментатору об исходе карточной игры так, чтобы долги могли быть улажены. Экспериментатор просил противника уйти и попрощаться с испытуемым, так как их объединенное участие было закончено. Каждому испытуемому тогда дали видоизмененную задачу так, чтобы он выиграл приблизительно \$2, независимо от его предыдущих ставок. Испытуемый исследовал сосуд с бобами и оценивал их количество, в то время как экспериментатор регистрировал колебания сопротивления кожи. Тогда испытуемому предлагали другую анкету, которая была психологической по характеру. После того, как его спросили, имело ли присутствие другого испытуемого какое-либо влияние на его психологические реакции, его попросили оценить другого испытуемого по шестибальной шкале в пределах от 1 (не очень компетентный в межличностном взаимодействии) к 6 (очень компетентного в межличностном взаимодействии). Оставшиеся вопросы были пунктами-наполнителями, которые имели отношение к психологическим вопросам. После того, как эти измерения были получены, всех испытуемых поблагодарили и попросили позвонить автору в следующем месяце, если они хотели узнать цель и результаты исследования.

Результаты

Перед исследованием, действительно ли количество поставленных на кон денег варьировалось как функция компетентности противника, важно удостовериться, что противник был действительно воспринят по-разному в двух условиях. Средняя оценка компетентности противника была 4.8, при условии противник-щеголь, и 3.17 при условии растяпа. Практически не было пересечений между двумя условиями. Различия между двумя средними значениями значительно ($t = 5.46$, $p < 0.005$). Поэтому, можно сказать, что испытуемые при условии противник-щеголь видели себя конкурирующими против более компетентного человека, чем испытуемые в условии растяпа.

Давайте вспомним, что испытуемые могли делать ставки в размере от 0 до 25 центов в каждом из четырех раундов пари. Эти четыре ставки были усреднены, чтобы дать отдельный счет для каждого испытуемого. Средняя ставка для испытуемых при условии противник-щеголь была 11.04 цента по сравнению с 16.25 центов для испытуемых в условии растяпы ($t = 2.39$, $p < 0.025$). Различия между этими двумя группами должно быть даже более очевидно, когда мы исследуем первые сделанные ставки, так как первый тур пари наиболее близко соответствовал экспериментальной манипуляции.

Средняя первая ставка для условия противник-щеголь была 9.28, в то время как средняя первая ставка для условия растяпы была 16.72 ($t = 3.16$, $p < 0.005$).

Концептуальное испытание манипуляции. Чтобы проверить предположение, что испытуемые из Йельского университета ожидают, что привлекательный противник поставит больше, чем непривлекательный, две анкеты были предъявлены случайным выборкам старшекласников Йельского университета. В первой анкете были описаны задача и участники, и испытуемых спросили, кто, как они думали, поставит на кон больше. Двенадцать из 16 испытуемых предположили, что привлекательный человек поставит на кон больше ($\chi^2 = 4$, $p < 0.05$). Во второй анкете описывалась задача, и у людей спросили, сколько бы они поставили на кон при каждом испытании. Все 15 человек опрошенных испытуемых ответили – максимальную ставку (25 центов).

Эксперимент 2: Влияние выбора на иллюзию контроля

И снова мы предположили, что, когда случайная ситуация уподобляется ситуации навыка, люди ведут себя, как будто они контролируют неуправляемое событие, даже когда факт, что успех или неудача зависят от случая, существенен. Лотерея обеспечивает идеальное средство для исследования иллюзии контроля, потому что, кроме решения покупать билет или нет, исход полностью управляется случаем. Если бы человек *мог* управлять исходом лотереи, можно было бы увеличить вероятность выбора билета. Этот билет имел бы тогда большую ценность, чем билет, принадлежащий человеку без этого контроля. И если бы он имел большую ценность, можно было бы требовать большую цену от потенциального покупателя.

В следующем исследовании лотерея проводилась чтобы, оценить влияние выбора, – важный фактор в ситуации навыка, – на иллюзию контроля. Было спрогнозировано, что испытуемые, которым давали выбрать лотерейный билет, потребуют более высокой цены за него.

Метод

Испытуемые. Билеты лотереи были доступны служащим офиса мужского и женского пола, которые работали в одной из двух фирм, расположенных в Лонг-Айленде, страховом агентстве и производственной компании.¹ Так как различные розыгрыши и спортивные фонды были не редкостью в этих офисах, оправдания для проведения данной лотереи были не нужны. За исключением четырех женщин, все служащие, к которым подходил агент, распространяющий лотерейные билеты, купили билет. Испытуемых наугад разделили на группы так, что в итоге было 24 мужчины и 3 женщины в условии выбора и 23 мужчины и 3 женщины в условии без выбора.

¹ Фирмы пожелали остаться анонимными

Материалы. Билеты лотереи были стандартными 4x2 дюйма (10.16 × 5.08 см) футбольными карточками. На каждом билете был изображен известный футболист, его имя и его команда. Карточки были упорядочены по алфавиту сначала по названию команды, а затем по имени отдельного игрока. Было два одинаковых набора билетов из 227 футбольных карточек. Каждый испытуемый хранил у себя билет из одного набора, а такой же билет из другого набора был отложен в картонную коробку, из которой победивший билет был бы позже отобран.

Процедура. Лотерея проводилась служащим страхового агентства мужского пола и служащим производственной фирмы женского пола за неделю до игры Суперкубка 1973 г. Оба экспериментатора не знали о гипотезе эксперимента. Каждый из них подходил к служащим соответствующего офиса и спрашивал их, не хотят ли они купить лотерейный билет, стоимостью \$ 1. Испытуемым сказали, что билеты продавались также в другом офисе (говорилось название другого офиса) и что полный выигрыш, приблизительно \$ 50, уйдет победителю. Испытуемым также сообщили дату розыгрыша. Согласившись участвовать в лотерее, первому испытуемому была дана коробка с билетами, и ему (ей) сказали выбрать билет(ы), какой(ие) ему(ей) больше нравился(ась). Испытуемые называли билет так, чтобы экспериментатор мог выбрать тот же самый билет из второго набора и положить его в закрытую картонную коробку. В это время экспериментатор также записывал имя испытуемого и выбранный им билет. Со вторым испытуемым повторили ту же процедуру, за исключением того, что после согласия принять участие в лотерее, ему давали карточку, которая соответствовала выбору предшествующего испытуемого. Испытуемые были, таким образом, в условии выбора или условии без выбора. На следующий день после того, как билеты были проданы в одном офисе, та же самая процедура была выполнена во втором офисе.

Зависимые измерения. Ко всем испытуемым индивидуально подходил экспериментатор, у которого они купили билет, в утро розыгрыша лотереи. Каждому из них сказали: "Один человек в другом офисе хотел принять участие в лотерее, но так как я больше не продаю билеты, он спросил меня, чтобы я выяснил, за сколько Вы бы продали свой билет. Для меня это совершенно не важно, но какую цифру я должен сообщить ему?" Указанная сумма составила зависимое измерение. Если испытуемый говорил, что он не будет продавать билет, экспериментатор был проинструктирован, чтобы подталкивать его, пока он не назовет сумму, и затем делать запись ответа "не продаст" рядом с количеством, которое он наконец предложил.

Результаты

Как было спрогнозировано, манипуляция выбора имела значительное влияние на цену билета лотереи. Среднее количество денег, за которое испытуемый хотел продать свой билет было \$ 8.67 в условии выбора, и только \$ 1.96 в условии без выбора ($t = 4.33$, $p < 0.005$). Хотя их спрашивали, за сколько они продадут свой билет, а не, хотят ли они его продать, 15 испытуемых первоначально ответили, что они его не продают. Из них 10 испытуемых были в условии выбора и 5 – в условии без выбора ($p < 0.10$). Упомянутое различие не являлось просто функцией сумм, указанных этими испытуемыми после понуждения, так как их ответы располагались от \$ 3 до полной суммы выигрыша в \$ 53, причем только 3 испытуемых были в последней категории.

Не будучи особенным образом проверенным до последующего исследования, один из результатов, полученных в этом исследовании, касается влияния осведомленности на иллюзию контроля. Женщины не столь знакомы с игрой в футбол, как мужчины. Следовательно, менее вероятно, что они согласятся играть в лотерею, а если они все же согласятся, они должны требовать меньшего количества денег, за свой билет. Нужно вспомнить, что только 4 человека отказались участвовать в лотерее, и что все они были женщинами. Из шести женщин, которые согласились участвовать в лотерее, четыре попросили \$ 1, и две попросили \$ 2 за свои билеты. Таким образом, средняя сумма для женщин была \$ 1.33, по сравнению с \$ 5.89 для мужчин ($t = 2.14$, $p < 0.05$)...

Выводы и их значение для практики

На основе только что предоставленных сведений, кажется, что испытуемые не отличают событий, определенных случаем, от событий, определенных навыками так, как предложено их определениями. Объективное непредвиденное обстоятельство не является критической переменной, определяющей поведение испытуемых. Вместо этого, они реагируют на событие так, как будто оно является управляемым, в значительной степени зависит от факторов таких как конкуренция, выбор, знакомство с ситуацией и вовлечение, которые могут быть ортогональными по отношению к фактическому непредвиденному обстоятельству. Как было показано, это очень важно даже в ситуациях, которые управляются случаем, таких как лотерея....

Почему это происходит? Люди заинтересованы в том, чтобы управлять своей окружающей средой. Важность контроля в этом контексте широко обсуждалась как психиатрами, так и исследователями в области социальных наук. Рассматривается ли это как потребность в компетентности (White, 1959), инстинкт обладания (Hendrick, 1943), стремление к превосходству (Adler, 1930) или стремления к личной причинной обусловленности (deCharms, 1968), большинство социологов согласны с тем, что человек заинтересован в том, чтобы управлять своей окружающей средой, и полная власть включила бы способность "победить случай", то есть управлять случайными событиями. Чем труднее проблема, тем более компетентным чувствует себя человек в способности решить ее. Поэтому самое большое удовлетворение или чувство компетентности следовали бы из способности управлять тем, что не поддается контролю....

В дополнение к заинтересованности в управлении, существует другая причина для отсутствия различия между управляемыми и не поддающимися контролю событиями. Это то, что навык и случайные факторы близко связаны в опыте людей. То есть существует не только заинтересованность в том, чтобы не различать их, но и существует трудность в различении, так как в каждой ситуации навыка присутствует элемент случая, и элемент навыка присутствует почти в каждой ситуации случая. Первое, очевидно и

не нуждается ни в каком дальнейшем объяснении. Примеры второго – это знание того, какую ставку лучше сделать в игре в кости (то есть, знание вероятности) или знание, какие игорные автоматы могут дать самый высокий выигрыш....

17. Результаты тестов – такие, какими Вы их себе представляете*

Лорен Дж. Чепман и Джин Чепман

Каждый день психиатры и клинические психологи должны принимать жизненные решения:

Какова его проблема? Нужно ли его поместить в психиатрическую больницу? Есть ли действительно риск, что он совершит убийство или самоубийство? Этого пациента уже можно выписывать или он должен остаться?

Для помощи в этих решениях клинические психологи почти всегда используют психологические тесты.

Согласно обзору Нормана Сандберга (Norman Sundberg), два наиболее широко используемые теста – тест чернильных пятен Роршаха и тест “Рисунок человека” (ТРЧ). Оба теста – проективные, основанные на предпосылке, что человек проектирует часть своей личности, когда он реагирует на неоднозначную, неструктурированную ситуацию. Например, так как нет никаких объективных форм в тесте Роршаха, то, что человек в нем видит, возможно, отражает его собственные стимулы, конфликты и индивидуальность. Точно так же, когда кто-либо рисует человека на чистом листе бумаги, он, как предполагают, проектирует часть себя в свое творение.

Собственная личность

Наше недавнее исследование говорит, что тесты Роршаха и ТРЧ могут быть проективными более, чем в одном направлении. В интерпретации результатов этих тестов, среднестатистический клинический психолог может проектировать свои собственные предубеждения и предположения в свое описание пациента.

* Эта книга первоначально появилась в *Psychology Today*, ноябрь 1971, стр. 18-22, 106-110. Авторское право ©. 1971 Ziff-Davis Publishing Co. Переиздано в соответствии с разрешением.

Наши первые исследования в этой области были с рисуночным тестом, в котором клинический психолог дает испытуемому карандаш и чистый лист бумаги и просит, чтобы он нарисовал человека. Карен Мачовер (Karen Machover) издала этот тест в 1949. Она описала рисунки, обычно нарисованные людьми с различными эмоциональными проблемами, и объяснила, как интерпретировать несколько характеристик рисунков как ключей к индивидуальности. Она сказала, например, что “человек, больной паранойей, придает много графического акцента глазам”, и “человеку того пола, у которого нарисована голова большого размера предоставлена большая интеллектуальная и социальная власть”.

Руководство к тестам Маховер отмечено далеко идущими обобщениями относительно того, какие виды людей рисуют какие виды рисунков, но она представила очень мало подтверждающих данных.

Части тела

Некоторые клинические психологи не желали принимать на веру работу Мачовер; они проверили ее утверждения экспериментально. Жюль Холзберг (Jules Holzberg) и Мюррей Векслер (Murray Wexler), например, пробовали определить, действительно ли люди, больные паранойей, лучше прорисовывают глаза. Они сравнили рисунки 18 пациентов, больных параноидальной шизофренией и 76 студентов, обучающихся на медсестер, но не нашли никакого различия в том, как эти две группы рисовали глаза.

Множество подобных исследований проверили прогнозы Маховер относительно других рисуночных характеристик – головы, ушей, губ, волос, одежды, рта, и т.д., — но снова и снова признаки ТРЧ не были подтверждены. Несколько экспериментаторов обнаружили, что более уравновешенные испытуемые имеют тенденцию рисовать более полные рисунки, но главное заключение из данных исследования – то, что определенное содержание рисунка не валидный показатель характеристик личности.

Признак

Необходимо указать, что этот тип исследования не требует совершенной дискриминации. Если 50 процентов гомосексуалистов рисуют фигуры одним способом, и только 25 процентов других людей рисует фигуры этим же способом, характеристика рисунка может все еще рассматриваться как валидный диагностический признак, так как в конечном счете, он может внести информацию к диагнозу гомосексуализма.

Большинство клинических психологов знает об исследованиях, показывающих, что признаки ТРЧ недействительны, все же многие тысячи продолжают использовать тест регулярно, потому что они заявляют, что признаки работают в их собственной клинической практике. “Я буду скорее доверять своим собственным чувствам, чем некоторой статье из журнала”, –

сказал один клинический психолог. “Я знаю, что больные паранойей, кажется, не рисуют большие глаза в исследовательских лабораториях, – сказал другой, – “ но они точно рисуют их в моем кабинете”.

Иллюзия

Некоторые критики говорят, что клинические психологи так захвачены своими теориями и традициями, что они не замечают фактов. Мы думаем, что существует другое объяснение. Клинический психолог, который продолжает доверять ТРЧ, даже если ему представить очевидные опровержения, может испытывать *ложную корреляцию*, явление, которое мы обнаружили несколько лет назад в исследовании относительно словесных ассоциаций.

Мы обнаружили, что слова, которые тесно связаны друг с другом, замечаются вместе чаще чем, это действительно бывает. В этих экспериментах испытуемый сидел на удобном стуле, в то время, как мы проектировали различные пары слов (например, *бекон — тигр*) на большой экран перед ним. Пары слов менялись каждые две секунды. Слово с левой стороны пары было всегда одно из четырех возможных слов: *бекон, лев, бутоны* или *лодка*. Каждое слово появилось так же часто, как любое другое (25%), но оно оказывалось всегда на левой стороне экрана. Слово на правой стороне пары было *яйца, тигр* или *тетрадь* с равной вероятностью появления.

Мы систематизировали пары слов так, чтобы каждое слово с левой стороны появилось равное количество раз с каждым словом с правой стороны. Например, когда слово *бекон* появлялось на левой стороне, слово *яйца* было в паре с ним в одной третьей части испытаний, *тигр* в другой трети испытаний, а *тетрадь* — в оставшейся трети. Но когда мы позже спросили испытуемых относительно пар слов, они сказали, что, когда слово *бекон* появлялось слева, слово *яйца* шло с ним в паре 47% от времени, и что, когда *лев* был слева, *тигр* был тем словом, которое наиболее часто появлялось справа. Даже при том, что каждая пара слов появлялась так же часто, как любая другая, испытуемые заявляли, что пары с сильной вербальной ассоциацией появлялись более часто, чем другие.

Мы назвали тенденцию видеть две вещи, появляющимися вместе более часто, чем они фактически появляются, *ложной корреляцией*.

Видимо, имеется существенное подобие между студентами, которые заявляют, что некоторые слова появляются вместе чаще, чем на самом деле, и клиническими психологами, которые утверждают, что видели валидность в признаках теста ТРЧ, когда исследование говорит, что ее нет.

Интерпретация

Признаки ТРЧ и интерпретации могут отличаться сегодня от того, какими они были, когда Махвер представила тест, более чем 20 лет назад. Поэтому

мы спросили современных ученых, как они используют тест. Мы послали анкетные опросы 110 клиническим психологам, которые являлись лидерами в диагностическом тестировании. Мы написали краткие описания шести типов пациентов и попросили, чтобы каждый психолог сообщил нам, какие характеристики он видел в рисунках каждого. Эти шесть описаний были: (1) “он обеспокоен тем, насколько он мужественный,” (2) “он подозрителен к другим людям”, (3) “он обеспокоен тем, насколько он умен”, (4) “он обеспокоен тем, что его кормят и о нем заботятся другие люди”, (5) “у него были проблемы с сексуальной потенцией” и (6) “он очень обеспокоен тем, что люди говорят о нем плохо”. Мы попросили психологов определить в каждом случае, какой пациент, нарисовал какого человека.

Мы получили 44 заполненных анкеты, и было ясно, что психологи соглашались друг с другом относительно характеристик рисунка, которые они видели в каждом случае. Например, большинство психологов (91%) сказали, что подозрительный пациент будет рисовать большие или необычные глаза. 82% сказали, что человек, обеспокоенный своим интеллектуальным уровнем, будет иметь тенденцию рисовать большую или выделенную голову (см. табл. 1).

Согласие не было абсолютным, но было внушительным. Вообще, психологи согласились по поводу двух или трех характеристик рисунка, которых они будут ожидать от каждого типа пациента.

Пары

Большинство психологов имело кандидатскую степень и в среднем 8,4 лет опыта в психодиагностике. Интересно, какие признаки ТРЧ обнаружили бы наблюдатели, у которых не было почти никакого опыта.

Чтобы выяснять это, мы собрали 45 рисунков мужских фигур – 35, нарисованных психическими больными в близлежащей государственной больнице и 10 рисунков, выполненные аспирантами в области клинической психологии. Мы измерили размер головы в каждом рисунке, размер глаз и т.д., и попросили независимых людей оценить рисунки на предмет более субъективных характеристик, таких как муже- или женоподобие.

К каждому рисунку мы приложили два из шести диагнозов, которые мы высылали психологам – например, “человек, который нарисовал это: (1) подозрителен к другим людям и (2) имел проблемы с сексуальной потенцией”. Имелось 15 отличных пар, которые могли быть получены из этих шести утверждений, так что мы использовали каждую пару на трех различных картинах.

Мы приписывали диагнозы систематически всем типам рисунков. Например, предложение: “он обеспокоен тем, насколько он умен” появлялась так же часто на рисунках с маленькими головами, как и на рисунках с большими головами.

Затем мы отобрали группу студентов колледжа и выбрали 108 человек, которые заявили, что они никогда не слышали о рисуночных тестах и не знают ничего относительно того, как они интерпретируются.

Мы опросили студентов в группах. Перед каждым опросом мы кратко объяснили суть ТРЧ. Мы сказали, что студент увидит ряд рисунков, наряду с краткими утверждениями относительно диагноза людей, которые нарисовали их. Мы сказали, что у многих людей были одинаковые проблемы, и что студенты должны тщательно исследовать картинку и поискать общие характеристики в рисунках людей с одинаковой проблемой. Студенты просмотрели рисунки в заранее спланированном случайном порядке, останавливаясь на каждом в течение 30 секунд.

Подтверждение

Хотя мы тщательно сбалансировали картинку и диагнозы так, чтобы не было никаких объективных связей между ними, почти каждый испытуемый сообщил, что он *видел* эти связи. И связи, которые студенты обнаружили, были подобны тем, которые видели психологи в своей ежедневной практике. Имелись некоторые различия, конечно, но студенты имели тенденцию описывать обычный рисунок каждого типа пациента в тех же самых терминах, которые использовали психологи. И в случае студентов, мы знаем, что признаки были иллюзорны, потому что о них не сообщалось в данных.

Наше предыдущее исследование относительно пар слов предлагает объяснение: вспомните, что мы обнаружили, что слова с сильными ассоциативными связями имеют тенденцию быть замеченными появляющимися вместе. Возможно, тот же самый механизм был и у признаков ТРЧ. Мы провели анкетный опрос по ассоциации слов, чтобы определить, как близко области симптома (подозрительность, интеллект, импотенция, и т. д.) связаны с различными частями тела (глаза, голова, половые органы, мускулы, и т. д.). Вопросы имели следующую форму: “Тенденция слова ПОДОЗРИТЕЛЬНОСТЬ ассоциироваться со словом ГОЛОВА (1) очень сильная, (2) сильная, (3) умеренная, (4) маленькая, (5) очень маленькая, (6) никакой тенденции нет вообще”.

Мы провели анкетный опрос 45 студентов, которые не участвовали в других частях эксперимента. Вербальные ассоциации, о которых они сообщили, четко совпадали с ложными корреляциями, которые “наивные” студенты видели между характеристиками рисунка и симптомами. Устные ассоциации даже более близко совпадали с корреляциями, о которых сообщали практикующие психологи.

Оплата

В нашем следующем эксперименте мы опросили 56 человек в течение трех дней подряд, чтобы увидеть, осознают ли они, что нет никакой истинной

характеристики рисунков	обеспокоены, на- сколько они мужественны			подозрительно относятся к другим людям			обеспокоены, на- сколько они умны			обеспокоены тем, что их кормят и о них заботятся другие люди			имели проблемы с потен- цией			обеспокоены тем, что люди говорят о них плохие вещи		
	П	С	И	С	И	С	С	И	С	С	И	С	С	И	С	С	И	С
1. Широкие плечи, мускулы	80	76	0	6	0	8	0	0	12	0	0	25	31	0	6			
2. Необычные глаза	0	0	91	58	0	6	0	0	3	0	0	2	2	43	26			
3. Голова большая или выделена	0	5	0	13	82	55	2	7	7	0	3	9	10					
4. Выделен рот	0	0	7	5	0	1	68	8	8	2	1	5	5					
5. Лучше проработана область половых органов	14	5	0	0	0	0	0	0	0	55	8	0	0					
6. Необычные уши	0	0	55	6	0	3	0	0	0	2	0	64	7					
7. Необычное выражение лица	0	17	18	44	2	21	2	21	21	2	14	18	52					
8. Женственность, детскость	23	22	7	12	2	11	32	39	39	23	25	11	13					

9. Хорошо прорисованы волосы	23	13	2	2	2	2	8	0	1	11	6	0	3
10. Рисунок деталей	20	8	2	6	34	13	0	0	3	7	3	2	6
11. Пассивная поза	5	4	2	8	0	2	36	21	1	2	2	0	8
12. Пуговицы на одежде	0	0	0	0	0	0	23	1	0	0	0	0	0
13. Область половых органов подкорректирована	0	0	0	0	0	0	0	0	0	18	27	0	0
14. Фаллический нос, конечности	9	0	0	0	0	0	0	0	0	23	2	0	0
15. Тучность	0	2	0	1	0	0	7	16	0	4	0	0	1

Примечание: П – психологи, С – студенты. Приведенные характеристики – те, которые были упомянуты, по крайней мере, 15% психологов или студентов, по крайней мере, для одного симптома. Практически для каждой характеристики рисунок, симптом, с которым ассоциировали ее психологи, является тем же симптомом, с которым ее ассоциировали студенты.

корреляции между признаками и рисунками, если предоставить им возможность увидеть материалы тестирования больше, чем один раз. Корреляции наблюдались так же отчетливо в третий день, как и в первый. Мы начали понимать, насколько сильной может быть ложная корреляция, и заинтересовались, какие условия, если таковые вообще имеются, позволят преодолеть ее.

Мы опросили еще 41 человека индивидуально и позволили каждому смотреть на рисунок, так долго, сколько он хотел. Чтобы поощрить их изучать рисунки тщательнее, мы предложили \$ 20 студенту, чьи оценки будут наиболее точными.

Это не сработало. Студенты видели ложные корреляции так же настоятельно, как всегда.

Наконец мы остановились и дали испытуемым все возможности, которые могли только придумать, чтобы они могли проверить свое собственное восприятие. Мы предложили каждому испытуемому все множество рисунков, чтобы он сам их изучил; мы сказали ему, что он может смотреть на них в любом порядке столько времени, сколько он хотел. Он мог раскладывать рисунки в стопки и делать прямые сравнения. Он мог поместить все рисунки, сделанные подозрительными людьми, в одну стопку и искать подобие в них. Мы дали каждому испытуемому бумагу для заметок, карандаш и линейку; мы снова предложили \$ 20 человеку, чьи оценки будут наиболее точны, и мы дали каждому испытуемому копию заключительного анкетного вопроса, так что он мог видеть то, на какие вопросы ему придется отвечать.

Мужественность

В этих условиях ложная корреляция понизилась значительно для большинства признаков, но она не исчезла. Например, в нормальных условиях 76% студентов видели связи между людьми, волнующимся о своей мужественности и тенденцией рисовать мускулистые фигуры; в новых условиях, 45% все еще заявляли, что они видели связи, которых там не было. Ложная корреляция сильна и примечательно стойка к любым попыткам изменить ее.

Студенты даже утверждают, что видели типичные корреляции, когда рисунки были сложены в противоположном направлении. В одном исследовании, например, мы поместили утверждение, “Его волнует его уровень интеллекта” только на рисунках с *маленькими* головами; утверждение относительно подозрительности появлялось исключительно на рисунках с *маленькими* глазами, и т. д. Это несколько уменьшило иллюзорную корреляцию, но не устранило это. Шестнадцать процентов все равно заявили, что пациенты, которых волновал их уровень интеллекта, рисовали фигуры с большой головой, а 50% все еще видели отношения между людьми, волнующимися относительно мужественности, и тенденцией рисовать мускулистые фигуры – даже при том, что истинные связи были противоположными.

Из нашего исследования очевидно, что клинические интерпретации ТРЧ содержат сильный компонент иллюзорной корреляции. И решения, которые принимают психологи относительно их пациентов, могут быть проекцией собственных предвзятых мнений.

Пятна

Интересно, присутствовали ли иллюзорные корреляции в наиболее популярном испытании из всех – чернильные пятна Роршаха – и если так, будут ли они замечены так же ясно, как и истинные корреляции, те немногие признаки Роршаха, которые считаются валидными индикаторами некоторых характеристик личности.

В 50-летней истории теста Роршаха, многие психологи сообщили, например, что определенные ответы даются более часто гомосексуалистами, чем другими людьми. В 1949, Уильям Уилер (William Wheeler) суммировал 20 признаков гомосексуализма по Роршаху. Другие исследователи проверили признаки Уилера, но только 2 из 20 признаков были сочтены валидными более, чем одним исследователем. Один из них (номер семь), – реакция на четвертое чернильное пятно, – “человек или животное, искаженное, чудовищное или угрожающее.” Другой валидный признак, – восьмой у Уилера, – сообщение о странной фигуре животного-человека на пятой карте.

Признаки

Чтобы исследовать, как психологи фактически используют тест Роршаха, для диагностики гомосексуализма, мы послали анкету 76 психологам, попросив их, описать две перцепции, которые гомосексуальные пациенты обычно видят в 10 пятнах Роршаха. Из психологов, которые возвратили заполненные анкетные опросы, 32 сказали, что они видели протоколы Роршаха, касающиеся гомосексуалистов. Эти 32 специалиста описали несколько признаков Роршаха, но из тех, что они упомянули, наиболее частыми были (1) ягоды или задний проход, (2) гениталии, (3) женская одежда, (4) человеческие фигуры неопределенного пола, без ясных мужских или женских признаков, и (5) человеческие фигуры, как с мужскими, так и с женскими признаками. Все это – признаки Уиллера, которые не были подтверждены исследованием. С другой стороны, только два психолога упомянули валидный признак номер 7 – искаженная, чудовищная фигура, и ни один не упомянул другой валидный признак, номер 8 – “полу-человеческая-полу-животная” фигура.

Некоторые психологи видели признаки в тесте Роршаха, которых не было, и не могли увидеть признаки, которые там были. И вновь наше исследование с вербальными ассоциациями предлагает ключ к разгадке. Два валидных признака не интуитивны: гомосексуализм не ассоциируется ни с рычащими животными, ни с человеко-животными гибридами. Но гомосек-

суализм имеет высокую вербальную ассоциацию с пятью признаками, о которых клиницисты сообщали наиболее часто. Так или иначе, разумно ожидать, что гомосексуалисты будут склонны видеть ягодицы, женскую одежду или фигуры неопределенного пола в чернильных пятнах.

Идеи

Мы проверили эти понятия объективно, попросив 34 независимых студентов оценить, как сильно слово “гомосексуализм” ассоциирует с различными идеями. Их оценки совпадали – популярные, но невалидные признаки имеют более сильную вербальную ассоциацию с гомосексуализмом чем, два непопулярных, но валидных признака. Это предполагает, что признаки гомосексуализма, которые как утверждают психологи, они видели в тестах Роршаха, могли просто отразить их собственные предположения и ожидания.

Мы проверили это утверждение с помощью проекта, подобного тому, который мы проводили в рисуночном тесте. Мы получили несколько карт Роршаха, и к каждой мы приложили ответ – некоторый образ, который человек, предположительно, заметил на карте. Область карты, которая относилась к ответу, была обведена кругом.

На некоторых чернильных пятнах ответ был валидным признаком гомосексуализма (например, “гигант с морщинистыми руками”), на других, ответ был невалидным признаком (например, “шнурованный женский корсет”), а на третьих, это был нейтральный признак (например, “карта Испании”). Ниже ответа были два описания человека, который проходил тест. Мы выбрали эти описания из всех возможных пар из группы, состоящей из четырех описаний: (1) “он проявляет сексуальное влечение к другим мужчинам”, (2) “он полагает, что другие люди сговорились против него”, (3) “он испытывает грусть и депрессию в течение длительного времени” и (4) “у него сильное чувство собственной неполноценности.” Мы, конечно, были больше всего заинтересованы первым утверждением.

Смешивание

Как в исследованиях ТРЧ, мы систематически приписывали утверждения признака к картам так, чтобы не было никаких последовательных связей между любым из утверждений и любым признаком.

После того, как студенты просмотрели несколько карт, мы спросили их, какие образы увидели пациенты с каждым из четырех типов признаков. Студенты сообщили, что гомосексуалисты более часто видели ягодицы, гениталии, и т.д., – то есть, те же самые пять невалидных признаков, о которых сообщили психологи. Ни один из студентов не увидел связи между гомосексуализмом и двумя валидными признаками.

В более поздней вариации теста мы преднамеренно ввели отрицательную корреляцию в материалы теста, так, чтобы утверждение “Он испытывает

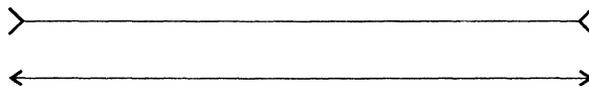
сексуальное влечение к другим мужчинам” никогда не появилось на карте, которая была воспринята как женская одежда, ягоды, и т.д. Это не уменьшило ложную корреляцию – студенты видели ее так же четко, как прежде.

Связь

Эти исследования показывают насколько легко полагать, что два независимых события связаны, особенно, когда существует некоторая субъективная вербальная ассоциация между ними. Наши испытуемые видели сильные ложные корреляции между симптомами и признаками проективных тестов в краткой, структурированной задаче. Задача психолога, безусловно, намного сложнее. Проблемы реального пациента многочисленны и неопределенны, – редко когда пациент имеет только два четко определенных признака. И реальные пациенты проявляют много различных реакций в проективных тестах, а не только одну. Также вероятно, что в реальной практике иллюзорные корреляции, которые психолог наблюдает, усиливаются сообщениями его коллег-психологов, которые непосредственно подвержены тем же самым иллюзиям. Согласие усиливает иллюзии. Нашим студентам, с другой стороны, не позволяли говорить друг с другом в течение теста, так что каждый увидел свои собственные ложные корреляции. Вот почему вероятно, что практикующие психологи имеют дело с ложными корреляциями, которые являются даже более сильными, чем те, о которых сообщили наши испытуемые.

Трудности

Мы не хотим сказать, что клинические психологи некомпетентны или слепы к фактам, как некоторые могли бы заключить. Наши данные указывают не на некомпетентность психолога, а на чрезвычайную трудность его задачи. Психологи подвержены тем же иллюзиям, что и другие люди. По аналогии, почти все говорят, что две горизонтальных линии имеют различные длины, когда они рассматривают рисунок Мюллера-Лайера (Miiller-Lyer):



но никто не называет плотника некомпетентным в оценке расстояния просто, потому что он также подвержен иллюзии.

Психологи должны знать о ложной корреляции, если они хотят избавиться от нее. В идеале, психолог первый, кто должен испытывать такие иллюзии. Нормальная политика обучения – требовать от каждого аспиранта в области клинической психологии, побыть наблюдателем в задачах подобных тем, которые мы описали. Он мог сам изучать величину и источник лож-

ных корреляций, которые он испытывает и, таким образом, как мы надеемся, научиться принимать меры против таких ошибок в своей клинической практике.

Опыт также напомнил бы ему, что он склонен ошибаться, основываясь на чувствах, что его клинические оценки должны постоянно проверяться с помощью объективных измерений, и что его профессиональная задача является одной из наиболее трудных и сложных в психологии.

18. Вероятностные рассуждения в клинической медицине: проблемы и возможности*

Дэвид М. Эдди

В значительной степени, качество и себестоимость здравоохранения определяются решениями, которые принимают врачи, чья окончательная цель состоит в том, чтобы разработать и управлять программой лечения для улучшения состояния пациента. Большинство решений включают много факторов, большую неопределенность и трудные вопросы, касающиеся ценностей.

Эта глава рассматривает один из аспектов того, как эти решения принимаются, изучая использование вероятностного рассуждения для анализа специфической проблемы: делать ли биопсию женщине, у которой имеется новообразование в груди, которое может быть злокачественным. В частности, мы изучим, как врачи обрабатывают информацию относительно результатов маммограммы и рентгеновского снимка для диагностики рака груди. Предоставленные данные показывают, что врачи не очень хорошо справляются с неопределенностью, что многие врачи допускают серьезные ошибки в вероятностном рассуждении, и что эти ошибки угрожают качеству медицинского обслуживания.

Проблема

Биопсия груди – не тривиальная процедура. Наиболее обычный вариант (около 80%) – ампутирующая биопсия, при которой подозрительная опухоль удаляется хирургическим путем для микроскопической экспертизы и гистологического диагноза патологом. Обычно пациент кладется в больницу и проходит полный набор дооперационных диагностических процедур. Биопсия почти всегда делается под общей анестезией (с вероятностью смерти от анестезии приблизительно в 2 из 10.000 случаев). Делается малень-

* Подготовка этой книги была поддержана грантом от семейного фонда Генри Дж. Кэйсера.

кий (от 1 до 2 дюймов) разрез и удаляется ткань размером от ореха до сливы. Во многих случаях (возможно 1 из 2) потеря ткани практически незаметна; в других – сохраняется небольшое углубление. В редких случаях (возможно 1 из 200) организм подвергается воздействию инфекции, которая может сохраняться в течение нескольких недель. Это стоит приблизительно \$700. Эта процедура может быть сделана в амбулаторных условиях и под местной анестезией. Как альтернативу к ампутиционной биопсии, некоторые хирурги предпочитают получать образец ткани, используя иглу. Это может быть сделано в амбулаторных условиях, без шрамов или других остаточных эффектов, и гораздо дешевле. Однако по мнению многих врачей, этот опыт является менее надежным, потому что злокачественное образование может быть пропущено.

Важный фактор, который определяет необходимость биопсии – возможность того, что опухоль в груди – раковая. Чтобы оценить эту возможность, врач может перечислить возможные болезни, оценить частоты, с которыми различные признаки и симптомы сопутствуют каждой болезни, сравнить эту информацию с результатами пациентки, оценить вероятность, что у нее есть одна из болезней в списке и провести биопсию, если вероятность рака или другого заболевания достаточно высока. Чтобы помочь врачу, многие учебники описывают, чем доброкачественные болезни отличаются от рака. Например, следующий абзац описывает одну такую доброкачественную болезнь – мастопатию.

Мастопатия часто путается с карциномой груди. Она обычно бывает у рожавших женщин с маленькой грудью. Наиболее часто она встречается в верхнем внешнем квадранте, но может находиться в других частях, и, в конечном счете, охватывать всю грудь. Она часто болезненна, особенно в предменструальный период, и обычно сопровождается менструальными расстройствами. Выделения из сосков, обычно серозные, наблюдаются в приблизительно 15% случаев, но отсутствуют какие-либо изменения непосредственно в соске. Новообразование не имеет четко ограниченной формы и не прилегает к коже. Многочисленные кисты плотные, круглой формы и колеблющиеся, а также могут просвечиваться, если содержат прозрачную жидкость. Большая киста при мастопатии на ощупь подобна опухоли, но обычно она более гладкая и хорошо ограниченная. Подмышечные лимфоузлы обычно не увеличиваются. При мастопатии редко наблюдаются большие синеватые кисты. Чаще наблюдаются многочисленные и маленькие кисты.¹ (Del Regato, 1970, с. 860-861)

Подобные описания существуют для фиброаденомы, жирового некроза, травмы и полдюжины других болезней груди, вплоть до рака.

Этот тип вероятностной информации может использоваться, чтобы помочь врачу проанализировать возможные причины новообразования в груди у пациентки. Рассматривая возможные исходы (например, верно диагностируя рак, делая ненужную биопсию доброкачественного новообразования, не делая биопсии и пропуская злокачественные опухоли, и правильно

¹ В этих и последующих цитатах добавлено выделение курсивом.

решая не подвергать биопсии доброкачественные опухоли), врач может оценить вероятность того, что пациентка, с ее специфическими признаками и симптомами, больна раком, и затем решить, как действовать дальше.

Использование маммографии

Существуют и другие диагностические процедуры, помогающие врачу оценить вероятность того, что грудное образование отдельно взятой женщины является злокачественным. Возможно, наиболее важная и обычно используемая из них – маммография. Суть этого испытания состоит в том, что компоненты раковых клеток поглощают рентгеновские лучи иначе, чем компоненты доброкачественных клеток. Изучая маммограммы, рентгенолог может увидеть некоторые признаки, которые наблюдаются с различными частотами в различных образованиях, и, исходя из этой информации, делается оценка относительно характера рассматриваемой болезни. Как правило, результат маммограммы классифицируется как положительный или отрицательный в отношении рака. Иногда используется расширенная схема классификации, такая, например, как схема, содержащая три класса: злокачественный, подозрительный и доброкачественный.

Эта диагностическая процедура не совершенна, так как в ней некоторые злокачественные повреждения неправильно классифицируются как доброкачественные, и некоторые доброкачественные опухоли классифицируются как злокачественные. Таким образом, один из факторов, который является очень важным для врача – точность диагностики.

Вероятностное рассуждение

Давайте рассмотрим это понятие более подробно. Цель диагностической процедуры состоит в том, чтобы обеспечить врача информацией о состоянии пациента. Врач использует ее для пересмотра состояния пациента и определения, как действовать, исходя из новой оценки. Это может быть решение пройти дальнейшие диагностические процедуры, или если врач достаточно уверен в состоянии пациента, может быть принято решение о терапии. Существенно то, что врач может иметь различные степени уверенности о состоянии пациента. Врач будет собирать данные, чтобы усилить уверенность в том, что у пациента есть или отсутствует рак, и когда эта уверенность станет достаточно сильной (в контексте серьезности болезни и изменений в прогнозе и лечении), будут предприняты действия.

Мы можем связывать вероятность, то есть субъективную вероятность врача, что у пациента рак, с этой степенью уверенности. Результаты различных диагностических процедур, таких, как, например, маммография, могут существенно изменить уверенность врача или субъективную вероятность того, что у пациента рак.

Понятие субъективной вероятности или степени уверенности появляется в различных формах в профессиональном медицинском жаргоне. Например, один автор пишет, что “так как старшая возрастная группа имеет самую большую пропорцию злокачественных новообразований, в ней повышенный коэффициент подозрения на рак, по мнению врача, который сталкивается с таким пациентом” (Gold, 1969, с. 162). Другой автор заявляет, что маммограмма может уменьшить количество биопсий груди “во многих случаях, когда сложившееся мнение осматривающего врача относительно доброкачественной опухоли поддерживается точным маммографическим диагнозом о доброкачественности” (Wolfe, 1964, стр. 253). Третий описывает это так: “Если субъективное впечатление врача дает достаточно причин для подозрения карциномы, он будет вынужден прибегнуть к биопсии, несмотря на отрицательный результат маммограммы” (Clark и др., 1965, с. 133). Другие высказывания, которые отражают это понятие: “уровень уверенности” (Burne, 1974, с. 37), “впечатление злокачественности” (Wolfe, 1967, с. 138), “более положительный диагноз” (Egan, 1972, с. 392), и т.д. Эти утверждения не точны, потому что мало врачей формально ознакомлены с понятиями субъективной вероятности и анализа при принятии решений. Тем не менее, существует вполне достаточное свидетельство того, что понятие степеней уверенности является естественным для врачей и используется ими для определения, как поступать дальше.

Интерпретация точности маммограммы

Теперь рассмотрим пациентку с новообразованием в груди, которое врач считает доброкачественным. Пусть эта вероятность будет 99 из 100. Вы можете интерпретировать фразу, что “врач считает, то вероятно (99 из 100) новообразование доброкачественно” следующим образом. Предположим, что у врача имеется опыт с множеством женщин, которые, во всех важных аспектах, таких как возраст, симптомы, история семьи и результаты анализов, подобны этой пациентке. И предположим, что врач знает из своего опыта, что частота рака в этой группе, скажем, 1 из 100. При отсутствии другой информации, врач припишет (возможно, подсознательно) субъективную вероятность 1% тому, что эта пациентка больна раком.

Теперь пусть врач посылает пациентку на маммограмму и получает сообщение, что, по мнению рентгенолога, опухоль является злокачественной. Это новая информация, и предпринятые врачом действия будут, очевидно, зависеть от его новой оценки вероятности того, что у пациентки рак. Врач, который обратится к литературе, может найти множество полезных утверждений, таких, как: “точность маммографии – приблизительно 90%” (Wolfe, 1966, с. 214); “(У пациенток с новообразованием в груди) положительный результат (маммограммы) карциномы высоко точен” (Rosato, Thomas & Rosato, 1973, с. 491); и “точность маммограммы при правильном диагностировании злокачественных опухолей груди составляет в среднем

от 80 до 85%” (Cohn, 1972, с. 98). Если необходимо больше подробностей, врач может обнаружить много утверждений, как “результаты показали, что 79,2% от 475 злокачественных новообразований, и 90,4% от 1. 105 доброкачественных новообразований были правильно диагностированы, и полная точность составила 87%” (Snyder, 1966, с.217).

На этом этапе Вы можете понять сложность проблемы врача, самостоятельно оценив новую вероятность того, что у этой пациентки рак: врач думает, что опухоль доброкачественна с вероятностью (99%), но рентгеновский диагноз был *положителен* с точностью, только что указанной.

Таблица 1. Точность маммограммы в диагностировании доброкачественных и злокачественных новообразований

Результаты рентгеновского обследования	Злокачественное новообразование (рак)	Доброкачественное новообразование (рак отсутствует)
Положительный	0.792	0.096
Отрицательный	0.208	0.904

Источник: числа взяты из работы Снайдера (Snyder, 1966).

Для оценки вероятности применяется формула Байеса. Согласно этой формуле

$$P(\text{рак} | \text{пол}) = \frac{P(\text{пол} | \text{рак}) \cdot P(\text{рак})}{P(\text{пол} | \text{рак})P(\text{рак}) + P(\text{пол} | \text{доброкач})P(\text{доброкач})}$$

Где

$P(\text{рак} | \text{пол})$ – вероятность того, что у пациентки рак, при условии, что у нее положительный результат рентгеновского обследования (апостериорная вероятность)

$P(\text{пол} | \text{рак})$ – вероятность того, что если у пациентки рак, рентгенолог правильно диагностирует его (истинно положительная оценка или чувствительность)

$P(\text{рак})$ – вероятность того, что у пациентки рак (априорная вероятность)

$P(\text{доброкач})$ – априорная вероятность того, что опухоль у пациентки доброкачественная ($P(\text{доброкач}) = 1 - P(\text{рак})$)

$P(\text{пол} | \text{доброкач})$ – вероятность того, что, если у пациентки доброкачественная опухоль, рентгенолог неправильно диагностирует ее как рак (ложно положительная оценка)

В табл.1 приводится итог чисел, представленных Снайдером (Snyder). Цифры в ячейках – соответствующие вероятности (например, $P(\text{пол}|\text{рак}) = 0.792$).

Используя оценку врача априорной вероятности того, что опухоль является злокачественной и, принимая во внимание новую информацию, полученную в результате процедуры, мы имеем

$$P(\text{рак}|\text{пол}) = \frac{(0.792)(0.01)}{(0.792)(0.01) + (0.096)(0.99)} = 0.077$$

Таким образом, врач должен оценить вероятность того, что у пациентки рак как приблизительно 8%.

Неправильное вероятностное рассуждение

К сожалению, большинство врачей (приблизительно 95 из 100 в неофициальной выборке, сделанной автором) неправильно интерпретирует утверждения относительно точности анализов и оценивает $P(\text{рак}|\text{пол})$, приблизительно как 75%. Другие исследователи получили похожие результаты (Casscells, Schoenberger & S; Grayboys, 1978). Когда их спрашивали, допустившие ошибку врачи обычно отвечали, что они оценили вероятность рака, при условии, что у пациентки положительный результат рентгеновского обследования ($P(\text{рак}|\text{пол})$) приблизительно равна вероятности положительного результата рентгеновского обследования у пациентки, больной раком ($P(\text{пол}|\text{рак})$). Последняя упомянутая вероятность измерена в клинических исследовательских программах и очень известна, но именно первая упомянутая вероятность необходима врачу для принятия решения. По-видимому, что многие, если не большинство врачей, путают эти две вероятности.

На самом деле существует два типа точности для любого анализа, предназначенного, чтобы определить, действительно ли присутствует определенная болезнь. *Ретроспективная точность* касается $P(\text{пол}|\text{рак})$ и $P(\text{отр}|\text{не рак})$. (Сокращение “не рак” относится к случаю, когда пациентка не больна раком. Это может происходить, потому что у нее либо доброкачественная опухоль, либо она вообще здорова.) Эта точность, на которую обычно ссылаются в литературе по маммографии, определяется, обращаясь к рентгеновскому диагнозу после того, как истинный (гистологический) диагноз стал известен. Давайте, используем термин *спрогнозированная точность*, для описания $P(\text{рак}|\text{пол})$ и $P(\text{доброкач}|\text{отр})$, точность, важная для врача, у которого есть рентгеновский снимок пока еще не продиагностированного пациента и который хочет сделать прогноз состояния болезни пациента.

Путаница между ретроспективной и спрогнозированной точностью. Обзор медицинской литературы по маммографии показывает сильную тенден-

цию приравнивать спрогнозированную точность положительного результата рентгеновского обследования к ретроспективной точности рентгеновского снимка; то есть приравнивать $P(\text{рак}|\text{пол}) = P(\text{пол}|\text{рак})$. Существует много причин подозревать, что эта ошибка постоянно совершается. Во-первых, формулировки многих утверждений в литературе предполагают, что их авторы приравнивают прогнозирующую точность ($P(\text{рак}|\text{пол})$) и ретроспективную точность ($P(\text{пол}|\text{рак})$), о чем они сообщают в своих исследованиях. Например, в статье *Рентгенология* в 1964 говорилось, что, “полная правильность диагноза рентгеновского снимка была 674 из 759, или 89 процентов” (том. 84, с. 254). Автор статьи в издании “*Клиническое акушерство и гинекология*” (*Clinical Obstetrics and Gynecology*) в 1966 заявил, “Аш (Asch) обнаружил 90% -ую корреляцию маммографии с патологическими результатами у 500 пациентов” (том. 9, с. 217). “Совпадение между рентгеновским и патологическим диагнозом было 91.6%” (Egan, 1972, с. 379). Все эти утверждения подразумевают, что, если у пациента был положительный диагноз, обследование будет правильным, и у пациента будет рак в 90% случаев. Это не так.

Во-вторых, некоторые авторы явно допускают ошибку. Приведем цитату из выпуска “*Клиническое акушерство и гинекология*” (*Clinical Obstetrics and Gynecology*) 1972 года в статье “Маммография в ее перспективе”, в которой была предпринята попытка исправить некоторую путаницу, существующую в литературе: “Среди женщин, у которых подтвердилась карцинома груди и которым делали маммограмму, не существовало рентгеновского диагноза злокачественной опухоли приблизительно для одной пациентки из обследованных пяти. В таком случае, если на основе отрицательной маммограммы, мы приходим к решению отсрочить биопсию твердого новообразования в груди, то, возможно, в одном случае из пяти мы отсрочиваем биопсию злокачественной опухоли” (том. 134, с. 98). Автор неправильно утверждает, что $P(\text{отр}|\text{рак}) = 0.2$ подразумевает, что $P(\text{рак}|\text{отр}) = 0.2$. Его ошибка становится очень серьезной, когда он заключает, что “отсрочить биопсию клинически доброкачественного твердого новообразования груди, которое было показано как доброкачественное на маммографии, значит, сделать шаг назад в уничтожении карциномы груди для женского населения”. Вероятность того, что у такой пациентки рак, зависит от априорной вероятности, но она меньше чем 1 из 100. Его анализ более, чем в 20 раз ошибочен.

Журнал “*Хирургия, гинекология и акушерство*” (*Surgery, Gynecology and Obstetrics*) опубликовал в 1970 (том 131, стр. 93-98) выводы другой исследовательской группы, которая вычислила “корреляцию рентгеновского диагноза с патологическим диагнозом” следующим образом. Они взяли всех пациентов с гистологически подтвержденными диагнозами и разделили их на три группы на основе рентгеновского диагноза – “доброкачественный”, “карцинома” и “подозрение на карциному”. В группе с рентгеновским диагнозом “доброкачественный” (“отрицательный” в нашей терминологии),

было показано, что у 84% пациентов на самом деле обнаружились доброкачественные опухоли. Было также отмечено, что 87.5% из группы с диагнозом “карцинома” (или “положительный”) имели подтвержденные биопсией злокачественные новообразования. Таким образом, $P(\text{рак}|\text{пол})=87.5\%$ и $P(\text{доброкач}|\text{отр})=84\%$. Но авторы приняли эту спрогнозированную точность за ретроспективную точность. Они заявили, что “правильный диагноз маммограммы был сделан в 84% от случаев с *доброкачественными опухолями* и в 87.5% от *опухолей с карциномой*”. Фактически, истинно положительная пропорция ($P(\text{пол}|\text{рак})$) в этом исследовании была равна 66%, а истинно отрицательная пропорция ($P(\text{отр}|\text{доброкач})$) была 54%.

В письме к редактору журнала “*Национальный обозреватель*” (*National Observer*), выпуск от 11 сентября 1976 врач представил пять “наблюдений и фактов” в поддержку своего мнения, что “установившаяся практика маммографии (то есть, проекционная маммография) не отвечает интересам населения в целом в любом возрасте”. Вот первые наблюдения.

(1) точность экспертизы маммографии, как сообщают, между 80 и 90%, в зависимости от таких факторов как возраст пациентки, действительно ли она имеет фиброцистическую болезнь, тип рентгеновского оборудования, опыт рентгенолога, и, каково наше определение “точности”... Даже если мы заключаем, что точность – 85% (и я уверен, что не каждый рентгенолог в нашей стране может достигнуть такого значения в собственной практике), то это означает, что *15% женщин, которым сделали рентген, окажутся с неправильными интерпретациями результатов, или более вероятно, их маммограммы просто не смогут показать болезнь. Это означает, что 15% женщин ложно внушат чувство безопасности, если им скажут, что их рентгеновские снимки нормальны, если действительно они уже больны раком.* Трудно оценить вред, причиненный этой группе, поскольку им было бы лучше не получить информацию, чем иметь ошибочную. Если женщине сказать, что ее маммограмма нормальная, и ей не нужно проходить анализы еще в течение года, женщина, больная раком, вполне может проигнорировать уплотнение в своей груди, из-за которого иначе она сразу же пошла бы к врачу.

Существует несколько ошибок в рассуждении этого автора. Во-первых, “точность” маммографии не может быть выражена как единственное число. Допустим, автор имеет в виду, что и истинно-положительные, и истинно-отрицательные пропорции равны 85%.

Во-вторых, эти пропорции (85%) соблюдаются, если маммография используется, чтобы установить дифференциальный диагноз известных признаков и симптомов. Такие опухоли крупнее, чем опухоли, которые изучают с помощью рентгена – тот случай, который описывает автор. Более разумные оценки истинно-положительных и истинно-отрицательных пропорций в рентгеновских снимках – 60% и 98% соответственно.

В-третьих, даже используя 85%, мы находим несколько погрешностей в рассуждении. Рассмотрим второе предложение. Существуют два способа неправильной интерпретации: (а) у пациентки может быть рак и отрицательные анализы, $P(\text{рак}, \text{отр})$; или (б) ее анализы могут быть положитель-

ными, но у нее может не быть рака, $P(\text{не рак, положит})$.² Из элементарной теории вероятности, мы знаем что

$$P(\text{рак, отр}) = P(\text{отр} | \text{рак})P(\text{рак})$$

$P(\text{отр} | \text{рак})$ дополняет $P(\text{пол} | \text{рак})$ и поэтому равняется 0.15 в этом случае. Мы не знаем точно $P(\text{рак})$, но для совокупности прошедших рентген, мы вполне уверены, что она меньше чем 0.005. То есть менее 5 из 1. 000 женщин имеют бессимптомный, но обнаружимый с помощью маммограммы рак груди.

Таким образом,

$$P(\text{рак, отр}) \leq (0.15) \text{ от } (0.005) = 0.00075$$

Также,

$$P(\text{не рак, пол}) = P(\text{пол} | \text{не рак})P(\text{не рак}) \geq (0.15) \cdot (0.995) = 0.14925$$

Полная вероятность неправильной интерпретации (то есть, $P(\text{рак, отр}) + P(\text{не рак, пол})$) – это сумма этих двух чисел, которая составляет 15%, как утверждает автор. Однако это не подразумевает, что “вероятнее, их маммограммы просто не смогут показать болезнь”. $P(\text{рак, отр}) = 0.00075$ не больше, чем $P(\text{не рак} | \text{пол}) = 0.14925$. Она приблизительно в 200 раз меньше.

Другая проблема – то, что “точность” 85% не подразумевает, что “15 процентам женщин будет сообщаться ложное чувство безопасности, если им скажут, что их рентгеновские снимки нормальны”. Автор, кажется, пытается оценить $P(\text{рак} | \text{отр})$. Теперь по формуле Байеса,

$$\begin{aligned} P(\text{рак} | \text{отр}) &= \frac{P(\text{отр} | \text{рак})P(\text{рак})}{P(\text{отр} | \text{рак})P(\text{рак}) + P(\text{отр} | \text{нерак})P(\text{нерак})} = \\ &= \frac{(0.15)(0.005)}{(0.15)(0.005) + (0.85)(0.995)} = 0.00089 \end{aligned}$$

То есть если 10. 000 женщин, у которых не обнаружено симптомов, пройдут рентген, и если мы используем неверную оценку точности автора, 8. 458 из них покинет клинику с отрицательными диагнозами. Автор считает, что

² $P(A, B)$ – совместная вероятность того, что произойдет как событие А, так и событие В.

приблизительно 1. 269 из них будет сообщено ложное чувство безопасности. Фактически, только приблизительно у 9 это будет так. Значение было завышено приблизительно в 150 раз.

Таблица 2. *Наличие рака и результаты рентгена у 1000 женщин, у которых наблюдались положительные результаты после физического осмотра*

	<i>Женщины, больные раком</i>	<i>Женщины, не больные раком</i>	<i>Общее количество</i>
Женщины с положительными результатами рентгеновского обследования	74	110	184
Женщины с отрицательными результатами рентгеновского обследования	6	810	816
Общее количество	80	920	1.000

Примечание: истинно-положительная пропорция 0.92 ($P(\text{пол}|\text{рак}) = 0.92$) подразумевает, что из 80 женщин, больных раком, у 74 будут положительные результаты рентгеновского обследования, а у 6 — отрицательные. Из всех женщин с положительными результатами, $74/184$ больны раком или $P(\text{рак}|\text{пол}) = 74/184 = 0.4$ (то есть 40%).

Источник: числа взяты из работы Вульфа (Wolfe, 1964).

Наконец, введение фразы “если действительно они уже больны раком” искажает значение предложения. Фразы “ложное чувство безопасности”, “если им сказали, что их анализы рентгена нормальны”, и “если они уже больны раком” переводятся символически в $P(\text{рак}|\text{отр}, \text{рак})$. Эта вероятность равняется 1, а не 0.15.

Важность вероятности $P(\text{рак})$. В дополнение к путанице между двумя точностями, многие авторы, кажется, не понимают, что для проверки постоянной ретроспективной точности, значение результатов анализов для врача (прогнозируемая точность) зависит от начального риска рака у пациентки, прошедшей маммограмму. Даже если допустить, что истинно-положительные и истинно-отрицательные пропорции постоянны для всех исследований, надлежащая интерпретация испытательных результатов — вероятность того, что пациент с положительными (или отрицательными) результатами маммограммы болен раком — будет зависеть от распространен-

ности рака в совокупности, из которой был выбран пациент, от предварительной вероятности того, что у пациента рак. Это может быть чрезвычайно важно, когда мы сравниваем использование обследования в диагностической клинике (где у женщин наблюдаются признаки и симптомы болезни груди) с его использованием в рентгеновской клинике для женщин, у которых нет симптомов.

Важность этого доказывается следующим примером. Предположим, что задача врача – сделать маммограмму женщинам, у которых “положительные” результаты физического обследования. Частота рака у таких женщин, как было обнаружено в одном исследовании, была приблизительно 8% (Wolfe, 1964). В одном множестве маммограмм в этой совокупности были получены истинно-положительная пропорция – 92% и истинно-отрицательная пропорция – 88% (Wolfe, 1964). Пусть врач столкнется с пациенткой, которая, по его мнению, репрезентативна по отношению к этой совокупности (то есть, пусть $P(\text{рак}) = 8\%$). Предположим, что он отправляет ее на маммограмму и получает положительное заключение рентгенолога. Его решение послать пациентку на биопсию должно быть основано на новой вероятности, что у пациентки рак. Эту вероятность можно подсчитать, и она будет равна 40% (см. табл. 2). Разве отрицательный результат исключил бы рак? Вероятность того, что эта женщина больна раком, учитывая отрицательный результат рентгеновского обследования, – немного меньше 1%. Правила получения этой оценки показаны в табл. 2.

Теперь, предположим, что врач посылает на рентген женщину с подозрением на рак, у которой нет никаких симптомов и физическое обследование было отрицательным. Распространенность рака, который может быть обнаружен маммограммой у таких женщин, – приблизительно 10% (например, Shapiro, Strax & Venet, 1967). Для этого примера, пусть ретроспективная точность рентгена, будет той же, то есть в этой совокупности пациентов у нее снова будет истинно-положительная пропорция 92% и истинно-отрицательная пропорция (для диагноза доброкачественных опухолей) 88%.³ Литература приводит данные только касательно ретроспективной точности обследования у женщин, у которых рак и доброкачественные опухоли. В одном исследовании приблизительно 60% этих женщин вообще были здоровы (Wolfe, 1965). Таким образом, в этом случае,

$$P(\text{рак}|\text{пол}) = \frac{P(\text{пол}|\text{рак})P(\text{рак})}{P(\text{пол}|\text{рак})P(\text{рак}) + P(\text{пол}|\text{доброкач})P(\text{доброкач}) + P(\text{пол}|\text{здорова})P(\text{здорова})}$$

³Это – неверное предположение, так как “точность” изменяется по мере того, как изменяется исследуемая совокупность. Например, истинно-положительная пропорция более низкая при использовании обследования в совокупности женщин, у которых не наблюдалось признаков болезни, потому что раковые новообразования имеют тенденцию быть намного меньшими, и их тяжелее обнаружить. Это предположение сделано только для того, чтобы продемонстрировать важность $P(\text{рак})$.

$P(\text{доброкач})$, $P(\text{здорова})$ и $P(\text{пол|здорова})$ не рассматриваются в литературе. Это приводит нас к подозрению, что их важность в анализе этих проблем не понята. В этом примере, мы будем использовать данные, представленные Вульфом (Wolfe, 1965) и предполагать, что $P(\text{здорова})$ – приблизительно 60%, и $P(\text{доброкач})$ – приблизительно 40%. Мы также сделаем предположение в пользу маммографии и позволим $P(\text{пол|здорова})$ быть 0%.

Продолжая использовать этот пример, предположим, рентгенолог сообщает, что маммограмма у данной женщины без симптомов заболевания, положительна. Учитывая положительный результат маммографии, вероятность того, что у пациента рак ($P(\text{рак|пол})$ – приблизительно 1 из 49, или приблизительно 2.0% (табл.3). В предыдущем примере, в котором были женщины с симптомами болезни, $P(\text{рак|пол})$ была 40%. Таким образом, в зависимости от того, кто исследуется, может существовать приблизительно *двадцатикратная* разница в вероятности того, что у женщины с положительной маммограммой рак.

Таблица 3. *Наличие рака и результаты рентгеновского обследования у 1000 женщин, у которых не обнаружено никаких симптомов*

	<i>Женщины с раковыми новообразованиями</i>	<i>Женщины с доброкачественными новообразованиями</i>	<i>Женщины, не больные раком</i>	<i>Общее количество</i>
Женщины с положительным результатом рентгена	1	48	0	49
Женщины с отрицательным результатом рентгена	0	352	599	951
Общее количество	1	400	599	1.000

Примечание: истинно-положительная пропорция 0.92 подразумевает, что рентген обнаружит рак у одной женщины, которая больна. Истинно-отрицательная пропорция 0.88 для доброкачественной опухоли подразумевает, что из 400 женщин с доброкачественными новообразованиями, у 352 будет отрицательный анализ рентгена, в то время как у 48 рентген будет положительным. Таким образом, у 49 женщин будет положительный рентген, но только у одной рак, или $P(\text{рак|пол}) = 1/49 = 0.2$ (то есть 2%).

Это поднимает главный вопрос относительно медицинского рассуждения – при попытке оценить симптомы пациента, как должен врач использовать информацию относительно основной частоты возможных болезней в этой совокупности в целом? Медики, кажется, запутались в этой проблеме. С одной стороны, врачи утверждают, что относительная распространенность болезни *не* должна затронуть оценку вероятности того, что отдельно взятый пациент болен. Это понятие появляется в нескольких аксиомах, таких, как, “каждый пациент – особый неповторимый случай” и “статистика – для мертвых”. При обсуждении определенных проблем, эта идея иногда выражается неявно, как в утверждении, “у молодых женщин, очевидно, меньше злокачественных новообразований, что, однако, не должно влиять на индивидуальный случай” (Wolfe, 1967, стр. 138). Это может также быть утверждено в явной форме и представлено как правило, которому нужно подчиняться. Например, следующее появилось в учебнике по клиническому диагнозу, когда имеется следующая рекомендация: “Когда пациент консультируется со своим врачом по поводу недиагностированной еще болезни, ни он, ни доктор не знают редкая ли это болезнь, пока диагноз не поставлен окончательно. Статистические методы могут применяться только к многотысячной совокупности. Редкая болезнь либо есть у человека, либо ее у него нет; возможность наличия двух болезней полностью не соответствует проблеме постановки диагноза” (DeGowin & DeGowin, 1969, стр. 6).

С другой стороны, эти утверждения часто противоречат поведению врачей, которые пытаются, использовать эту диагностическую информацию. Посмотрите, какие принципы преподаются в медицинских учебных учреждениях: “Когда Вы слышите цокот копыт, думайте о лошадях, а не о зебрах”, “Обычные вещи происходят наиболее часто”, “Следуйте закону Саттона (Sutton): идите туда, где есть деньги”, и т.д. Кажется, что многие врачи осознают ценность информации относительно предшествующей вероятности болезни, но формальные уроки теории вероятности усвоены плохо. Без формальной теории, врачи склонны совершать те же самые ошибки в вероятностном рассуждении, которые наблюдались в других контекстах (Kahneman и Tversky, 1973, 4; Lyon и Slovic, 1976).

Выводы: Маммограммы и биопсии

Эти проблемы могут иметь важные практические последствия. Например, в упомянутых примерах два автора основывали свои заключения на неправильном вероятностном рассуждении. Один неправильно утверждал, что женщина с опухолью в груди, которая кажется доброкачественной при физическом обследовании и по результатам рентгеновского снимка все еще имеет вероятность 20% того, что у нее рак, и ей стоит порекомендовать пройти биопсию. Другой автор основывал свои рекомендации против рентгена на неверной оценке частоты, с которой женщины получили бы ложное чувство безопасности (то есть, рак, не обнаруженный маммограммой). Оба ав-

тора, возможно, пришли бы к тому же заключению, если бы они рассуждали правильно, но могли бы и не прийти.

Ценность диагностической информации. Ценность маммографии у женщин, которые имеют признаки и симптомы болезни груди, лежит в ее способности обеспечить диагностическую информацию, которая повлияет на решение врача относительно биопсии. Точнее, исход обследования должен изменить оценку вероятности того, что у пациента рак. Как пишет один автор:

Маммография может помогать врачу в различении доброкачественных и злокачественных новообразований... У некоторых новообразований, особенно небольших, может отсутствовать достаточно характеристик, которые дают врачу высокую степень подозрения, чтобы оправдать биопсию. Именно здесь ... маммограмма может обеспечить дополнительные объективные данные. Таким образом, в случае неопределенной опухоли в груди, маммография может помочь врачу в принятии решения, применять ли биопсию (Clark & Robbins, 1965, стр. 125).

Чтобы любое диагностическое обследование было полезным, оно должно предоставлять информацию, которая может потенциально изменить решение о том, что делать с пациенткой – направить на биопсию тех, кого иначе бы туда не направили, и, мы должны надеяться, не допускать биопсии у некоторых женщин, которые иначе были туда направлены. Это понятие развито в статистической теории принятия решений и использовалось для анализа некоторых медицинских проблем в исследовательских условиях (например, Lusted и др., 1977)

Многие врачи признают, что результаты рентгена несут полезную информацию, которая должна помочь при общении с пациентом, но как именно информация должна использоваться, не указывают. Объяснения, данные большинством авторов, содержат мало определенных направлений. “Маммография не предназначена для предписания лечебных процедур, но она может обеспечивать, в некоторых случаях, только ту частицу более точной информации, которая помогает избежать нежелательных последствий” (Egan, 1972, с. 3192). “Маммография – ценное вспомогательное средство для хирурга при диагностировании и лечении опухолей в молочной железе” (Lyon, 1975, с. 231). “Маммография может помочь в разъяснении непонятных результатов, полученных после физического осмотра” (Egan, 1969, р. 146). Она “играет поддерживающую или вспомогательную роль...” (Block & Reynolds, 1974, с. 589). Точность и степень поддержки обычно оставлены на усмотрение врача.

Маммограммы и биопсии: Практика. Кажется, что роль маммографии в таких случаях понята только частично. Чтобы понять это, исследуем воздействие, которое, как клинические исследователи предсказывают, маммография будет оказывать на потребность в биопсии больной груди. В то время как утверждения, приведенные выше, подразумевают, что использование

рентгена должно помочь выбрать пациентов для биопсии, равное число утверждений говорит, что маммография не может, и на самом деле, не должна исполнять эту функцию. “Любая ощутимая опухоль требует проверки хирургическим вмешательством и биопсией, независимо от результатов рентгена” (Lesnick, 1966, стр. 2007). “В то время как результаты маммографии обычно однозначны, она не может заменить биопсию” (Egan, 1969, с. 148). “Никоим образом эта процедура не должна умалять важность биопсии. Фактически, использование установившейся практики маммографии вновь подтверждает важность биопсии, так как результаты рентгена, говорящие о злокачественном новообразовании, требуют биопсии для подтверждения.... Это никоим образом не умаляет важность биопсии.... Биопсия так же необходима для подтверждения результатов рентгена, как и для подтверждения физических симптомов” (Gershon-Cohen, и Borden, 1964, с. 2753, 2754). “Очевидно, что маммография не может заменить хирургию” (DeLuca, 1974, с. 318). “Позвольте нам решительно заявить, что маммография – не замена биопсии” (McClow и Williams, 1973, с. 618).

Одно из наиболее точных стратегических утверждений политики того, как маммография должна использоваться для выбора пациентов для биопсии, появилось в журнале “*Архивы хирургии*” в 1966 (*Archives of Surgery* том 93, с.853-856). Тщательное исследование различных направлений показывает, что используется только половина потенциала этой диагностической процедуры. Схема использования маммографии, “для того чтобы назначить лечение или установить, что делать с пациентом дальше” включает три категории пациентов:

Категория А: “пациентки с ‘уплотнением’ или ‘доминирующей опухолью’ в груди – прежде всего хирургическая проблема и не должно быть никакой задержки для прохождения биопсии. Маммография, в этом случае, является строго *дополнительным* средством. ... Она может показать скрытые опухоли” (стр. 854).

Категория В: “у пациенток наблюдаются симптомы, имеющие отношение к молочной железе, но никакого различимого уплотнения или ‘доминирующей опухоли’ нет. ... В этой категории, хирург и врач получит больше всего сведений от маммографии, потому что существует *подтверждающая* модальность”. Здесь маммограмма даст подтверждение и поддержку, “если впечатление от клинического осмотра — что опухоль доброкачественная. Однако она не должна изменить предшествующее мнение касательно биопсии” (с. 855).

Категория С: у пациенток отсутствуют признаки или симптомы, отсутствуют клинические признаки для биопсии, и маммограмма может только увеличивать количество биопсий.

Таким образом, автор выделил план, который сводит к нулю ценность маммографической информации в отборе пациентов, у которых биопсии можно избежать. Используется только та часть информации, где подразу-

меваются биопсия. Информация, которая могла бы воспрепятствовать проведению биопсии, игнорируется.

Маммограммы и биопсии: Потенциал. Чтобы оценить, как проблемы в вероятностном рассуждении могут затрагивать процесс медицинского обслуживания, давайте исследуем роль, которую маммография могла бы играть в дифференциальном диагнозе и в выборе пациентов для биопсии. Как описано выше, цель процедуры – изменить субъективную оценку человека, принимающего решения, относительно вероятности того, что у пациента рак. Если эта вероятность достаточно высока (как определено врачом и пациентом), рекомендуется биопсия. Назовем эту вероятность, *порогом биопсии*.⁴ Теперь рассмотрим влияние процедуры на управление двумя группами пациентов.

Первая группа состоит из тех пациенток которые, на основе истории болезни и физического осмотра, как думают врачи, больны раком. Используя данные, опубликованные Фридманом (Friedman, 1966), пусть предшествующая вероятность (частота) рака в этой группе, будет 90%. Если бы маммограмма была выполнена у такого пациента, положительный результат увеличил бы вероятность рака ($P(\text{рак}|\text{пол})$), возможно, до 95%. Отрицательный результат маммограммы все еще оставил бы для пациента вероятность наличия рака 71%. Эта высокая вероятность побуждает такие утверждения как: “Если субъективное впечатление врача дает достаточно причин для подозрения на рак, врач будет вынужден прибегнуть к биопсии” (Clark и др., 1965, с. 133). Вероятность злокачественности 71% – все еще достаточно высокая вероятность для того, что почти каждый захотел бы пройти биопсию.

Теперь рассмотрим вторую группу пациентов, у которых имеется опухоль, злокачественность которой не очевидна. В одном исследовании вероятность, что такая опухоль является злокачественной, была 14% (Friedman и др., 1966). При отсутствии дальнейшей информации, клиника проводит биопсию новообразования: “Если опухоль развивается, она должна быть удалена и тщательно исследована” (del Regato, 1970, с. 861). При использовании этой установки, давайте предположим, что порог биопсии пациента – 10%. То есть если, по предположению врача, вероятность, что у его пациента рак выше 10%, тогда пациент и врач соглашаются, что биопсия должна быть проведена.⁵ Используя порог биопсии 10%, мы можем опреде-

⁴ Любой, кто хочет убедиться в существовании порога биопсии может рассуждать, следующим образом. Вы согласны, что никто не захочет проходить биопсию, если вероятность рака – 1 из 30 триллионов? И мы можем согласиться, что фактически каждый хочет подтвердить диагноз и пройти лечение, если эта вероятность – 98 из 100? Если это так, то где-то между 1 из 30 триллионов и 98 из 100 у человека находится порог биопсии. Конечно, если женщина отказывается от биопсии и лечения даже, если вероятность рака велика, тогда у нее нет такого порога.

⁵ Порог биопсии – очень интересное и важное число. Его изменение оказывает большое влияние на количество женщин, подвергающихся биопсии, частоту непродуктивных биопсий, стоимость лечения пациента, а также прогноз о пациенте. Из-за рискованности, а также

лить влияние маммограммы на руководство 1.000 таких пациентов. Без обследования, все пациентки должны были бы пройти биопсию, причем 860 из них — непродуктивную. Приблизительная судьба изначальных 1.000 пациенток с новообразованием, при использовании маммограммы, представлена на Рисунке 1.⁶

потому, что, они принимают решение за другого человека, врачи, устанавливает порог биопсии весьма низко. Утверждение “если есть хоть какая-то вероятность, что опухоль злокачественна, биопсия должна быть сделана” является типичным. “Если врач полностью не уверен, что новообразование доброкачественно, оно должно быть подвергнуто биопсии без промедления” (Allen, 1965, с. 640). Согласно полученным сведениям, сами женщины вообще устанавливают порог выше, чем врачи, хотя бывает и наоборот.

Например, мы можем исследовать данные, взятые из большого клинического испытания, в котором использовались маммография и физическое обследование молочной железы, чтобы выявить рак у женщин, у которых не наблюдалось никаких симптомов. (Shapiro, Strax и Venet, 1971). В зависимости от того, как было обнаружено новообразование в молочной железе (т. е. с помощью какой процедуры или комбинации процедур), вероятность того, что болезнь молочной железы у женщины — рак, варьировалась от 15 до 54%. На основе положительного физического обследования, врачи порекомендовали 545 женщинам с отрицательными результатами маммограммы пройти биопсию. Несмотря на то, что частота рака в этой группе составляла 15%, 31% женщин отказался проходить биопсию. Частота рака у женщин, результаты физического обследования которых были отрицательными, а маммограммы — положительными равна 20%, но 29% таких женщин отказались проходить биопсию. У женщин с положительными результатами, как физического обследования, так и маммограммы вероятность рака 54%, но 5% женщин предпочли не проходить биопсию в рекомендованное время. Таким образом, из этой предварительной информации следует, что у более чем 31% женщин порог биопсии выше 15%, у 29% женщин порог биопсии выше 20%, а у 5% женщин порог биопсии превышает 54%.

⁶ Чтобы показать воздействие маммограммы на этих пациенток (а также пациентов с другими признаками и симптомами) необходимо большое количество информации, которая непосредственно недоступна в литературе. К счастью, в одном исследовании (Friedman и др., 1966) данные относительно частоты рака и ретроспективной точности маммографии представлены отдельно для трех групп пациенток — с очевидной карциномой, новообразованиями и пациенток с другими симптомами болезни молочной железы. Опубликованные данные, однако, неполны, и данные относительно частоты сомнительного рентгеновского диагноза доброкачественных и раковых опухолей не включены. Использовались данные, предоставленные в исследовании Фридмана, и для этого примера, были сделаны следующие предположения: (1) опухоли, не подвергнутые биопсии, были на самом деле доброкачественными, (2) опухоли, не подвергнутые биопсии, были закодированы как отрицательные, (3) половина доброкачественных новообразований, которые не были закодированы как отрицательные, были закодированы как положительные (другая половина была закодирована как неопределенная), и (4) половина злокачественных новообразований, которые не были закодированы как положительные, были закодированы как отрицательные. Первые два предположения — наиболее оптимистическая интерпретация точности маммографии. Третье и четвертое предположения очень важны, и по мере того, как ложно-положительная (или ложно-отрицательная) пропорция стремится к нулю, мощность положительного (отрицательного) результата рентгеновского обследования, включающее (исключающее) рак, растет. Аналогично, по мере того, как ложно-положительные или ложно-отрицательные пропорции растут, обследование теряет свою прогнозирующую мощность. Интерпретация данных Фридмана даже усложнилась при представлении ее в терминах молочных желез, а не пациенток. Тем не менее, существует много информации в этом сообщении, и разумно использовать ее в этом примере, при условии, что читатель понимает, что это — иллюстра-

Пациенты с положительным результатом маммограммы имеют 53% вероятности наличия рака и, так как мы предположили, что они имеют порог биопсии 10%, они должны были бы пройти биопсию. Поскольку вероятность того, что пациентка с сомнительной маммограммой больна раком равна 34%, эти пациенты также должны пройти биопсию. Пациенты с отрицательным результатом маммограммы имеют вероятность 4% наличия рака, и, так как это ниже их принятого порога биопсии (10%), они не хотели бы подвергнуться биопсии, но предпочтут тщательное наблюдение. Общее количество немедленных биопсий было уменьшено от 1.000 до 240. По крайней мере, еще 30 биопсий будут сделаны в конечном счете, потому что 30 из 760 оставшихся пациенток больны раком.

Таким образом, ожидаемая польза от наличия маммограммы (например, сокращения вероятности ненужной биопсии от приблизительно 86% до немного более 13%) может быть сравнена с затратами (например, опасностью поражения лучевой болезнью и приблизительно \$ 75), и небольшим уменьшением в ожидаемом выживании (существует 3% вероятность того, что диагноз злокачественного новообразования будет отложен на месяц или около того). Если использовалось понятие порога биопсии и некоторая простая теория вероятности, многим пациентам в этой группе, у которых был отрицательный результат маммограммы, не будет сделана биопсия. При отсутствии этого типа анализа существует "согласие между хирургами, что все пациенты (в этой группе) должны пройти биопсию, независимо от результатов маммограммы" (Friedman и др., 1966, с. 889)

1,000 пациентов	Положительные новообразования	140	Рак	74
			0,53	
			Нет рака	66
			0,47	
	0,14			
	Неопределенные новообразования	100	Рак	34
			0,34	
			Нет рака	66
			0,66	
	0,10			
	Отрицательные новообразования	760	Рак	30
			0,04	
			Нет рака	730
			0,96	
	0,76			

Рис. 1. Вероятность рака у женщин с доминирующими новообразованиями

ция, а не формальный анализ. Формальный анализ этих вопросов требовал бы лучших данных. Значения для точности, используемой в тексте для оценки пациентов в группе 2 следующие: $P(\text{пол}|\text{рак}) = 0.52$, $P(\text{неопр}|\text{рак}) = 0.24$, $P(\text{отр}|\text{рак}) = 0.24$, $P(\text{пол}|\text{доброкач}) = 0.075$, $P(\text{неопр}|\text{доброкач}) = 0.075$, и $P(\text{отр}|\text{доброкач}) = 0.85$.

Важность порога биопсии в этом примере должна быть подчеркнута. Если врач и его пациент установили порог в 1% – то есть, если бы пациент чувствовал, что вероятность 1 из 100 наличия рака достаточна, чтобы применить биопсию – тогда отрицательный результат маммограммы не устранил бы потребность в биопсии (вероятность рака 4% превысит этот порог). Маммограмма, возможно, дала бы клиницисту некоторую информацию, но эта информация не повлияла бы на решение о биопсии. Использование маммографии в этом случае должно было бы быть оправдано на других основаниях.

Пересмотр практики. Этот тип анализа помогает прояснить потенциальную полезность маммографии при постановке дифференциального диагноза различных опухолей. Это также помогает нам оценить следующие утверждения:

1. “Маммография оказывает мало влияния на решение относительно клинически (т.е. физически) осязаемой опухоли молочной железы, которая, на основе своих собственных характеристик, требует биопсии” (из *Archives of Surgery*, 1974, том. 108, стр. 589). В обследовании пациентов с новообразованиями, биопсия требовалась только на одних лишь клинических основаниях. Использование маммографии разбивает группу на подгруппы с частотами рака в пределах от 53% до 4%. Биопсии можно было бы избежать в последней группе, и количество биопсий могло бы быть уменьшено на 73% (от 1.000 из 1.000 до 270 из 1.000).

2. “Для клинических целей маммография должна обеспечить приблизительно 100 процентный уровень точности прежде, чем она одна может определять лечение” (из *Archives of Surgery*, 1974, том 108, с. 589). В совокупности, подобной второй группе, которая обсуждалась выше, могло бы быть весьма рационально позволить маммографии определять пациентов для биопсии. Вспомним, что истинно-положительная пропорция, используемая в том примере, была 52%, и что более точное испытание будет даже более ценно.

3. “Маммография – не замена биопсии” (из *Oncology*, 1969, том 23, с. 148). Цель и маммографии, и биопсии состоит в том, чтобы обеспечить информацию относительно состояния пациента. Некоторые пациенты, при отсутствии маммографии, требуют биопсии. У некоторых из этих пациентов отрицательный результат маммограммы устранил бы необходимость биопсии, и в этих случаях, маммограмма заменяет биопсию.

4. “Каждому решению о прохождении биопсии должна предшествовать маммограмма” (из *Oncology*, 1969, том 23, с. 146). Рассмотрим клинически очевидную карциному. Вероятность рака будет превышать порог биопсии каждого, почти независимо от того, каков результат маммограммы. Первичное оправдание этой политики, в таком случае, должно лежать в вероятности того, что клинически очевидная опухоль является доброкачественной (иначе пациентка была бы должна пройти мастэктомию (удаление молочной железы), и что имеется скрытое неосязаемое злокачественное ново-

образование. Вероятность этого сложного события равна сумме вероятностей двух событий, которые являются чрезвычайно маленькими (приблизительно 1 из 5.000).

5. “Отсрочить, биопсию *клинически доброкачественного* новообразования груди, которое было определено *доброкачественным на маммографии*, значит, сделать шаг назад в уничтожении карциномы груди” (из *Surgery, Gynecology and Obstetrics*, 1972, том 134, с. 98). Пусть “клиническая доброкачественность” будет представлена P (рак) – 5%. После отрицательного результата маммограммы, вероятность, что такой пациент имеет рак – приблизительно 1%. Из 100 биопсий, 99 были бы непродуктивны. Является ли отказ от биопсии здесь шагом назад или вперед? Другой момент – то, что, если бы этой стратегии следовали, все новообразования, начиная от “клинически доброкачественных” и заканчивая клинически очевидными карциномами, требовали бы биопсии, независимо от того, каков был исход обследования. Это, кажется, противоречит утверждению автора, что “когда используется в своем надлежащем виде, маммография является превосходным дополнением для врача в лечении карциномы груди” (из *Surgery, Gynecology and Obstetrics*, 1972, том 134, с. 98).

6. “Маммография никогда не должна использоваться вместо биопсии, когда мы имеем дело с ‘доминирующим новообразованием’ в груди, и никогда не должна изменить основной хирургический подход в болезнях груди, то есть, ‘опухоль – это опухоль’ и она должна быть подвергнута биопсии либо путем разреза, либо путем аспирации” (из *Archives of Surgery*, 1966, том 93, стр. 854). Пациентки с доминирующими новообразованиями и порогами биопсии выше чем 5% не согласились бы с этим утверждением.

7. “В сомнительных случаях полагаться на маммограмму ошибочно. Она необходима после обследования и прощупывания молочной железы, чтобы решить, будете Вы или нет делать биопсию, если рентген не доступен. Если Вы решили проводить биопсию – проводите ее. Если Вы уверены, что нет никаких симптомов для хирургического вмешательства или физического обследования, направьте пациентку на маммограмму. Как только врач говорит к себе, и особенно, если он говорит пациентке, “Я не совсем уверен, давайте пройдем рентген”, он подсознательно полагается на отрицательный результат маммограммы, когда нужно рассчитывать только на положительный. Это – психологическая ловушка, в которую все мы склонны попадать, и которая намного серьезней, чем некоторое количество ложно-положительных диагнозов, достигнутых при помощи маммографии” (Rhoads, 1969, с. 1182). Ни одной биопсии не удастся избежать, применяя эту стратегию. Это – позор, ведь сам автор вышеупомянутого утверждения говорит, что, “существует мало областей, в которых требуется так много хирургии, которой можно было бы избежать лучшими методами диагноза, чем в области болезней молочной железы”.

Мы теперь готовы оценить следующую историю, которая появилась в “Хрониках” в Сан-Франциско (Kushner, 1976). У женщины-репортера недавно обнаружили опухоль в молочной железе, и она описывает консультацию с врачом.

“Я хотел бы, чтобы Вы прошли ксеромаммограмму. Это новый способ делать маммограммы - снимки молочной железы.”

“А она дает точные результаты?”

Он пожал плечами, “Вероятно, столь же точные, как любой снимок. Вы знаете, - предупредил он, - “даже, если результат отрицателен - что означает, не злокачественное новообразование - единственный способ убедиться в этом состоит в том, чтобы вырезать опухоль и рассмотреть ее под микроскопом”.

Женщина обсудила эту проблему со своим мужем.

“Что сказал доктор?”

“Он хочет сделать ксеромаммограмму. Тогда, независимо от результата, опухоль нужно будет вырезать”.

“Так, зачем сначала нужно делать рентген?”

“Я думаю, потому, что так надо. И наш доктор говорит, что обследование дает правильные результаты приблизительно в 85% случаев.... Так, сначала я записалась на термограмму – она будет либо положительная, либо отрицательная, и если этот результат совпадает со снимками ксеромаммограммы, то статистика говорит, что диагноз был бы на 95% надежный”.

Подводя итоги, казалось бы разумно спросить, что, если цель маммографии состоит в том, чтобы помочь врачам отличить доброкачественную от злокачественной опухоли молочной железы, таким образом не допуская некоторых пациентов на такую более обширную и травмирующую процедуру как биопсия, тогда мы должны позволить обследованию выполнить эту функцию. Если с другой стороны, врач всегда должен придерживаться предшествующего решения о биопсии, и на него не влияет исход маммограммы, мы не должны настаивать на том, что цель этого обследования состоит – помочь отличить доброкачественную опухоль от злокачественной, так как этот вывод будет сделан окончательно исходя из результатов биопсии. Наконец, если цель обследования состоит в том, чтобы искать скрытый и клинически неподозреваемый рак в другой области молочной железы (далеко от прощупанной опухоли, которая нуждается в биопсии так или иначе), мы должны признать, что вероятность такого события чрезвычайно мала и что использование этого вида обследования равно использованию рентгена.

Моя цель не состоит в том, чтобы приводить доводы в пользу политики маммографии или биопсии – чтобы так поступать, нужны более качественные данные и лучшая оценка ценностей пациентов. Я только хочу предположить, что мы не развили формального способа вероятностного рассуждения относительно этого типа проблемы, что клинические оценки могут быть неверными, и что текущая клиническая стратегия может быть непоследовательна или неправильна.

Обсуждение

Эти примеры были представлены, чтобы проиллюстрировать сложность принятия решения в медицине и продемонстрировать, как некоторые врачи управляют одним из аспектов этой сложности – вероятностями. Случай, который мы исследовали – относительно простой, использование единственного диагностического обследования для распределения опухолей на две группы злокачественные и доброкачественные. База данных для этой проблемы относительно хороша. Точность и диагностическая ценность процедуры были исследованы и проанализированы во многих учреждениях в течение многих лет. Как выразился один исследователь, “я не знаю ни о какой другой медицинской процедуре, которая была бы больше проверена и перепроверена, чем маммография” (Egan, 1971, с. 1555).

Вероятностные методы, обсужденные в этой главе, были известны в течение столетий. В последних двух десятилетиях они применялись все больше и больше для медицинских проблем (например, Lusted, 1968), и использование систематических методов для управления неопределенностью возрастало в учебных планах медицинских факультетов, статьях журналов и образовательных программах для аспирантов. В настоящее время, однако, применение этих методов было спорадическим и еще не отфильтровалось, чтобы затронуть мышление большинства практикующих врачей. Как проиллюстрировано в этом социологическом исследовании, медицинские проблемы сложны, и сила формального вероятностного рассуждения обеспечивает большие возможности для улучшения качества и эффективности медицинского обслуживания.

19. Получение знаний из опыта и условно - оптимальных правил при принятии решения*

Хиллел Дж. Айнхорн

Текущая работа в исследовании принятия решения ясно изменила направление от описания процессов выбора с помощью нормативных моделей (и их модификации) к постановке акцента на эвристических процессах, развитых в пределах общей структуры когнитивной психологии и теорий обработки информации (Payne, 1980; Russo, 1977; Simon, 1978; Slovic, Fischhoff & Lichtenstein, 1977; Tversky & Kahneman 1974, 1, 1980). Перемещение внимания от вопросов, насколько хорошо люди действуют к тому, как они действуют, безусловно, важно (например, Hogarth, 1975). Однако, полезность изучения двух вопросов одновременно нигде так не очевидна, как в исследовании эвристических правил и стратегий. Причина для этого – то, что сравнение эвристических и нормативных правил позволяет исследовать несоответствия между фактическим и оптимальным поведением, которое поднимает вопросы, почему такие несоответствия существуют. В этой главе, я концентрирую внимание на том, как человек формирует оба типа правил из своего опыта. Интерес к изучению с помощью экспериментов ставит множество проблем, которым не было уделено достаточно внимания; например, при каких условиях изучаются эвристики? Как они проверяются и поддерживаются в данном эксперименте? При каких условиях мы не сможем узнать о предубеждениях, которые могут следовать из их использования?

Важность изучения для понимания эвристик и выбора поведения может быть обнаружена, принимая во внимание следующее:

* Это - сокращенная версия книги, которая появилась в редакции T. S. Wallsten. Когнитивные процессы при поведении выбора и принятии решений. Hillsdale, N.J.: Lawrence Erlbaum Assoc., Inc. Переиздано в соответствии с разрешением.

Это исследование было поддержано грантом от Отделения психического здоровья и задержек в развитии штата Иллинойс, Исследования и Развитие № 740-02. Я хотел бы поблагодарить Робина Хогарта за его комментарии по поводу более ранней версии этой книги.

1. Способность делать прогноз о том, в каких условиях будет использоваться отдельное правило, в настоящее время неадекватна (Wallsten, 1980). Однако интерес к тому, как и при каких условиях правило усваивается, должен увеличить способность человека делать прогноз того, когда оно, используется. Например, если правило усвоено в ситуациях, где дано мало времени на то, чтобы сделать выбор, прогнозирование использования такого правила увеличено знанием о давлении времени, используемом в задании.

2. Сопутствующее обстоятельство (1) – это то, что должно возможно влиять на то, как люди делают оценки и принимают решения, разрабатывая ситуации, в которых задачи включают или подобны начальным условиям изучения. Значение этого как для помощи, так и для управления людьми, огромно (Fischhoff, Slovic и Lichtenstein, 1978; 1980).

3. Изучение сосредотачивается на переменных окружающей среды и структуре задачи. Поэтому, переменные, такие как количество подкреплений, план подкрепления, количество испытаний (= количество проб), и т.д., должны рассматриваться применительно к пониманию способов оценивания и принятия решений (см. Estes, 1976). Хотя важность задачи для понимания поведения непрерывно подчеркивалась (Brunswik, 1943; Castellan, 1977; Cronbach, 1975; Dawes, 1975b; W. Edwards, 1971; Einhorn & Hogarth, 1978; Simon & Newell, 1971), психологи, по-видимому, так же склонны к тому, что Росс (Ross, 1977) называет фундаментальной ошибкой атрибуции, как и все люди (недооценка факторов окружающей среды при приписывании причин).

4. Главная переменная в понимании эвристик – обратная связь исхода. Так как обратная связь исхода – главный источник информации для оценки качества наших правил принятия решений/оценивания, знание того, как переменные задачи влияют как на исходы, так и на способ, которым результаты закодированы и хранятся в памяти, становится чрезвычайно важным в объяснении того, как эвристика изучается и используется.

5. Область изучения представляет собой фокусную точку для рассмотрения относительных достоинств психологических, а не практических объяснений поведения выбора. Некоторые экономисты спорили, что, хотя человек не действует “рационально” все время, он осваивает оптимальное правило посредством взаимодействия с окружающей средой. В поддержку этого аргумента предоставляются неопределенные утверждения относительно равновесия, эффективности и эволюционных концепций. Поэтому, исследование, как (и насколько хорошо) люди учатся из своего опыта, важно в освещении относительных достоинств психологических и практических теорий выбора.

Получение знаний из опыта: Как?

Очевидно, что принятие решения ориентировано на действие; каждый должен выбрать, какое действие совершить, чтобы удовлетворить основные

потребности и нужды. Поэтому, для любого организма важно изучить степень, в которой действия будут вести к желательным или нежелательным результатам. Это означает, что получение знаний из опыта, главным образом, должно включать в себя изучение связей действие-исход. Кроме того, так как действия и исходы следуют друг за другом, люди склонны видеть связи между ними как выражение причинно-следственных отношений (Michotte, 1963). Поэтому, сильная тенденция видеть причинные отношения может рассматриваться как следствие потребности принимать меры для удовлетворения основных потребностей. Кроме того, как показали Канеман и Тверски (Kahneman и Tversky, 1979b), изучение причинных отношений и организация событий в причинные “схемы” позволяет людям достигать связанной интерпретации их опыта. Наконец, изучение связей действие-исход важно для понимания того, как люди изучают собственные вкусы или потребности. Например, рассмотрим ребенка, который выбирает некий овощ, ест его, испытывает неприятный вкус, и, таким образом, учится связывать отрицательную полезность с этим продуктом. Обратите внимание, что именно путем выбора испытываются последствия и усваивается полезность. Поэтому, усвоение связей действие-исход и усвоение полезности тесно связано.

Хотя мы получаем знания из опыта, предпринимая действие, как человек первоначально учится выбору альтернатив? Несомненно, большая часть первоначального изучения происходит методом проб и ошибок; то есть люди беспорядочно делают выбор и наблюдают исходы (ср. Campbell, 1960). Процесс, с помощью которого обучение путем проб и ошибок ведет к развитию стратегий или правил, не известен (ср. Siegler, 1979). Однако можно предположить, что и подкрепления из эмпирического опыта, и обобщения (как по стимулам, так и по реакциям) играют важную роль (Staddon и Simmelhag, 1971). В любом случае, правила, которые мы разрабатываем, кажется, непосредственно связаны с получением знаний того, какие исходы будут вытекать из определенных действий. Как описано выше, получение знаний из опыта в основном индуктивно по характеру, то есть человек попадает в некоторые ситуации или случаи, и разрабатывает эвристику, чтобы обеспечить некоторый общий способ справиться с ними. Индуктивный характер получения знаний из опыта имеет несколько значений относительно эвристики:

1. *Специфика правил.* Если получение знаний происходит индуктивно с помощью определенных случаев, то эвристические правила должны сильно зависеть от контекста. Большое количество сведений говорит о том, что это действительно так (Grether и Plott, 1979, Lichtenstein и Slovic, 1971; Simon и Hayes, 1976; Tversky и Kahneman 1980). Способ, которым проблема сформулирована или показана, или ожидается определенная реакция, все это, кажется, имеет огромное значение для того, как обрабатывается информация и порождаются ответы. Яркий пример этой специфики можно обнаружить в работе Саймона и Хайеса (Simon & Hayes, 1976), посвященной “проблемным изоморфам” (problem isomorphs). Они показали, что различные

поверхностные формулировки идентичных по структуре проблем (то есть, проблем, которые могут быть решены, используя идентичные принципы) сильно изменяют то, как люди представляют проблему в памяти и, следовательно, решают ее. Важное значение этого результата заключается в следующем: чтобы сделать эвристические модели более прогнозирующими, человек должен столкнуться с задачей, как она была ему предоставлена, и не обязательно с ее структурой, как видит ее экспериментатор. Особенно ярким примером важности этого явления при прогнозировании поведения является наблюдение, что поведение зависит от того, представлено ли сокращение налогов как выгода или как меньший убыток (Kahneman и Tversky, 1979b).

2. *Общность правил.* Если эвристики – это правила, изученные с помощью индукции, необходимо сгруппировать задачи по подобию, иначе было бы столько правил, сколько и ситуаций. Так как последняя возможность недопустима, эвристика должна иметь некоторую общность по задачам. Однако это заключение противоречит тому, что было сказано выше относительно зависимости от контекста и специфичности правил. Этот парадокс может быть решен, если рассмотреть диапазон задач, к которым правило может применяться. Например, рассмотрим правило “Никогда не заказывайте рыбу в ресторане, специализирующемся на мясе.” В то время как такое правило является общим относительно некоторого типа ресторанов, оно, безусловно, более определено, чем правило “Оценивайте вероятность, с которой событие В происходит из процесса А их степенью подобия” (Tversky и Kahneman, 1974, 1). Очевидно, последняя эвристика имеет намного более высокий уровень общности. Фактически, эвристики, такие как репрезентативность, доступность, привязка, и регулировка – “метаэвристики,” то есть они – правила о том, как производить правила. Поэтому, когда человек сталкивается с задачами, с которыми он не встречался прежде (подобно оценке вероятности событий), или задачам, специфика которых заставляет их казаться новыми, метаэвристика определяет способ, с помощью которого определенные правила могут быть сформулированы для решения проблемы. Идея о метаэвристиках позволяет сохранить общность, которую любое правило обязательно подразумевает, и в то же время учитывает важные эффекты контекста, формулировки, способа ответа, и так далее. Чтобы проиллюстрировать это, рассмотрим исследование Словика, Фишхоффа и Лихтенштейна (Slovic, Fischhoff и Lichtenstein, 1976; см. также Главу 33), в котором людей просили оценить относительные вероятности смерти от необычных причин. Например, что более вероятно: быть убитым молнией или умереть от эмфиземы*? Когда люди сталкиваются с таким вопросом, существует много способов попытаться ответить. Одно из правил, которое могло бы использоваться, гласит: “Вспомнить знакомых мне людей, которые умерли от двух причин, и выбрать ту, из-за которой умерло больше людей”. В моем

* Эмфизема – заболевание легких (*прим. редактора*).

собственном случае, я выбрал бы эмфизему (что имеет более высокую вероятность, хотя большинство людей выбирает удар молнии). Однако, я мог бы также легко разработать правило, которое приведет к противоположному ответу; например, “Вспомнить о всех случаях смерти от удара молнии и от эмфиземы, о которых я когда-либо *слышал* (газеты, телевидение, и т.д.)”. Если бы это было моим правилом, я выбрал бы удар молнии как более вероятный. Обратите внимание, что в обоих случаях я использовал эвристику доступности. Ясно, что способ, которым вопрос был сформулирован, мог стимулировать определенные правила, что привело бы к различным результатам, и все же эти определенные правила могли бы быть систематизированы под отдельную, более общую стратегию или метаэвристику (также см., Einhorn, Kleinmuntz и Kleinmuntz, 1979).

3. *Сила эвристики.* Если эвристика усваивается индуктивно, то усвоение происходит в течение многих испытаний с многократным закреплением. Мы увидим, благодаря способу, которым осуществляется обратная связь, и благодаря методам, которые мы используем, для проверки правила опытом, положительное закрепление может происходить даже для неправильных правил (Wason, 1960). Кроме того, в дополнение к большому количеству закрепления, которое мы испытываем, размер или интенсивность закрепления может быть также большим. Например, получая значительную сумму денег после использования некоторого правила в выборе акций, мы должны испытывать значительный эффект закрепления. Поэтому, в отличие от лабораторных исследований процессов обучения, где этические соображения предотвращают получение большого положительного и отрицательного подкрепления, наш собственный опыт не ставит никаких подобных ограничений.

Получение знаний из опыта: Насколько хорошо?

Вопрос, насколько хорошо мы получаем знания из опыта, переключает внимание на сравнение эвристических правил с оптимальными правилами. Поэтому, возникает вопрос, как последние изучаются и каковы их значения для применения в нашем собственном опыте? Оптимальные правила, такие как теорема Байеса, оптимизация, и т.д., изучаются *дедуктивным путем*. Фактически, многое из того, что может называться формальным изучением, имеет дедуктивный характер, то есть нам преподают научные законы, логические принципы, математические и статистические правила, и т.д. Такие правила по своей природе очень абстрактны и не зависят от контекста. Кроме того, когда контекст может влиять на форму правила, нам часто говорят, что правило сохраняется, “при прочих равных условиях”. Конечно, в нашем собственном опыте “прочие условия” редко равны, что делает изучение оптимальных правил с помощью индукции столь трудным. (Первоначальные исследователи или изобретатели оптимальных правил преодолели эти трудности; однако, это отличает их от остальной части нас.)

Абстрактный характер дедуктивных правил имеет большое значение относительно трудностей, с которыми сталкиваются люди при применении оптимальных методов в определенных ситуациях. Эта трудность сосредотачивается вокруг способности различать структуру задач, встроенную в большое количество подробностей. Поэтому, когда перед нами стоит определенная проблема, богатая на детали, и в которой детали могут быть несоответствующими или избыточными, внимание к специфическим особенностям, вероятно, отвлечет внимание от общей структуры проблемы. Фактически, именно абстрактность дедуктивно изучаемых оптимальных правил не позволяет им восстанавливаться из памяти (см. Nisbett и др., 1976, см. гл. 7). Абстрактные правила, поэтому, могут быть “не доступны” в определенных случаях. Однако спорный вопрос не решен, так как важно знать, *почему* эти правила не доступны.

Рассмотрим способ, которым комбинации действие-исход, вероятно, организованы и сохранены в памяти. В частности, рассмотрим, будет ли такая информация организована и сохранена в памяти по содержанию или по структуре задачи. Кажется более легким и “естественным” организовать комбинации действие-исход скорее по теме, чем по структуре; например, опыт со школой, родителями, людьми противоположного пола и т.д., чем задачи Байеса, ситуации выбора, проблемы оптимизации и так далее. Факт, что содержание может отличаться, в то время как структура остается той же, весьма трудно увидеть (Einhorn и др., 1979; Kahneman & Tversky, 1979b; Simon & Hayes, 1976). Поэтому, я думаю, вряд ли большинство людей организует свои опыты по структуре задачи. Я не говорю, что человек не может научиться действовать таким образом. Фактически, большая часть профессионального обучения именно такова; например, человека обучают обращаться с проблемами как принадлежащими к некоему классу проблем, имеющих данную структуру и (иногда) известное решение. Оптимальные правила, таким образом, могут быть “доступны” с помощью обширного обучения. Конечно, существует опасность, что такие правила будут *чересчур* легко доступны; то есть проблемы вынуждены складываться в несоответствующую структуру, потому что решение существует в пределах этой структуры. Банально, что при столкновении с проблемой, профессионалы рассматривают ее в пределах структур, видеть которые они были обучены. Поэтому, хотя обучение профессионала действительно включает структуры, такое обучение находится в пределах узко определенной содержательной области.

Дальнейшие сведения, иллюстрирующие потребность в группировании проблем скорее по содержанию, чем по структуре, приводятся при рассмотрении способа, с помощью которого организовано и изучается знание людей о мире. Например, специализированное образование, профессиональное обучение, каталогизация информации в библиотеках и энциклопедиях, и так далее, иллюстрирует организацию информации скорее по содержанию, чем по структуре. В то время как существуют большие преимущества в организации знания таким образом, имеются также и недостатки. Трудность применения оптималь-

ных правил, разработанных в одной содержательной области, для подобных по структуре проблем в других содержательных областях может быть одним таким недостатком. Однако на уровне отдельного ученика рассматриваются другие трудности, которые могут иметь еще большие недостатки.

Хотя структуру задачи трудно различить, а исходы – легко, они четко видимы, доступны и часто однозначны. Рассмотрение закрепления с помощью обратной связи исхода необходимо в понимании того, как эвристика сохраняется, несмотря на опыт. Кроме того, если исходы – в значительной степени функция структуры задачи, и у человека, принимающего решения, недостаточно знания такой структуры, то правила, которые являются несоответствующими или даже недостаточными, могут все еще укрепляться положительной обратной связью исхода. (Например, “суеверное” поведение животных при обучении; см. Staddon и Simmelhag, 1971.)

Представлены два примера, где нормативно неподходящая эвристика может привести к хорошим исходам и где может отсутствовать понимание плохого качества правила. Рассмотрим человека, который в магазине подходит к банкам с соком со следующими ценами и качеством (взято из Tversky, 1969):

<u>Торговая марка</u>	<u>Цена</u>	<u>Качество</u>
X	60 центов	Высокое
Y	55 центов	Среднее
Z	50 центов	Низкое

Предположим, что я использую следующее правило, чтобы выбрать одну из трех марок: если различие в цене – пять центов или меньше, то я выбираю марку с более высоким качеством; если различие в цене больше, чем пять центов, я выбираю согласно цене. Такое простое правило (которое является лексикографическим полу-упорядочиванием) подразумевает: $X > Y$, $Y > Z$, но $Z > X$. Поэтому, это правило ведет к нетранзитивным выборам, которые являются иррациональными. Однако, обратите внимание, что после того, как я выбираю X, а не Y, то я устранию Y из оставшегося множества и сравниваю X с Z. Поэтому, я останавливаюсь на Z, который может оказаться весьма приемлемым после того, как я попробую сок. Затем я поздравляю сам себя с тем, что я такой хороший покупатель – я сэкономил деньги и сделал хорошую покупку. Важный пункт, на который стоит обратить внимание – то, что не делается сравнения Y с Z, и я не сознаю, что мое правило приводит к нетранзитивному выбору. Все, что я знаю – то, что я сделал выбор с минимальной суетой и напряжением, и исход был удовлетворителен. Положительная обратная связь исхода, таким образом, укрепляет нормативно неподходящее правило, и отсутствует понимание, что что-то неправильно.

Второй пример – вероятностный (ср. Schum, 1980). Вообразите, что Вы – военный генерал в политически напряженной области, и Вы обеспокоены тем, что Ваши враги вторгнутся в Вашу страну. Кроме того, из прошлого опыта Вам известно, что, когда вражеские отряды подходят к границе, вероятность

вторжения – 0.75. Однако у Вас нет прямого доступа к информации относительно вражеских отрядов, и Вы вынуждены полагаться на сообщение ваших разведчиков о таких действиях. Как оказывается, каждый раз ваши источники разведки сообщают, что отряды сосредотачиваются и что они – действительно там. Предположим, что Вы теперь получили сообщение из ваших источников, что вражеские отряды – на границе. Какова вероятность вторжения? Более формально, пусть

H = гипотеза, что будет вторжение

D = отряды, сосредотачивающиеся на границе

D^* = сообщение об отрядах, сосредотачивающихся на границе

В задаче дано, что $p(H|D) = 0.75$ и $p(D|D^*) = 1.0$ и спрашивается относительно $p(H|D^*)$. Если Вы подобны большинству людей, Вы, вероятно, ответите 0.75. Однако, данной информации не достаточно, чтобы ответить на этот вопрос нормативно правильным способом. Фактически, возможно, что в вышеупомянутой задаче $p(H|D^*) = 0$. Так как большинство людей считает, что в это трудно поверить, рассмотрим Рисунок 1, в котором эта задача проиллюстрирована посредством диаграммы Венна. Обратите внимание, что пересечение H с D^* пустое, так, что условная вероятность, $p(H|D^*)$, нулевая. Причина, по которой люди считают этот результат таким неожиданным, состоит в том, что они сделали логическую ошибку: если $D^* \Rightarrow D$, тогда $D \Rightarrow D^*$. Хотя D происходит всякий раз, когда дано D^* , наоборот происходит не всегда. Фактически, интуитивный способ видения проблемы состоит в том, чтобы думать, что враг настолько хитер, что ваша разведка видит их отряды только тогда, когда не планируется никакого вторжения. Однако когда вторжение запланировано, и отряды на границе, они скрыты настолько хорошо, что ваши источники не сообщают о них.

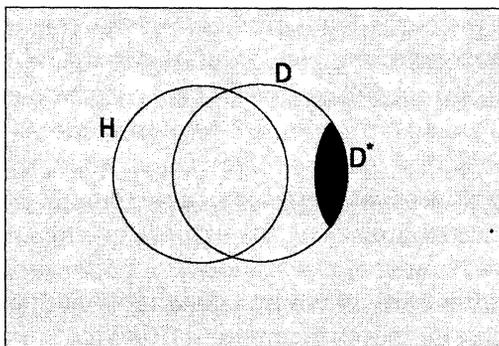


Рис. 1. Диаграмма Венна, показывающая отношения между гипотезой (H), данной величиной (D) и сообщением о данной величине (D^*)

Этот пример иллюстрирует трудность применения оптимальных правил (в этом случае, правил формальной логики) к определенной задаче. В то время как очень мало людей сделали бы логическую ошибку, когда она представлена в очевидной форме, важность примера – показать, как определенные особенности задачи скрывают ее реальную структуру так, чтобы оптимальные правила были легко нарушены (см. Tversky и Kahneman, 1980). В отношении этого примера также можно сказать о следующем.

Предположим, что генерал делает логическую ошибку и оценивает вероятность войны как 0.75. Тогда он посылает *свои* войска к границе, таким образом, вызывая вражеское вторжение. Следовательно, неправильное рассуждение генерала подкреплено обратной связью исхода: “В конце концов, – он мог бы сказать, – они напали на нас, как мы раньше и предполагали”.

Эти два примера иллюстрируют основной момент этой главы: без знания структуры задачи, обратная связь исхода может быть несоответствующей или даже вредной для исправления неправильной эвристики. Кроме того, положительная обратная связь исхода без знания задачи не дает понять, что наши правила являются неподходящими, так как нет желания задавать вопрос, как был достигнут успех. Условия, при которых обратная связь исхода не играет исправляющей роли в отношении эвристики и стратегий, обозначены как несоответствующие исходу структуры получения знаний (НИСПЗ). Такие структуры могут быть намного более распространены, чем мы думаем. Перед тем, как исследовать одну из таких структур подробно, рассмотрим вероятностные оценки в пределах структуры НИСПЗ, так как многие работы по эвристике непосредственно рассматривают эти типы оценок. Предположим, что Вы оцениваете вероятность некоторого события, как 0.70. Пусть событие не происходит. Что этот исход говорит вам о качестве правил, использованных для оценивания? Можно поспорить, что *любого единственного* исхода недостаточно, чтобы оценить “совершенство” (то есть, степень калибровки) вероятностных оценок. Поэтому, в важном смысле, непосредственная информация исхода не подходит для исправления неподходящей эвристики. Только в том случае, если Вы ведете “таблицу” относительной частоты исходов при оценке события с данной вероятностью, можно получить значительную обратную связь от исходов. Однако это является необходимым, но не достаточным условием для создания совершенных оценок. Во-первых, какой период времени человек рассматривает таблицу перед тем, как решить, является ли оценка откалиброванной или нет? Кроме того, насколько близким является “достаточно близко”, чтобы говорить, что оценка точна (в смысле хорошо калибрована)? Обратите внимание, что весь этот способ оценки исходов включает подкрепление, которое отсрочено на длительный период времени. Таким образом, ясно, что такая обратная связь не произведет значительный эффект само-исправления. Во-вторых, чтобы научиться определять совершенство правил для оценки вероятности, таблица должна включать не только оценки и конечные исходы, но также и

правила для получения этих оценок. Например, если я хранил записи исходов для 100 случаев, в которых я дал оценки 0.7, что даст мне информация о том, что 53 из этих событий произошли, относительно качества правил, которые я использовал? Так как вероятно, что много различных правил могло использоваться, чтобы оценить вероятности в 100 различных ситуациях, информация исхода нерелевантна, и обратная связь исхода не значительна, если только человек не знает правил и не хранит запись об их использовании (см. Nisbett и Wilson, 1977, о том, знаем ли мы наши собственные когнитивные процессы).

Я не подразумеваю, что невозможно научиться делать совершенные оценки вероятности. Если человек делает *много* оценок вероятности в *одной и той же ситуации*, как, например, синоптики или те, кто пытается предсказать исходы конных скачек, и обратная связь исхода быстро получена, такие условия не могут быть нерелевантными исходу, и обратная связь может иметь само-исправляющий характер. Однако такие условия были бы скорее исключением, чем правилом для большинства из нас.

Хотя вероятностные оценки типично происходят в НИСПЗ, как же насчет не вероятностных оценок? Конечно, если человек делает прогноз, он может проверить и увидеть, является ли прогноз правильным или нет. Поэтому, казалось бы, исходы должны быть релевантными для обеспечения само-исправляющейся обратной связи. Оставшаяся часть этой главы обсуждает эту проблему в пределах контекста одной общей и распространенной структуры задачи, хотя определенное содержание таких задач может быть весьма различно.

Задача выбора ¹

Рассмотрим общую задачу, вовлекающую не вероятностные оценки, так как информация исхода кажется как доступной, так и уместной в обеспечении само-исправляющейся обратной связи. Задача, которую нужно рассмотреть – та, в которой оценки сделаны с целью выбора между альтернативными действиями. Например, рассмотрим ситуацию с двумя возможными действиями, А и В. Обозначим за x целое, оценочное суждением которое может само быть функцией различных типов и количества информации. Кроме того, пусть x_c такой предел, что

Если $x \geq x_c$, совершаем действие А;

если $x < x_c$, совершаем действие В

Хоть оно и упрощено, уравнение 1 применяется ко многим ситуациям, а также ситуациям оценки/решения, например: трудоустройство, продвижение по службе, поступление в школу, предоставление кредита, назначение на коррекционные программы, участие в социальных программах, приня-

¹ Многие из этой главы взято из Айнхорн и Хогарт (Einhorn и Hogarth, 1978)



Рис. 2. Комбинации действие-исход, которые следуют из использования оценок для решения относительно принятия-отклонения

тие статьи в журнал, предоставление гранта, и т.д. В этих случаях оценка степени “достойности” обычно определяет, какое действие должно быть совершено, так как предпочтительное действие не может быть применено ко всем. Чтобы сравнить оценку со стандартом, существует критерий, обозначенный y , и служащий основанием для определения точности оценки. В то время как практические трудности обнаружения и развития адекватных критериев огромны, в данном случае внимания сосредотачивается на теоретических: концепция критерия – то, что необходимо для этого анализа. Чтобы быть совместимым с формулировкой оценки, далее допускается, что критерий имеет предел (y_c) такой, что $y \geq y_c$ и $y < y_c$ служат основанием для определения исходов оценки. Таким образом, поскольку мы рассматриваем учение об оценках, представление исходов в памяти имеет часто категориальную форму, то есть успехи и неудачи (см. Estes, 1976).

Очень важно обратить внимание, что структура задачи – та, в которой оценки (прогнозы) приводят к различным действиям и что исходы используются как обратная связь для определения точности прогнозов. Формальная структура может быть замечена при рассмотрении регрессии y от x , а также четырех секторов, которые следуют из пересечения x_c и y_c как показано на Рисунке 2. Обозначим правильные прогнозы как положительные и отрицательные попадания и два типа ошибок как ложные позитивы ($y < y_c | x \geq x_c$) и ложные негативы ($y \geq y_c | x < x_c$). Чтобы оценить отношения между x и y (то

есть, корреляцию между x и y , p_{xy}) необходимо иметь информацию относительно каждой комбинации оценка-исход. Предположим сначала, что такая информация становится доступной через какое-то время (то есть, последовательно), и рассмотрим экспериментальные сведения относительно получения знаний об отношениях между x и y при таких обстоятельствах. Исследование способности оценивать случайности между x и y , исходя из информации в таблице 2×2 (Jenkins и Ward, 1965; Smedslund, 1963; 1966; Ward и Jenkins, 1965) указывает, что люди судят о силе связей с помощью частоты положительных попаданий (в терминологии рис. 2), игнорируя информацию в трех других ячейках. Эти результаты чрезвычайно важны, так как они говорят, что *даже если* вся соответствующая информация исхода доступна, люди не используют ее. Это означает, что в лабораторных исследованиях, которые имеют релевантные относительно исхода структуры изучения, люди преобразовали их в нерелевантные для исхода структуры изучения. Чем это может объясняться?

Приведем объяснение: наш опыт в задачах реальной жизни является таким, что мы развиваем правила и методы, которые, как кажется, вполне хорошо “работают”. Однако эти правила могут быть весьма неправильными, и наше понимание их несоответствия глубоко. Этот недостаток понимания существует, потому что положительная обратная связь исхода может происходить, скорее, несмотря, чем благодаря нашей прогнозирующей способности. Чтобы проиллюстрировать это, рассмотрим исследование Васона (Wason, 1960), в котором он предоставлял испытуемым последовательность с тремя числами, например: 2, 4, 6. Испытуемых просили обнаружить правило, которому эти три числа подчинялись (правило, расположения по возрастанию). Чтобы обнаружить правило, им разрешали создавать наборы трех чисел, которые экспериментатор классифицировал как подходящие или не подходящие под это правило. В любой момент испытуемые могли остановиться, когда они думали, что они обнаружили правило. Правильное решение этой задачи должно включать поиск опровергающего свидетельства, а не накопление подтверждающих. Например, если кто-то предполагал, что правило имеет некоторое отношение к четным числам, это могло быть проверено только, перебирая последовательность, включающую нечетные числа (то есть, накапливание большого количества подтверждающих примеров последовательностей четных не будет вести к правилу). То, что только 6 из 29 испытуемых нашли верное правило с первого раза, иллюстрируют опасность индукции простым перечислением. Как указывает Васон (Wason, 1960), решение этой задачи должно включать “готовность пытаться фальсифицировать гипотезы, и таким образом, *проверять те интуитивные идеи, которые так часто придают чувство уверенности*” (с. 139, нами добавлено выделение курсивом).

Важно подчеркнуть, что в эксперименте Васона, где не было действий, возможен поиск опровергающего свидетельства. Однако когда действия основаны на оценке, изучение, основанное на опровергающем примере, ста-

новится более трудным. Рассмотрим, как можно было ошибочно усвоить неверное правило для оценивания, и рассмотрим гипотетическую ситуацию с менеджером, изучающим свою прогнозирующую способность относительно “потенциала” кандидатов на работу. Важный фактор здесь – то, что действия (например, принимать / не принимать) – зависят от оценок. В следующий раз менеджер может только экзаменовать *принятых* кандидатов, чтобы проверить, сколько из них “успешны”. Если успехов много, что вероятно, тогда все эти случаи подтверждают правило. Действительно, важный момент – то, что было бы трудно опровергнуть правило, даже при том, что оно могло быть ошибочным. Одним из способов, которым менеджер может проверить правило, будет принять несколько человек, которые, как он оценивал, имели низкий потенциал и затем понаблюдать за их успехами. Если бы их успехи были столь же высоки, как и у кандидатов с оцененным высоким потенциалом, правило было бы опровергнуто. Однако систематический поиск опровергающих свидетельств редок, и против него можно было бы возражать на прагматических и даже этических основаниях, то есть человек был бы должен отказать в предпочтительном действии тем, кого он оценил как более достойных, и передать его тем, которые были оценены как наименее достойные. Поэтому, прагматические и/или этические соображения могут препятствовать даже мыслям о возможности сбора опровергающей информации. Обратите внимание, что тенденция не проверять гипотезы контрпримерами – прямое следствие структуры задачи, в которой действия приняты на основе оценок. Васон (Wason, 1960) указывает, “В реальной жизни нет силы, позволяющей делать оценку на основании выводов: выводы могут только быть проверены данными” (с. 139). В результате, много положительной обратной связи может привести к укреплению невалидного правила.

Хотя влияние исходов на несовершенное действие может не рассматриваться, возможно, что человек будет исследовать количество положительных попаданий и ложных положительных как способ проверить точность прогнозов. Поэтому, в то время как такая информация неполна для точной оценки отношений между прогнозами и исходами, такая информация – то, что доступно большинству людей. Поэтому важно рассмотреть факторы, которые затрагивают эти переменные.

Факторы, воздействующие на положительные попадания и ложные положительные

Рассмотрим Рисунок 2 снова и обратим внимание, что существуют три фактора, которые затрагивают пропорции положительных попаданий и ложных позитивов; местоположение x_c , y_c , и “наклон” эллипса (который является корреляцией между x и y). Например, если x_c смещен направо, а значения x_c и r_{xy} постоянны, имеется точка, в которой не будет никаких ложных позитивов. Конечно, будет соответствующее увеличение ложных негативов.

Однако если у человека нет информации относительно этих случаев (как ситуации вообще), опыт успеха может убеждать, что качество оценивания высоко. Поэтому, когда критерий для предоставления предпочтительного действия поднят (увеличивая x_c), вероятность того, что $p(x \geq x_c)$ (также называемое отношением выбора, φ), уменьшится, и это приведет к высокому положительному попаданию и низким пропорциям ложных позитивов. Другой фактор, y_c , очевидно влияет на исходы, так как уровень y_c определяет успех и неудачу. Обратите внимание, что когда значение y_c понижается, вероятность, что $p(y \geq y_c)$ (также называемая базовым значением, br), увеличивается, и опыт успехов может быть высок независимо от способности к оцениванию; то есть если бы вы *наугад* назначили людей для выполнения различных действий, вы бы получили пропорцию успеха, равную $p(y \geq y_c)$. Поэтому, чтобы оценить способность к прогнозированию, должно быть проведено сравнение пропорции положительного попадания с $p(y \geq y_c)$, и способность к оценке определена при незначительном увеличении в успехе. Третий фактор, p_{xy} , прямо пропорционален исходам; то есть, чем больше p_{xy} , тем больше процентное отношение положительных попаданий.

Влияние этих трех факторов на процентное отношение положительных попаданий хорошо известно. Тейлор и Расселл (Taylor & Russell, 1939), например, показали, что можно повысить процентное отношение положительных попаданий, для любых данных p_{xy} и базового значения, путем уменьшения отношения выбора (φ), то есть, предоставляя предпочтительное действие меньшему проценту людей (допустим, $p_{xy} \neq 0$). Таким образом, даже если p_{xy} низкое, возможно получить высокое процентное отношение положительных попаданий, зависящее от величин φ и br . Тейлор и Расселл (1939) приводят таблицы процентных отношений положительных попаданий для большого разнообразия значений p_{xy} , φ и br . Исследование этих таблиц показывает, что низкая корреляция между оценками и критериями вполне совместима с большим процентным отношением положительных попаданий.

В дополнение к трем уже упомянутым факторам необходимо рассмотреть и четвертый фактор. Он может быть проиллюстрирован с помощью следующего воображаемого эксперимента. Пусть о некоторых людях сделано некоторое множество оценок. Половину из тех, которые были оценены выше x_c , *наугад* назначаем выполнять действие А, а вторую половину – действие В. Подобным образом поступаем с теми, чья оценка не превысила x_c . Через некоторое время измеряем поведение и подсчитываем пропорцию людей с $y \geq y_c$ в каждой клетке (каждому человеку приписано 1 или 0, чтобы обозначить, ниже он или выше предела y – пропорция, большая y_c в таком случае просто среднее клетки). Это факторный план 2×2 , причем один фактор – это “оценка”, а другой “тип действия”. Заметьте, что так как критерий не может быть измерен непосредственно перед решением (и действительно, если это было бы так, то в оценке не было бы необходимости), люди, получающие действия А и В, также получили различное экспериментальное воздействие.

Если бы этот эксперимент был проведен, можно было бы протестировать основное влияние оценки (что измеряет его точность); основное влияние на действие, то есть, влияет ли само получение А или В на поведение; и взаимодействие между оценкой и действием. Обратите внимание, что преимущество эксперимента состоит в том, что он позволяет нам узнать точность оценки из влияния отношения действия. Как бы там ни было, такой эксперимент редко проводится, даже умозрительно, и, конечно же, он не проводится людьми, не прошедшими углубленное обучение планированию экспериментов. Таким образом, точность оценок всегда будет связана с возможным влиянием отношения, обусловленного действием. Более того, ссылаясь на более раннее обсуждение, можно сказать, что этот эксперимент позволяет проверить опровергающую информацию. В отличие от большинства реальных оценочных задач, он позволил бы опровергнуть гипотезу точности оценки, *также как и* оценить любое влияние отношения, обусловленного действием.

Пример влияния отношения показан на Рисунке 3. Эллипс, изображенный пунктиром, тот же, что и был на Рисунке 2, и он показывает "истинное" взаимоотношение между оценками и исходами. Заштрихованная область обозначает наблюдаемые исходы; таким образом, показаны только значения, для которых $x \geq x_c$. Влияние отношения проявляется в том, что исходы (т.е. поведение) всех людей, которым было дано действие А, увеличены на некоторую константную величину, так что количество положительных попаданий больше, чем если бы она была при отсутствии эффекта отношения. С психологической точки зрения главный аспект на Рис. 3 – это то, характер обратной связи искажен, количество положительных попаданий резко увеличилось, а количество ложных позитивов резко сократилось.



Рис. 3 Влияние отношения на пропорцию наблюдаемых положительных попаданий

Чтобы измерить влияние четырех вышеупомянутых факторов на пропорцию положительных попаданий, Айнхорн и Хогарт (Einhorn & Hogarth, 1978) смоделировали эксперимент, в котором различные уровни воздействующих факторов отношения выбора, базового значения и способности к прогнозированию изменялись по факторному плану. Зависимой переменной было процентное отношение положительных попаданий. Результаты этого эксперимента могут быть представлены следующим образом: (а) в общем, пропорция положительных попаданий больше, чем 0.50. Если существует влияние отношения, то процентное отношение положительных попаданий может быть высоко даже при $p_{xy} = 0$; (б) когда $\varphi < br$, процентное отношение положительных особенно высоко. Кроме того, пропорция положительных попаданий зависит от влияния отношения при низких значениях p_{xy} . Это означает, при ситуациях выбора плохая способность к прогнозированию, закрепится положительной обратной связью исхода. (с) При $\varphi > br$, отношение положительных попаданий – наименьшее. Однако, малые воздействия факторов имеют существенное влияние на увеличение процентного отношения положительных попаданий в этих ситуациях.

Результаты смоделированного эксперимента показывают, что положительная обратная связь может существовать, если способность к прогнозированию крайне мала, и что осознание этого обычно мало из-за неудачи соответственно понять структуру задачи. Поэтому, хотя можно было предположить, что не вероятностные оценки сделаны в уместной исходу структуре изучения, когда оценки делаются с целью решения, какое действие предпринять, информация исхода может быть нерелевантной для обеспечения само-исправляющейся обратной связи.

Заключение ²

Основной темой этой главы было, что информация исхода, без знания структуры задачи, может быть несоответствующей для обеспечения само-исправляющейся обратной связи относительно неподходящей эвристики. Также обсуждалось, что знание структуры задачи трудно достичь из-за индуктивного способа, которым мы получаем знания из опыта (см. Hammond, 1978, Галилейевский способ мышления против Аристотелевского). Эти выводы поднимают две проблемы, которые будут кратко обсуждены.

Может быть, что даже со знанием структуры задачи, человек действует таким способом, который мешает получению знаний. Например, рассмотрим официанта в шумном ресторане. Поскольку у него нет времени, чтобы обеспечить хорошее обслуживание всем клиентам, он делает прогноз, относительно того, какие клиенты оставят хорошие или плохие чаевые. Хоро-

² Я бы хотел поблагодарить J. E. R. Staddon за поднятие проблем, обсужденных в этой главе.

шее или плохое обслуживание распределяется в зависимости от прогноза. Если качество обслуживания оказывает влияние на размер чаевых, исходы “подтверждают” первоначальные прогнозы. Обратите внимание, что официант мог бы провести эксперимент, чтобы не допускать зависимости качества обслуживания от своих предсказаний, если бы он знал структуру задачи; то есть он мог бы предоставлять плохое обслуживание некоторым людям, которые, как он судил, оставят хорошие чаевые, и хорошее обслуживание тем людям, которые, как он оценил, оставят плохие чаевые. Однако обратите внимание, что официант должен захотеть рискнуть возможной потерей дохода, если его суждение точно, взамен на знание, что его оценка является неправильной. Последняя информация может принести ему отдаленную выгоду, так как она может побудить его делать прогнозы лучше или, если это невозможно, использовать стратегию предоставления хорошего или плохого обслуживания наугад, таким образом, экономя много умственных усилий. В принятии решений в организациях, отдаленные выгоды от знания точности прогнозов могли быть существенны. Например, если отборочные собеседования не прогнозируют выполнение работы (независимо от влияния отношения), зачем тратить деньги и время, используя их? Поэтому, затраты и выгоды от краткосрочных стратегий для действия по сравнению с отдаленными стратегиями получения знаний должны быть более тщательно исследованы.

Вторая проблема возникает при рассмотрении следующего вопроса: изучают ли люди, и продолжают ли они использовать несоответствующие правила, и не противоречит ли это эволюционной концепции выживания сильнейших? Под этим вопросом я подразумеваю, что те, кто используют неадекватные правила, менее вероятно, должны выжить, чем те, кто используют лучшие правила (они - сильнее). Однако, использование лучших правил все еще весьма далеко от использования оптимальных правил. Концепция наибольшего “соответствия” включает относительное упорядочение, в то время как оптимальность подразумевает некоторый абсолютный уровень. Следовательно, тот факт, что условно оптимальные правила сохраняются, несмотря на опыт, не противоречит теории Дарвина. Возможно, наиболее краткий способ выразить это процитировать Эразма: “В стране слепых, и одноглазый человек – король.”³

³ Смысл этой цитаты состоит в том, чтобы указать, что *относительные* преимущества в отношении окружающей среды важны. Мы не оскорбляем слепых людей. Том Воллстейн сделал следующий комментарий, “В стране слепых, одноглазый человек может выжить, только закрыв свой глаз, так как окружающая среда будет устроена так, что он будет полагаться на другие чувства”. Хотя это - очаровательный комментарий, я не соглашаюсь с ним, потому что одноглазый человек имел бы другие чувства в дополнение к преимуществу зрения.

Часть VI
Чрезмерная уверенность

20. Чрезмерная уверенность в суждениях, основанных на конкретных примерах*

Стюарт Оскамп

Это обычное явление в клинической практике: когда психолог накапливает исследовательский материал относительно человека, он начинает думать, что хорошо знает его. Следовательно, рано или поздно в процессе сбора информации психолог чувствует себя достаточно уверенным, чтобы делать диагностические заключения, описывать главную динамику клиента, и возможно даже, чтобы отважиться сделать прогноз его будущего поведения. Хотя заключения психолога могут остаться предварительными, усиление уверенности от времени первого столкновения с пациентом ко времени написания своего сообщения обычно очень заметно.

Это исследование рассматривает, оправдано ли это увеличение уверенности соответствующим увеличением точности заключений. Хотя уверенность психолога в его заключениях часто упоминалась как важный предмет научного исследования (Meehl, 1957), она редко интенсивно исследовалась. Кроме того, когда она исследовалась, часто получались довольно удивительные результаты. Например, Л.Р. Голберг и Оскамп (L. R. Goldberg, 1959 и Oskamp, 1962) показали, что диагностическая достоверность опытных психологов *меньше*, чем у менее опытных людей. Те же самые исследования и многие другие показали, что профессиональные психологи являются не лучшими экспертами в межличностных отношениях, а иногда и худшими, чем необученные люди (Taft, 1955).

Другим редко исследуемым фактором, который может обеспечить хороший показатель опытности эксперта, является отношение между уровнем его достоверности и уровнем точности. Эта величина показывает, например, является ли судья самонадеянным при принятии решений или, наоборот,

* Эта глава первоначально появилась в Журнале консультационной психологии (The Journal of Consulting Psychology), 1965, 29, 261-265. Авторское право © 1965 принадлежит Американской Психологической Ассоциации. Переиздано в соответствии с разрешением.

он не слишком уверен в себе. Относительно этой величины, которая может быть названа правомерностью достоверности, было обнаружено, что опытные “оценщики”, превосходят неопытных (Oskamp, 1962).

Многие исследования (Hamlin, 1954; Hathaway, 1956; Kostlan, 1954; Soskin, 1954; Winch & More, 1956) рассматривали влияние различного количества информации стимула на клиническое принятие решений. В настоящем эксперименте этот фактор изучался путем предоставления каждому испытуемому четырех множеств кумулятивно увеличивающегося количества информации, как основания для принятия им решений. Таким образом, было смоделировано постепенное наращивание информации при работе психолога над типичным случаем.

Исследовательские гипотезы были следующие:

1. После некоторого момента в процессе сбора информации, точность прогнозирования достигает максимума.
2. Однако, уверенность в решениях продолжает устойчиво расти, поскольку получено большее количество информации.
3. Таким образом, к концу процесса сбора информации, большинство испытуемых самонадеянны относительно своих оценок.

Процедура

Так как было желательно, смоделировать обычную клиническую ситуацию настолько точно, насколько возможно, в качестве информации, которая предложена испытуемым, было выбрано реальное исследование. В конечном итоге, случай был выбран из-за его обширности, из-за того, что он охватывал многие подходящие жизненные случаи, и потому, что он описывал относительно нормального человека (то есть случай неадаптированности молодого человека, который никогда не помещался в психиатрическую лечебницу). Это был случай Джозефа Кидда (Joseph Kidd), о котором сообщил Уайт (White, 1952) в своей книге “Жизни в развитии” (Lives in Progress).¹

Материалы относительно этого случая были организованы в хронологические наборы информации, которые были представлены испытуемых в четыре последовательных этапа. Этап 1 содержал только следующую краткую демографическую информацию относительно этого случая, чтобы проверить спрогнозированную точность на уровень “психологической вероятности” (Patterson, 1955):

Джозефу Кидду (псевдоним) – 29 лет. Он белый, не состоит в браке, является ветераном Второй Мировой Войны. Он – выпускник колледжа, работает бизнес-ассистентом в студии цветочной декорации.

На этапе 2 было добавлено полторы страницы материала, напечатанных через один интервал, относительно детства Кидда до 12 лет. Этап 3 (2 страницы) рассмат-

¹ Использование этого случая имело один недостаток. Несколько испытуемых помнили, что они читали этот материал во время их обучения, но все кроме одного, сообщили, что их более раннее знакомство никак не помогло им в этом исследовании. Так как их точность, подтвердила это впечатление, их результаты были сохранены в анализе данных.

ривал его годы, проведенные в школе и колледже, а этап 4 (1¹/₃ страницы) охватывали его службу в армии и более поздние действия до возраста в 29 лет.

Исследование на конкретном примере

Чтобы иметь основание для определения точности испытуемых, было построено испытание множественного выбора, используя метод, подобный методу Соскина (Soskin, 1954). Пункты опросника описывали обычные образцы поведения Кидда, отношения, интересы и типичные реакции на события жизни. Примеры некоторых этих пунктов даются в табл. 1.

Таблица 1. *Примеры пунктов опросника*

5. Будучи учащимся колледжа, когда Кидд был в знакомой и благоприятной социальной ситуации, он часто:

- a. Пытался направить группу и навязать ей свои желания.
- b. Оставался в стороне, отдельно от группы.
- c. Его не волновало то, как люди реагировали на него.
- d. Принимал активное участие в группе, но тихим и скромным способом.
- e. Вел себя, как клоун и хвастался^a

10. Позже во время службы в армии, будучи офицером и командующим отделением, отношение Кидда к наказаниям было:

- a. Он очень тревожился об этом, потому что предпочитал быть на одном уровне с другими людьми, а не выше их^a
- b. Он не любил этого, потому что он никогда не мог принять решение относительно того, что делать.
- c. Он избегал этого настолько полно, насколько возможно, потому что чувствовал, что неправильно наказывать людей, независимо от того, что они сделали.
- d. Он был счастлив, потому что это давало ему шанс контролировать ситуацию и быть выше других.
- e. Он испытывал садистское наслаждение, наказывая других, в качестве компенсации за то время, когда наказывали его.

15. Настоящее отношение Кидда к своей матери:

- a. Любовь и уважение к ее идеалам.
- b. Нежная терпимость к ее недостаткам.
- c. Как уважение, так и негодование^a.
- d. Отвержение ее и всех ее убеждений.
- e. Почтительная, но формальная привязанность.

20. В беседах с мужчинами, Кидд:

- a. Предпочитает располагать их говорить о работе и опыте^a.
- b. Предпочитает говорить, главным образом, о знакомых вещах.
- c. Предпочитает спорить с ними о религии или их философии жизни.

d. Любит хвастать относительно своей службы в армии или учебы в колледже.

f. Говорит главным образом о спорте, сексе и отпускает грязные шутки.

25. Отношение Кидда к жизни как бизнес-ассистента вызвано его недавним решением:

a. Остаться на этой работе, по крайней мере, еще несколько лет.

b. Расширить бизнес, построив другой магазин в близлежащем городе.

c. Оставить работу и открыть собственный цветочный магазин.

d. Подать заявления о приеме на работу в несколько более крупных компаний в сфере деятельности, подобной его настоящей работы.

e. Совсем уволиться и найти работу другого рода ^a

^a Правильный ответ.

Варианты ответов были построены только там, где имелись довольно объективные информационные критерии, представленные в случае, фактические данные или хорошо зафиксированные выводы. Четыре неправильных альтернативы в каждом пункте были составлены аспирантами по психологии посредством дополнения предложений к основе пункта. Они были построены таким образом, чтобы быть явно неправильными, основанными на опубликованном материале случая, но быть убедительными и “привлекательными” альтернативами. Ни один из пунктов не содержал ответ, который можно было обнаружить в полученном вводном материале случая. Испытуемые должны были следовать обычной процедуре в клиническом принятии решений (С. McArthur, 1954), формируя картину индивидуальности Кидда из представленного материала и затем прогнозируя его отношения и типичные действия, основываясь на картине его личности.

Эксперты

Эксперты были взяты из трех групп с разным опытом в психологии: (а) 8 психологов, работающих в государственной больнице штата Калифорния, причем у каждого было несколько лет опыта терапии, и 5 имели кандидатскую степень; ²(b) 18 аспирантов в области психологии;³ и (с) 6 студентов стар-

²Еще один психолог участвовал в тестировании, но его результаты пришлось аннулировать, так как он не понял инструкций и не смог им следовать. Эта проблема не случилась ни с одним студентом.

³ Приблизительно половина этих аспирантов имела некоторый опыт терапии или консультирования, и один или два из них, возможно, по уровню опыта в психологии равнялись психологам

ших курсов, занимающихся исследованием личности. Ни один из испытуемых не был знаком с гипотезами исследования.

Испытуемые принимали участие в эксперименте в маленьких группах размерами от четырех до девяти человек, но каждый работал своим собственным индивидуальным темпом и с собственной пачкой материалов. После чтения материалов каждого этапа случая, испытуемый отвечал на все 25 вопросов случая, прежде чем перейти к следующему этапу. В дополнение к ответу на вопросы, испытуемый также указывал в каждом пункте, насколько он был уверен в том, что его ответ был правильный.

Оценки уверенности

Оценки уверенности были сделаны, используя шкалу, изобретенную Адамсом (Adams, 1957), которая определяет уверенность в терминах ожидаемого процента правильных решений. Так как было пять альтернатив для каждого пункта теста, шкала начиналась с 20% (что представляло полностью случайный уровень достоверности) и завершалась 100% (что указывало на абсолютную уверенность в правильности). В дополнение к обеспечению понятного объективного значения уверенности, эта шкала имеет большое преимущество, т.к. позволяет напрямую сравнить уровень точности и уровень уверенности. Таким образом, например, если у испытуемого 28% пунктов правильные, а средний уровень уверенности 43%, он, безусловно, самоуверен.

Результаты

Эта задача оценки, оказалась достаточно трудной, по крайней мере, для этого материала о случае. Ни один испытуемый не достиг точности 50%, и средняя результирующая точность была меньше 28%, со случайностью 20% (незначительное различие). Однако этот низкий уровень точности служит для того, чтобы обеспечить еще более серьезное испытание гипотез исследования.

Был выполнен предварительный анализ, чтобы сравнить оценки трех групп испытуемых, хотя не было сформулировано гипотез об их относительных результатах. Эти результаты ясно указали, что не было никаких существенных различий между тремя группами испытуемых ни в точности, ни в уверенности, ни в общем количестве измененных ответов. Полученные результаты уверенности на этапе 4 сравнивались с результатами предыдущих исследований (Goldberg, 1959; Oskamp, 1962), показывая, что более опытные испытуемые, *менее* уверены, чем менее опытные испытуемые, но в этом исследовании эти результаты не приближались к значимым.

Таблица 2. Результаты 32 испытуемых в исследовании конкретного случая по 25 пунктам.

Измерение	Средние оценки				F	p
	Этап 1	Этап 2	Этап 3	Этап 4		
Точность (%)	26,0	23,0	28,4	27,8	5,02	0,01
Уверенность (%)	33,2	29,2	46,0	52,8	36,06	0,001
Количество измененных ответов	-	13,2	11,4	8,1	21,56	0,001

Главные результаты исследования показаны в Таблице 2, где следующие друг за другом столбцы показывают средние результаты испытуемых, по мере того, как они получали последовательно увеличивающееся количество информации. В результате предыдущих статистических тестов, результаты для всех 32 испытуемых объединены в этой таблице.

Первая строка в Таблице 2 показывает, что колебание в точности по четырем этапам было существенно. Однако, статистический тест Дункана (A. E. Edwards, 1960, с. 136) показал, что это разница возникла прежде всего благодаря снижению точности на этапе 2. При сравнении точности на этапе 1 с точностью на этапе 4 не было обнаружено никаких существенных изменений ($t = 1.13$, $df = 31$). Таким образом, первая гипотеза относительно максимума точности была не только поддержана, но в этом эксперименте не было никакого существенного увеличения в точности с увеличением информации!

Гипотеза 2 проверена во второй строке табл. 2. Там мы видим, как и было спрогнозировано, резкое и очень существенное увеличение достоверности с 33% на Этапе 1 до 53% на этапе 4.

Наконец, результаты гипотезы 3 обозначены сравнением первых и вторых строк таблицы. На этапе 1 средняя величина самоуверенности была 7 пунктов; на этапе 4 она была 25 пунктов, различие, существенное далеко за пределами уровня 0.001 ($t = 5.14$, $df = 31$).

Иногда средние величины группы существенны, но они могут вводить в заблуждение, потому что они могут скрывать ответы отдельных испытуемых, которые противоречат прогнозам. То, что дело здесь было не в этом, ясно показано следующими числами для отдельных испытуемых. Из 32 испытуемых, 14 увеличили точность ответов от этапа 1 к этапу 4, в то время как для 6 точность осталась той же, и для 12 уменьшилась - полностью случайный результат. Наоборот, все испытуемые, кроме 2, увеличили уверен-

ность, и большинство — заметно⁴. На этапе 1 почти половина испытуемых (13 из 32) не были самоуверенными; к этапу 4 осталось только 2 испытуемых, которые были не очень уверены в себе - высоко значимое изменение ($\chi^2 = 9.1$, $p < 0.01$).

Другой интересный результат исследования содержится в последней строке Таблицы 2, в которой показано среднее число пунктов, на которых испытуемые изменили свои ответы на каждом этапе случая. Эта величина показывает, что, поскольку представлялось большее количество информации, число измененных ответов *уменьшилось* заметно и существенно. Это заключение говорит о том, что испытуемые часто формировали довольно твердые стереотипные выводы из первого отрезка информации и затем неохотно меняли свои выводы, когда они получали новую информацию. Во всяком случае, заключительный этап информации, служит скорее для того, чтобы подтвердить предыдущие впечатления испытуемых, чем побудить их обновить полную картину личности Кидда.

Обсуждение

Необходимо избегать небрежного обобщения этих результатов. Имеются три главных фактора относительно этого исследования, которые могли бы ограничить общность результатов, (а) случай может быть не похож на те, с которыми большинство психологов привыкло работать. (б) Пункты теста не могут представлять те виды поведения, которые психологи привыкли прогнозировать. (с) Испытуемые, возможно, не были типичными представителями людей, принимающих психологические решения. В ответе на эти возможные возражения должно быть указано, что случай, пункты теста и испытуемые-психологи – все были выбраны с намерением приблизиться к ситуациям, встречающимся в фактической психологической практике.

Даже если эти возможные возражения могут быть предоставлены, некоторые четкие заключения могут быть сделаны. Независимо от того, казалась ли задача странной или материалы случая нетипичными, оценки уверенности испытуемых показывают, что они *убедились в увеличении своего понимания случая*. Поскольку они получили большее количество информации, их уверенность резко возросла. Более того, их уверенность относительно собственных решений стала обратно пропорциональна фактической правильности этих решений.

Таким образом, хотя этот результат не может повторяться для каждого психолога и каждого типа решений, можно четко заключить, что усили-

⁴Один из двух испытуемых, у которых уменьшилась уверенность, старшекурсник, позже заявил, что обычно у него увеличилась бы уверенность, но он только что участвовал в научно-исследовательской работе с компьютерами, в которой компьютер неоднократно давал неправильные результаты, причем настолько, что он полностью потерял свою уверенность даже в компьютерах.

вающееся чувство уверенности психолога во время прорабатывания ситуации - *не* надежный признак увеличения точности его заключений. Так называемая клиническая надежность, основанная на личных чувствах уверенности клинициста, не является адекватным свидетельством валидности оценки в диагностировании или прогнозировании человеческого поведения.

21. Сообщение о процессе обучения оцениванию вероятности*

Марк Алперт и Ховард Раиффа

В исследованиях принятия решений в условиях неопределенности, людей, принимающих решения и их квалифицированных консультантов часто просят оценить распределения вероятности величин, значения которых им неизвестны. Эта глава рассматривает некоторые эмпирические результаты, касающиеся таких вопросов как: Насколько хорошо могут необученные люди давать подобные оценки? Проявляют ли они некоторые повторяющиеся предубеждения? Как можно калибровать людей, делающих оценки? Как их можно научить оценивать лучше?

Эта глава рассматривает только оценки неопределенных величин, которые могут принимать континуум возможных значений. Следовательно, мы будем работать исключительно с одновариантными функциями плотности и их кумулятивными функциями распределения. Существует несколько различных процедур для оценки распределений вероятности непрерывных, одновариантных случайных переменных, но мы рассмотрим только одну специфическую процедуру, которую мы и наши коллеги часто использовали на практике. Она называется методом прямой оценки квантиля.

Процедура прямой оценки квантиля

Пусть x^* – истинное, целевое значение некоторой переменной, и предположим, что оно неизвестно оценивающему. K -тый оценочный квантиль x^* (для k в интервале от 0 до 1) – это число x_k , которое означает, что оценочная вероятность того, что он припишет событию $\{x^* \leq x_k\}$ равна k ; в терминах вероятности, $P\{x^* \leq x_k\} = k$. На числа $x_{0.50}$, $x_{0.25}$ и $x_{0.75}$ будут ссылаться как на оценочную медиану, нижний квартиль и верхний квартиль, соответственно.

* Эта глава первоначально появилась в 1969 как неопубликованное сообщение. Она была пересмотрена специально для этой книги.

Чтобы получить $x_{0.50}$, испытуемый должен представить такую величину, которая настолько же вероятна, что x^* как ниже $x_{0.50}$, так и выше $x_{0.50}$. Таким образом, $x_{0.50}$ делит континуум на два оценочных, одинаково вероятных интервала. Более низкий квартиль, $x_{0.25}$, делит интервал $(-\infty, x_{0.50})$ на два оценочных одинаково вероятных интервала; и верхний квартиль, $x_{0.75}$, делит интервал $(x_{0.50}, +\infty)$ на два оценочных интервала с одинаковой вероятностью. Для каждой неопределенной величины, наших участников эксперимента попросили определить их оценочную медиану и квартили. Последовательность (или “когерентность”, как предпочитают писать некоторые авторы) требует, чтобы испытуемый считал, (а) что каждый из этих четырех интервалов

$$(-\infty, x_{0.25}), (x_{0.25}, x_{0.50}), (x_{0.50}, x_{0.75}), (x_{0.75}, +\infty)$$

с равной вероятностью, будет содержать истинное значение x^* , и (б) вероятность того, что истинное значение x^* не будет содержаться в интервале $(x_{0.25}, x_{0.75})$. В будущем, мы будем называть интервал $(x_{0.25}, x_{0.75})$ оценочным, межквартильным диапазоном. Испытуемых, которые участвовали в наших упражнениях, обучили проверять эти требования последовательности и их проинструктировали, что в случаях непоследовательности необходимо заново оценить свои оценки квантиля, чтобы ее достичь.

В дополнение к оценочной медиане и двум квартилям, испытуемых просили оценить различные квантили на нижнем и верхнем пределе их распределений. Об этом мы расскажем дальше. Как только оценивающий определил несколько (x_k, k) пунктов в своем совокупном, левостороннем оценочном распределении вероятности, он может использовать прямой, произвольный процесс для того чтобы “подогнать” остаток кривой. В этой главе, однако, мы будем касаться только непосредственно оцененных (x_k, k) пунктов, а не полной кривой.

Возможность внешней валидации

Если все, что мы имеем от данного испытуемого, это одно распределение вероятности для единственной неопределенной величины, было бы бессмысленно говорить, что его распределение “неправильно”. Мы могли бы надеяться, что наш испытуемый более хорошо осведомлен о рассматриваемой величине, но его распределение вероятности только формальное выражение того, что он знает или не знает об этой величине. Мы не можем сказать, например, что его распределение “слишком компактное”, или “слишком свободное”, или “слишком смещено вправо”. Но, в отличие от этого случая, допустим, что наш испытуемый дает нам тысячу распределений тысячи различных неопределенных величин. Если каждое из фактических истинных значений должно попадать либо ниже его соответствующего 0.01 квантиля, либо выше его 0.99 квантиля, то мы сможем сказать, что он внешне не

ограничен, что его распределения имеют тенденцию быть слишком компактными. Или, напротив, если бы было так, что каждый из его оцененных межквартильных диапазонов содержал бы истинную величину, то это обозначало бы его тенденцию быть слишком разреженным. Допустим, что это экстремальные случаи, но они устанавливают то, что возможно и уместно говорить о внешней валидизации множества распределений вероятности.

Цели упражнений и состав групп испытуемых

Не вдаваясь в подробности, позвольте нам предоставить краткий обзор того, что мы сделали. Посредством ряда анкет мы опросили большое количество испытуемых (приблизительно тысячу) и записали их оценочные квантили для нескольких величин, неизвестных им во время оценки. Затем мы сравнили их оценки с фактическими истинными значениями; идентифицировали некоторые постоянные предубеждения; исследовали формальные процедуры для изменения их оценочных входных данных; проинформировали каждого испытуемого о качестве ответов группы и его собственного специфического набора ответов; и, наконец, мы предложили способы, которыми каждый испытуемый мог получить информацию о прошлых индивидуальных особенностях и, таким образом, изменить свой последующий набор оценок.

Мы провели четыре отдельных, но связанных упражнения в течение академического года 1968-1969, и для идентификации, перечислим их:

Группа 1: 139 студентов факультативного курса “Модели планирования в условиях неопределенности”, второй курс программы на степень магистра бизнес-управления (МБУ) Гарвардского университета.

Группа 2: 800 студентов, включающая всех первокурсников программы МБУ Гарварда.

Группа 3: 67 “добровольцев” с углубленного курса менеджмента на бизнес-факультете Гарварда.

Группа 4: 60 студентов, проходящих курс анализа принятия решений, который осуществлялся в аспирантуре точных и гуманитарных наук в Гарварде. Это были студенты различных факультетов в Гарварде и Технологическом институте штата Массачусетс и изучали экономику, статистику, инженерию, математику, юриспруденцию, управление, дизайн, общественные отношения, операционный анализ.

До принятия участия в этом исследовании, все испытуемые в этих четырех группах изучали основные принципы анализа принятия решений: построение и анализ схем принятия решений, предшествующие и последующие распределения вероятности, теорию полезности (или предпочтений) и величину информации.

На группе 1 проводилось пилотажное исследование. От каждого студента требовалось независимо оценить распределения для 20 неопределенных величин. Для этой группы мы изменяли наши инструкции о вероятности “хвоста” распределения; это обсуждается позже в этой главе. Все 20 оценок были проведены за одно испытание без обратной связи.

С группами 2, 3 и 4 было проведено нечто подобное. В первом испытании каждому испытуемому предлагали оценить 10 неопределенных величин. Затем испытуемых информировали относительно результатов всей группы и их собственных результатов. Затем осуществлялось следующее испытание с 10 различными неопределенными величинами.

Действия группы 1 и групп 2, 3 и 4 в первых испытаниях были в значительной мере сходны. Действия групп 2, 3 и 4 во втором испытании, после информирования и обратной связи, были снова чрезвычайно подобны; но заметные улучшения были отмечены между их первыми и вторыми испытаниями. Из-за подобия действий всех групп мы обсудим и сделаем замечания по детальным результатам самой большой из этих групп.

Упражнение и результаты, полученные до обратной связи

Мы воспроизводим инструкции, которые были предоставлены половине учащихся курса МБУ (группа 2). Эта подгруппа получила то, что мы назвали формой В; другая половина получила форму А, которая была идентична по содержанию, но в которой были приведены другие неопределенные величины.

Инструкции: форма В

Цель этого упражнения состоит в том, чтобы увидеть, насколько хорошо Вы как представитель группы и группа в целом можете оценивать распределения вероятности для неопределенных величин (НВ). Мы приведем ниже 10 НВ, и Вас попросят оценить медиану, 0.25 квантиль, 0.75 квантиль и значения экстремума каждой НВ. (Для вашего удобства краткие определения этих терминов даются ниже.) Благодаря типу используемой НВ, Вы сможете сравнить Ваши оценки с истинными значениями. Таким образом, Вы сможете увидеть, склонны ли Вы делать распределение “слишком плотным”, “слишком разреженным” или смещенным вверх или вниз в определенных типах вопросов. Позже Вас попросят повторить это упражнение с другими НВ. Для этого упражнения Вы должны независимо ответить на вопросы, не просматривая исходные материалы, хотя для некоторых вопросов желательно произвести простые вычисления. На интересует Ваше знание независимо от того, насколько оно неопределенно или насколько неудобно Вы себя чувствуете.

Определения квантилей ...

Ниже приведено 10 величин, которые, как мы надеемся, для Вас являются неопределенными. Вы должны записать Вашу оценку на каждом бланке ответа. Заполните каждый пункт настолько хорошо, насколько Вы можете с нынешними знаниями, но не ищите дальнейшую информацию и не обсуждайте вопросы о соответствии со своими партнерами даже после того, как Вы сдадите свои ответы. Вам необходимо сдать один бланк ответов перед занятиями завтра, а второй оставьте

себе так, чтобы у Вас была запись Ваших ответов. (Студенты использовали сохраненный экземпляр, чтобы оценить себя. Это будет описано позже.)

Значения некоторых из переменных будут определены с помощью ответов студентов-первокурсников на следующие вопросы:

Вопросы

А. Вы предпочитаете бурбон или скотч?

В. Вы одобряете отсрочки от армии для всех аспирантов независимо от области деятельности?

С. Приняли бы Вы участие в азартной игре "50 на 50", где Вы могли бы либо проиграть \$ 50 или выиграть \$ 100?

Список неопределенных величин

1. Процент ответов первокурсников (исключая тех, которые вообще не пьют), предпочитающих бурбон скотчу.

2. Процент ответов первокурсников, которые одобряют отсрочку от армии для всех аспирантов независимо от области деятельности.

3. Процент ответов первокурсников, которые сыграли бы в азартную игру в вопросе С, приведенном выше.

4. Процент респондентов, которые принимали участие в Опросе общественного мнения института Гэллапа в июле 1968 (рассматривающем репрезентативную выборку взрослых американцев) и выразили мнение, что если начнется полномасштабная война на Ближнем Востоке, США должны послать вооруженные формирования для помощи Израилю.

5. Процент респондентов, участвовавших в опросе общественного мнения в марте 1968 (рассматривающем репрезентативную выборку взрослых американцев), которые ответили, что учителям государственных школ нужно разрешить вступать в профсоюзы.

6. Количество терапевтов и хирургов, перечисленных в телефонном справочнике "Желтые страницы" Бостона и его окрестностей в 1968.

7. Общее количество студентов, которые в настоящее время зарегистрировались в докторской программе бизнес-факультета Гарварда.

8. Полное производство яиц (в млн штук) в США в 1965.

9. Количество иностранных автомобилей, импортированных в США в 1967, выраженное в тысячах.

10. Пошлина, собранная на Панамском канале в 1967 году, выраженная в миллионах долларов.

(Новая страница)

Форма В (бланк, который нужно сохранить)

Секция _____

Номер студента _____

Пожалуйста, выберите один ответ для каждого из вопросов А, В и С:

А. Напиток бурбон __ (1) Скотч __ (2) Не пью вообще __ (3)

В. Отсрочка от армии одобряю __ (1), выступаю против __ (2)

С. Азартная игра 50-50 принимаю __ (1), отклоняю __ (2)

Пожалуйста, оцените все 5 величин для каждой из этих 10 переменных, приведенных ниже. Заметьте, что самое высокое число в каждой строке находится справа. (Десятичные дроби в ответах допускаются. Если необходимо указать процент, ответы должны иметь форму 97.2 для 97.2%, а не 0.972.)

Неопределенная величина

Квантиль

1. Бурбон (%)..... (0.01) (0.25) (0.50) (0.75) (0.99) .

Информационное сообщение об обратной связи

Приблизительно через неделю после того, как студенты заполнили форму В, студентам был предложена следующее информационное сообщение. Оно обсуждалось в группе до того, как студентам дали для заполнения форму А.

Цель этого информационного сообщения:

1. Описать в совокупности поведение людей, заполнивших форму В,
2. Указать систематические предубеждения в ответах,
3. Позволить Вам проверить себя прежде, чем Вы повторите это упражнение в форме А.

Таблица 1. Ответы на вопросы в форме В

<i>Номер</i>	<i>Название</i>	<i>Истинное значение</i>
1	Бурбон	42.5
2	Отсрочка от армии	65.5
3	Азартная игра "50 на 50"	55.2
4	Израиль	10.4
5	Профсоюз учителей	63.5
6	Список докторов	2.600
7	Докторанты	235
8	Произведенные яйца (миллионы)	64.558
9	Импортированные автомобили (тысячи)	697
10	Доход с Панамского канала (миллионы)	82.3

Как вы помните, для каждого вопроса, Вы оценили пять квантилей: 0.01, 0.25, 0.50, 0.75, 0.99. Эти квантили делят интервал линии на шесть категорий:

- Категория 1: Все числа ниже квантиля 0.01
 2: Все числа между 0.01 и 0.25 квантилями
 3: Все числа между 0.25 и 0.50 квантилями
 4: Все числа между 0.50 и 0.75 квантилями
 5: Все числа между 0.75 и 0.99 квантилями
 6: Все числа выше 0.99 квантиля.

Для любого из этих десяти вопросов, как только Вы определяете Ваши пять квантилей (или, что то же самое, как только Вы определяете эти шесть категорий) следует указать, в какую категорию попадает фактическая истинная величина. Например, студент Джон Доу определил следующие квантили для вопроса о скотче и бурбоне:

Оценка:	(0.01)	(0.25)	(0.50)	(0.75)	(0.99)
	5	15	30	40	45

Фактический процент студентов, одобряющих бурбон, как оказалось, был 42.5; следовательно, в этом вопросе ответ г-на Доу попадает в категорию 5.

Теперь прежде, чем мы дальше будем обсуждать, как хорошо (или плохо) группа выполнила это упражнение, пожалуйста, обратитесь к *вашему* бланку ответов и для каждого вопроса отметьте, какая из ваших шести категорий содержит истинную величину.

В Таблице 1 мы перечислили истинные ответы.

Анализ межквартильных диапазонов

Давайте сначала рассмотрим колонки 3 и 4 в Таблице 2. Для любого вопроса Вы должны были выбрать 0.25, 0.50, 0.75 квантили так, что, по *вашему* мнению, настолько же вероятно, что истинный ответ попадет в диапазон от 0.25 до 0.75 (то есть, в категории 3 и 4), как если бы он в него не попал. В общем количестве 1.000 вопросов (100 x 10) мы получили бы (математическое) ожидание 500 ответов в категориях 3 и 4. Мы получили только 334 таких ответов. Неплохо. Это несоответствие могло быть статистическим отклонением, но мы в этом сомневаемся. Например, подобный анкетный опрос нескольких недель назад был предоставлен многочисленной группе второкурсников МБУ – они ответили на 20 вместо 10 вопросов – и 33% их ответов попали в категории 3 и 4. Мы не хотим сказать, что 33% “универсальная полуконстанта”, но это довольно поразительное совпадение, которое стоит учитывать. (Соответствующий процент для первого заполнения формы А был 33%; студенты, занимающиеся углубленным изучением менеджмента (группа 3) также выдали число 33%, но студенты факультета точных и гуманитарных наук (группа 4) предоставили ответ 36%.)

В совокупности, интервалы от 0.25 квантиля до 0.75 были слишком плотными. Столько истинных значений должно было попасть за пределы диапазонов интерквартилей (от 0.25 до 0.75 диапазона), сколько и попасть в них, но фактически *вдвое* больше попало за его пределы, чем внутрь. Но нам этого еще не достаточно, чтобы говорить, “расширьте диапазон интерквартилей”, потому что существует большая изменчивость в зависимости от вопроса и человека.

Сравним вопросы 1 и 10. В вопросе 1 Вы могли почувствовать, что много знаете о том, какие напитки предпочитают Ваши одноклассники, по крайней мере, весьма много по сравнению с тем, что Вы знаете относительно доходов от Панамского канала. Поэтому, Вы могли бы почувствовать, что у Вас есть больший шанс (ожидаемый) захватить в вашем интерквартильном

Таблица 2. Распределение ответов испытуемых в форме В по категориям.

Номер	Название	Ниже 0.01	0.01 до 0.25	0.25 до 0.50	0.50 до 0.75	0.75 до 0.99	Выше 0.99	всего
1	Бурбон	3	16	20	40	11	10	100
2	Отсрочка от армии	15	12	35	19	10	9	100
3	Азартная игра 50-50	11	8	28	29	13	11	100
4	Израиль	51	41	6	1	1	0	100
5	Профсоюз учителей	1	1	13	28	29	28	100
6	Список докторов	24	14	12	13	10	27	100
7	Докторанты	1	3	11	9	15	61	100
8	Производство яиц	9	2	13	10	8	58	100
9	Импортирован- ные машины	25	15	18	9	7	26	100
10	Доход с канала	18	8	8	12	16	38	100
Всего		158	120	164	170	120	268	1.000
Ожидае- мая частота		10	240	250	250	240	10	1.000

диапазоне истинную пропорцию бурбона, чем истинные доходы от Панамского канала. *Но разве так вы должны думать?* Если Вы это ожидали, Вы должны были увеличить Ваш межквартильный диапазон для доходов от Панамского канала. При заполнении формы А убедитесь, что Вы чувствуете, независимо от вашей информированности, что:

а. Для любого вопроса одинаково вероятно, что истинное значение попадет как в межквартильные пределы, так и вне их;

б. Одинаково вероятно, что истинное значение попадет в Ваш межквартильный диапазон для вопроса, о котором Вы знаете много и для вопроса, о котором Вы знаете мало.

(Этот последний пункт относительно вопросов 1 и 10 мы необдуманно обсудили со студентами факультета точных и гуманитарных наук скорее до, чем после того, как они закончили свой первый круг оценок. Это могло бы объяснить несоответствие между 33% и 36%. Мы увидим намного более значительное изменение, когда посмотрим на вероятности размещения “хвоста” распределения.)

Некоторые из Вас действительно производят слишком плотные распределения. Например, 13 из 100 респондентов охватили только 1 или ни одно-

го истинного значения в своих 10 межквартильных диапазонах. Для каждого человека, пусть его *межквартильное множество* будет числом попаданий истинных значений в его межквартильные диапазоны. Множество, близкое к 0, подразумевает, что человек “слишком напряжен”; множество близкое к 10 предполагает, что человек “слишком свободен”; а множество около 5 подразумевает, что он – “почти прав”. Распределение множества приводится в табл.3. В третьей колонке таблицы мы перечислили ожидаемое количество респондентов, которые бы предоставили любое данное межквартильное множество при упрощающем предположении, что существует вероятность 0.33, что любой межквартильный диапазон содержит истинное значение, и что эти дихотомические ответы являются независимыми для разных вопросов. Убедитесь, что у вас есть *свое* межквартильное множество.

Таблица 3. *Распределение межквартильного множества для формы В*

<i>Межквартильное множество</i>	<i>Фактическое количество человек</i>	<i>Ожидаемое количество человек, использующих $p = 0.33$</i>
0	3	1.9
1	10	9.0
2	22	19.9
3	20	26.1
4	23	22.5
5	11	13.3
6	9	5.5
7	2	1.5
8	0	0.3
9	0	0.0
10	0	0.0
Всего	100	100.0

Анализ экстремумов

Возможно, Вы устанавливаете Ваши квантили 0.01 и 0.99 так, чтобы Вы были “удивлены” действительно низким ответом (ответ категории 1) и действительно высоким ответом (ответ категории 6). В 1.000 вопросах мы должны были “ожидать” в общем 20 таких сюрпризов, 10 слева и 10 справа (см. табл. 2). *Было в общем 426 сюрпризов!* Это должно не просто удивить нас, а шокировать! Вы можете немного успокоиться, хотя бы потому, что Вы не столь плохи как группа второкурсников (группа 1), о которой мы упомина-

ли. Приблизительно 35 студентов попросили, как и Вас, перечислить квантили 0.01 и 0.99 для 20 вопросов. Они записали 46% сюрпризов – немного хуже, чем ваши 42,6%. Других 35 студентов попросили записать 0.001 и 0.999 квантили и вместо “ожидаемых” 0,2% сюрпризов они записали 40%. Других 35 попросили в довольно неопределенных терминах, перечислить “минимальные” и “максимальные” значения – мы действительно не знаем, что это означает – и они сделали запись 47% сюрпризов. Наконец других 35 попросили дать “удивительно низкие” и “удивительно высокие” значения, и они отметили 38% сюрпризов. Очевидно, “удивительно низко”, ниже, чем “минимум”.

Ради Бога, *Расширьте свои экстремумные квантили!* Будьте честными с самим собой! Признайте то, что Вы не знаете!

Рассмотрим 6-й вопрос 2.600 докторов в области медицины перечислены в телефонном справочнике Бостона и его окрестностей. Половина из Вас была удивлена этим результатом. Некоторые были удивлены тем, что это такая большая цифра, другие, что – маленькая. Пятьдесят восемь процентов были удивлены фантастически *большим* количеством произведенных яиц – но все-таки 9% были одинаково удивлены фантастически *маленьким* количеством.

Таблица 4. *Распределение индексов неожиданности для формы В*

<i>Индекс неожиданности</i>	<i>Фактическое количество человек</i>	<i>Ожидаемое число человек, использующих $p = 0.43$</i>
0	1	0.4
1	9	2.7
2	15	9.3
3	13	18.7
4	17	24.6
5	15	22.3
6	16	14.0
7	4	6.0
8	9	1.7
9	1	0.3
10	0	0.0
Всего	100	100.0

Давайте определим для каждого человека *индекс неожиданности*, который показывает количество раз (из 10), когда он делает запись ответа кате-

гории 1 или 6. В среднем по поводу многих различных, независимых вопросов Вы должны быть удивлены 1 раз из 50. Индексы неожиданности 0 или 1 – то, что мы хотели бы видеть, и появляется причина для беспокойства, если индекс неожиданности достигает 3 или больше. Распределение индексов неожиданности дается в табл. 4. В третьей колонке таблицы мы перечислили ожидаемое количество испытуемых, которые бы зарегистрировали данные индексы неожиданности при упрощенном предположении, что существует вероятность удивления 0.43 в каждом вопросе, и вопросы друг на друга не влияют. Сорок пять (45) человек были удивлены по крайней мере в половине из этих десяти вопросов!

Для типичной неопределенной величины, мы предположим, что большинство из Вас хочет оценить унимодальную кривую плотности (или, что то же самое, совокупное распределение с левым хвостом в форме буквы S), где самая высокая точка на кривой плотности (или самая крутая часть совокупной S-кривой) находится где-то в интервале ваших квантилей от 0.25 до 0.75. Многие из Вас, как мы неосторожно полагаем, сделали запись бимодальных кривых. Это случилось бы, если длина вашего интервала категории 2 меньше, чем ваш интервал категории 3, а длина вашего интервала категории 5 – меньше, чем длина интервала категории 4¹. Давайте рассмотрим следующий набор оценок квантиля:

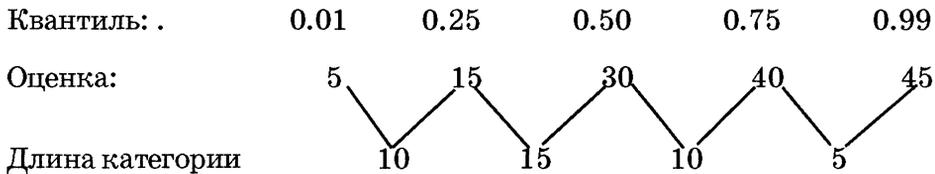


Таблица 5. Сравнение результатов в испытаниях 1 и 2

	форма А		форма В	
	Испытание 1	Испытание 2	Испытание 1	Испытание 2
	Значения, попадающие внутри межквартильных диапазонов (%)	32.9	40.3	33.4
Значения, не попадающие в диапазоны от 0.01 до 0.99 (%)	38.8	24.9	42.6	22.2

¹ Пусть l_1 определяет длину категории 1. Мы имели бы бимодальность, если бы l_2 ; и l_5 , каждый был бы меньше, чем $(l_3 + l_4) / 2$.

Таблица 6. Сравнения результатов испытаний 1 и 2

	Группа 3		Группа 4	
	Испытание 1	Испытание 2	Испытание 1	Испытание 2
Значения, попадающие внутри межквартильных диапазонов (%)	33.4	44.8	35.6	42.5
Значения, не попадающие в диапазоны от 0.01 до 0.99 (%)	35.6	22.8	20.8	8.7

В вышеупомянутой иллюстрации отношение длин вероятностного интервала 0.98 (от квантилей 0.01 до 0.99) к вероятностному интервалу 0.50 (от квантилей 0.25 до 0.75) равняется $(45-5) / (40-15)$ или 1.6. Это слишком маленькое число для унимодальной кривой. Например, для нормальной параболы это отношение – приблизительно 3.5. Отношения от 2.25 до 4.5 могут служить как схематическая директива, но директивы должны нарушаться время от времени. Помните: расширяйте свои распределения!

Результаты, полученные после обратной связи

После обсуждения в группе докладной записки об обратной связи, мы поменяли формы и повторили упражнение. После анализа ответов, мы раздали следующее информационное сообщение:

Информационное сообщение № 2

Каждый из Вас уже принимал участие в следующем:

а. (испытание 1): Вам давали одну из двух форм (А или В) и Вы представили квантиль оценки для десяти неопределенных величин;

б. (обратная связь): Вам дали истинные значения этих десяти квартилей и попросили записать два фактора: (1) сколько из Ваших ответов попали в ваши интерквартильные диапазоны, и (2) сколько Ваших ответов *не* попали в ваши диапазоны 0.01 и 0.99. Вы были также информированы об общем выполнении работы группой: Приблизительно вдвое больше ответов не попали в интерквартильные диапазоны, чем попали, и было действительно шокирующее число сюрпризов – около 40% ответов попало вне диапазонов 0.01 и 0.99.

с. (Испытание 2): Вам дали другую форму и попросили представить оценки квантилям.

Мы вычислили Ваши ответы для части (с) и готовы прокомментировать Вашу совокупную работу. Но, прежде всего, мы перечислим истинные значения неопределенных величин так, чтобы Вы могли проверить себя (таблица, представляющая эту информацию, была показана в докладной записке).

Таблица 7. Пропорция неожиданностей для различных интерпретаций “низко” и “высоко”

<i>Интерпретация “низко”</i>	<i>Интерпретация “высоко”</i>	<i>Размер выборки</i>	<i>Пропорция неожиданностей</i>
0.01 квантилями	0.99 квантилями	44	0.46
0.001 квантилями	0.999 квантилями	25	0.40
Минимальное значение	Максимальное значение	35	0.47
Удивительно низко	Удивительно высоко	35	0.38

Насколько хорошо группа справилась с этим заданием в целом? Помогла ли обратная связь? Она помогла, отчасти, но не настолько, насколько мы надеялись. Большинство ваших распределений слишком плотно. В Таблице 5 приведены результаты.

Если мы объединяем формы А и В, то процент попадания истинных величин в интерквартильные диапазоны, возрос от 33% до 43%. Намечено направление изменения, но нам все же хотелось бы получить 50%. Процент непопадания истинных значений в экстремумы значения (то есть, 0.01 и 0.99 диапазонов) снизился от шокирующего 41% к угнетающим 23%. Это – далеко от нашей цели 2%. Что-то необходимо сделать с “хвостами” распределения! Мы нуждаемся в другой сессии обратной связи и третьем круге практики; но, к сожалению, это не выполнимо. Мы предлагаем начать эксперименты с другими группами, чтобы получить более реалистические оценки вероятностей “хвоста” распределения.

(С целью сравнения в табл. 6 показаны действия группы студентов, углубленно изучающих менеджмент (группа 3) до и после выполнения, а также поведение студентов факультета точных и гуманитарных наук (группа 4)

Еще несколько слов об экстремумных “хвостах”

В экспериментальном исследовании (группа 1) мы изменили инструкции относительно экстремумных “хвостов”, чтобы увидеть, какое влияние это будет иметь на статистические распределения ответов. Всех наших субъектов, однако, попросили дать медиану оценки и два квартиля. Наши результаты собраны в табл. 7 и 8, которые мы объясним по очереди.

Из 139 испытуемых, сорок четырех попросили дать 0.01 и 0.99 квантили, и эта подгруппа зарегистрировала 46% неожиданностей (а не 2%); 25 попросили дать 0.001 и 0.999 квантили, и эта подгруппа зарегистрировала 40% неожиданностей (а не 0.2%). Тридцать пять испытуемых попросили дать “минимальное значение” и “максимальное значение”, и когда некото-

рые испытуемые попросили разъяснить, что эти термины означают, мы уклончиво заметили, что “конечно, это неопределенные термины, но попытайтесь ответить на вопросы так или иначе”. Мы дали тот же самый неоднозначный совет оставшимся 35 испытуемым, которых спросили относительно “удивительно низких” и “удивительно высоких” значений. Последние две группы зарегистрировали 47% и 38% соответственно. Различные инструкции имели определенное влияние, но, как Вы видите, не очень сильное.

Таблица 8. Распределение “высоких” оценок для вопроса 10

<i>Очки для победителя</i>	<i>Интерпретация "высоких оценок"</i>			
	<i>0.99</i>	<i>0.999</i>	<i>Максимум</i>	<i>Удивительно высоко</i>
< 29	8	1	6	2
29-35	15	6	7	6
36-42	10	5	7	11
43-49	2	6	5	4
50-56	7	3	8	7
57-63	2	0	2	4
64-70	0	4	0	1
Всего	44	25	35	35

Другим способом исследования влияния этих четырех инструкций на экстремумы является подробное изучение Таблице 8, которая дает распределение ответов на “высокие” значения для следующей величины: количество очков победившей команды в следующем футбольном матче Гарвард - Дартмаут. (Опрос проводили за неделю до игры.) Например, из 25 человек, которых попросили дать 0.999 квантиль, 1 сделал запись счета меньше, чем 29, 6 человек зарегистрировали счет между 29 и 35, 5 человек – между 36 и 42, и так далее. Кажется, как будто неопределенный термин “максимум” интерпретируется как 0.99 квантиль, а неопределенный термин “удивительно высоко” интерпретируется как 0.999 квантиль.

Мы не повторяли в точности эти инструкции относительно экстремумных значений для групп 2, 3 и 4; а, мы последовательно спросили этих испытуемых о 0.01 и 0.99 квантилях.

22. Калибровка вероятностей: положение дел к 1980г.*

*Сара Лихтенштейн, Барух Фишхофф и
Лоренс Д. Филлипс*

С субъективистской точки зрения (de Finetti, 1937/1964), вероятность – степень веры в утверждение. Она выражает только внутреннее состояние; нет никакой “правильной”, “верной” или “объективной” вероятности, существующей где-нибудь “в реальности”, с которой степень веры может быть сравнена. Во многих обстоятельствах, однако, возможно проверить правильность или ошибочность предположения, к которому вероятность была применена. Сегодня, человек оценивает вероятность предположения “завтра будет дождь”. Завтра, смотрит на прибор для измерения осадков, чтобы увидеть действительно ли шел дождь. Если есть возможность, такая проверка может использоваться для определения адекватности оценок вероятности.

Уинклер и Мерфи (Winkler и Murphy, 1968b) идентифицировали два вида “качественности” оценок вероятности: нормативная качественность, которое отражает степень, в которой оценки выражают истинные убеждения экспертов и соответствуют аксиомам теории вероятности, и действительная качественность, которое отражает знания объема темы, представленное в оценках. В этой главе рассматривается литература о другом аспекте качественности, называемой калибровкой.

Если человек оценивает вероятность, того что предположение истинно на 0.7, а позже обнаруживает, что предположение ложно, это само по себе не лишает оценку валидности. Однако если оценивающий приписывает вероятность 0.7 на 10.000 независимых событий, только 25 из которых впоследствии оказываются истинными, тогда что-то не так с этими оценками. То, чего им не хватает — калибровки (проверки); другими словами реализма (Brown и Shuford, 1973), внешней валидности, (Brown и Shuford, 1973),

* Это – пересмотренная версия статьи, которая первоначально появилась под редакцией Н. Jungermann и G. DeZeeuw. Принятие решения и изменения занятий людей. Dordrecht-Holland: D.Reidel Publishing, Co., 1977. Переиздано в соответствии с разрешением.

* Калибровка – (тех.) это процесс или результат настройки технического устройства для повышения точности его работы.

реализма уверенности (Adams и Adams, 1961), уместности уверенности (Oskamp, 1962), вторичной валидности (Murphy и Winkler, 1971) и надежности (Murphy, 1973). Формально, оценивающий “проверен”, если, по многим испытаниям, для всех предположений с заданной вероятностью, оцененная вероятность равна заданной. Калибровка экспертов может быть определена опытным путем, наблюдая их оценки вероятности, проверяя взаимосвязанные предположения, а затем наблюдая пропорцию, истинную в каждой категории ответа.

Экспериментальная литература по калибровке людей, экспертов по вероятности дискретных предположений, рассмотрена в первом разделе этой главы. Второй раздел описывает калибровку функций плотности вероятности, проведенную для неопределенных числовых величин. Хотя калибровка – по существу свойство человека, большинство исследований, рассмотренных здесь, представляет данные, сгруппированные по экспертам для обеспечения большего количества данных, необходимых для устойчивых оценок калибровки.

Дискретные предположения

Дискретные предположения могут быть охарактеризованы согласно количеству альтернатив, которые они предлагают:

Безальтернативные: “Что такое абсент?” Эксперт дает ответ, и затем оценивает вероятность того, что данный ответ является правильным. Возможен полный диапазон ответов вероятности, от 0 до 1.

Одна альтернатива: “Абсент – это драгоценный камень. Какова вероятность, что это утверждение является истинным?” Опять же, возможен диапазон шкалы вероятности – от 0 до 1.

Две альтернативы: “Абсент – (а) драгоценный камень; (б) спиртное”. Используя метод *неполного диапазона*, оценивающий сначала выбирает более вероятную альтернативу, а затем заявляет вероятность (≥ 0.5) того, что этот выбор является правильным. Используя метод *полного диапазона*, оценивающий определяет вероятность (от 0 до 1), что данная альтернатива правильна.

Три или больше альтернативы: “Абсент – (а) драгоценный камень; (б) спиртной напиток; (с) карибский остров; (d) ...” Могут использоваться две разновидности этой задачи: (1) оценивающий выбирает единственную наиболее вероятную альтернативу и заявляет вероятность, что она является правильной, используя ответ $1/k$ для k альтернатив или (2) оценивающий назначает вероятности для всех альтернатив, используя диапазон от 0 до 1.

Для всех этих разновидностей, калибровка может быть изображена посредством *калибровочной кривой*. Такая кривая получена следующим образом:

1. Собирается множество оценок вероятности для пунктов, правильный ответ по которым известен или будет вскоре известен экспериментатору.

2. Группируются сходные оценки, обычно в пределах диапазонов (например, все оценки между 0.60 и 0.69 помещены в одну категорию).

3. В пределах каждой категории, вычисляется доля правильных ответов, (то есть, доля пунктов, для которых предположение истинно, или альтернатива правильна).

4. Для каждой категории, среднего ответа (по абсциссе) с долей правильных ответов (по ординате).

Проведенную калибровку можно показать всеми точками, попадающими на линию идентификации.

Для задач с половиной диапазона, плохо калиброванные оценки могут быть либо *самонадеянными*, за счет чего исправленные пропорции – меньше, чем оцененные вероятности, и калибровочная кривая попадает ниже линии идентификации, либо *неуверенные*, за счет чего исправленные пропорции – больше, чем оцененные вероятности, и кривая калибровки попадает выше линии идентификации.

Для задач с полным диапазоном без альтернатив или одной альтернативой, самонадеянность имеет два возможных значения. Эксперты могли быть самонадеянными в истинности ответа; такая самонадеянность была бы обозначена кривой калибровки, попадающей всегда ниже линии идентификации. С другой стороны, эксперты могут быть самонадеянными в их способности отличить истинные предположения от ложных. Такая самонадеянность была бы показана кривой калибровки ниже линии идентификации в области более чем 0.5 и выше линии идентификации в области ниже 0.5.

Было предложено несколько числовых измерений калибровки. Мерфи (Murphy, 1973) исследовал общий случай пунктов с k альтернативами, начиная с оценки Байера (Brier, 1950), общее измерения полной качественности или вероятности оценки, такой, что чем меньше оценка, тем лучше. Оценка Байера (Brier) для N пунктов:

$$B = \frac{1}{N} \sum_{i=1}^N (r_i - c_i)(r_i - c_i)'$$

Где r_i является вектором оцененных вероятностей для k альтернатив пункта i , r_i , равняется $(r_{i1} \dots r_{ik})$, c_i , является связанным вектором исхода, $c_i = (c_{i1} \dots c_{ji} \dots c_{ki})$, где c_{ji} равняется единице для истинной альтернативы и нулю в другом случае, а штрих (') обозначает вектор-столбец. Мерфи показал, что оценка Байера может быть разделена на три составные части. Чтобы это сделать, распределим N векторов ответа на T подгрупп так, чтобы все векторы ответа r_i в подмножестве t являются идентичными. Пусть n_t будет

количество ответов в подмножестве t , и пусть C_t будет вектор исправления пропорции для подмножества t :

$$\bar{c}_t = (\bar{c}_{1t}, \dots, \bar{c}_{jt}, \dots, \bar{c}_{kt}), \text{ где } \bar{c}_{jt} = \sum_{i=1}^{n_t} c_{jt} / n_t$$

Пусть \bar{c} – вектор доли правильных ответов для всех,

$$\bar{c} = (\bar{c}_1, \dots, \bar{c}_j, \dots, \bar{c}_k), \text{ где } \bar{c}_j = \frac{1}{N} \sum_{t=1}^N c_{jt}$$

Наконец, пусть u будет вектор единства (the unity vector), вектор строки, чьи k элементы – все единицы.

Тогда разделение Мерфи оценки Байера:

$$B = \bar{c}(u - \bar{c}') + \frac{1}{N} \sum_{t=1}^T n_t (r_t - \bar{c}_t)(r_t - \bar{c}_t)' - \frac{1}{N} \sum_{t=1}^T n_t (\bar{c}_t - \bar{c})(\bar{c}_t - \bar{c})'$$

Первое слагаемое – не функция оценок вероятности; скорее, оно отражает относительную частоту истинных событий для k альтернатив. Например, предположим, что все оцениваемые пункты имели одни и те же две альтернативы: (дождь, нет дождя). Тогда первое слагаемое – функция базового значения дождя для N пунктов (или дней). Если бы всегда (или никогда) шел дождь, это слагаемое было бы нулевым. Его максимальное значение, $(k-1)/k$, указало бы на максимальную неопределенность относительно возникновения дождя. Второе слагаемое – мера калибровки, оцененное среднее число взятой в квадрат разности между ответами в категории и исправлением пропорции для этой категории. Третье слагаемое, называемое решением, отражает способность оценивающего сортировать события в подкатегории, для которых исправление пропорции отличается от полного исправления пропорции.

Разделение Мерфи было предназначено для повторного прогнозирования того же самого набора событий (например, “дождь” вместо “нет дождя”). Когда альтернативы не имеют никакого общего значения для всех пунктов (например, в вопросах с множественным выбором), тогда все, на что указывает первое слагаемое, это степень, в которой правильный ответ появляется одинаково часто как и первая, вторая, и т.д., альтернатива.

Когда записан только один ответ для одного пункта, разделение Мерфи (Murphy, 1972) сокращается до:

$$B' = \bar{c}(1 - \bar{c}) + \frac{1}{N} \sum_{t=1}^T n_t (r_t - \bar{c}_t)^2 - \frac{1}{N} \sum_{t=1}^T n_t (\bar{c}_t - \bar{c})^2$$

Где \bar{c} – полная доля правильных ответов и \bar{c}_t – доля правильных ответов в подкатегории t . Когда записанные ответы – это ответы, которые больше или равняются 0.5 (как с задачей с двумя альтернативами неполного диапазона), первое слагаемое отражает способность субъекта выбрать правильную альтернативу и, таким образом, могло бы называться знанием. Как прежде, второе слагаемое измеряет калибровку, а третье – решение.

Подобные измерения калибровки были предложены Адамсами и Оскампом (Adams и Adams, 1961, Oskamp, 1962). Ни одно из этих измерений калибровки не различает самонадеянности и неуверенности в себе. Свойства выборки этих мер не известны.

Метеорологическое исследование

В 1906, В. Эрнест Кук, правительственный астроном в Западной Австралии, высказал точку зрения, что каждый метеорологический прогноз должен сопровождаться единственным числом, которое “указывало бы, приблизительно, степень вероятности, которую сам метеоролог приписывает этому предсказанию” (Cooke, 1906b, с. 274). Он сообщил о результатах 1.951 предсказаний (Cooke, 1906a, 1906b). Из тех, которым он приписал самую высокую степень вероятности (“почти уверен, что правильно”), 0.985 были правильными. Для средней степени вероятности (“средняя вероятность”), 0.938 были правильными, в то время как для самой низкой степени вероятности (“сомнительный”), 0.787 были правильными.

В 1951, Уильямс (Williams) попросил восемь профессиональных метеорологов в Солт-Лейк-Сити оценить вероятность выпадения осадков для каждого из 1095 12-часовых прогнозов, используя одно из чисел 0, 0.2, 0.4, 0.6, 0.8 и 1.0. Почти по всему диапазону, доля дней с осадками была ниже, чем приписанная вероятность. Это могло бы отражать довольно естественную форму перестраховки в публичных утверждениях. Люди, намного более вероятно, будут критиковать прогноз погоды, который оставит их в дождь без зонтика, чем прогноз, прислушавшись к которому они носили зонтик в дни без осадков.

Подобные результаты показало исследование Мерфи и Уинклера (Murphy и Winkler, 1974). Их метеорологи дважды оценили вероятность выпадения осадков на следующий день, как до, так и после наблюдения результатов компьютеризированной системы предсказания погоды (PEATMOS). Эти 7.188 оценок (как до, так и после PEATMOS) показали ту же самую переоценку вероятности дождя, обнаруженную Уильямсом (Williams).

Сандерс (Sanders, 1958) собрал 12.635 прогнозов, используя 11 ответов 0, 0.1, ..., 0.9, 1.0, для следующих дихотомизированных событий: направление

ветра, скорость ветра, порывы, температуры, облачность, высота облачности, видимость, выпадение осадков, тип осадков и гроза. Эти данные показали только небольшую тенденцию оценок вероятности синоптиками превышать пропорцию погодных событий, которые произошли.¹ Рут (Root, 1962) сообщил о симметричности калибровки 4.138 прогнозов осадков: приписанные вероятности были слишком низки в низком диапазоне и слишком высоки в высоком, относительно наблюдаемых частот.

Уинклер и Мерфи (Winkler и Murphy, 1968a) сообщили о калибровочных кривых за год прогнозов осадков в Хартфорде, штат Коннектикут. Каждый прогноз был либо для 6-часового, либо 12-часового периода времени, с временем подготовки от 5 до 44 часов. К сожалению, было неясно, включили ли синоптики “признаки осадков” (меньше чем 0.01 дюйма) в свои предсказания. Данные были проанализированы дважды, предполагая, что “осадки” включают или не включают признаки. Включение или исключение признаков имело существенное влияние на калибровку, как и временной период. 6-часовые прогнозы с признаками и 12-часовые прогнозы, и без признаков, показали превосходную калибровку. Кривая калибровки для 12-часовых прогнозов с признаками располагается выше линии идентификации; кривая для 6-часовых прогнозов, без признаков, лежит значительно ниже ее. Разница во времени подготовки не затрагивала калибровку.



Рис. 1. Данные калибровки для прогнозов осадков. Количество прогнозов показано для каждой точки. (Источник: Murphy и Winkler, 1977a.)

¹ Ссылки на Кук (Cooke, 1906), Вильямс (Williams, 1951) и Сандерс (Sanders, 1958) были предоставлены нашему вниманию Раиффом (1969).

Национальная метеорологическая служба выражает свои прогнозы возникновения осадков в вероятностных терминах, начиная с 1965. Была опубликована калибровка для некоторых частей этой массивной базы данных (Murphy и Winkler, 1977a; Американское метеорологическое бюро, 1969). За эти годы калибровка улучшилась. Рисунок 1 показывает калибровку для 24.859 прогнозов осадков, сделанных в Чикаго в течение четырех лет, заканчивая июнем 1976. Он показывает очень хорошую калибровку; Мерфи (Murphy, 1980) утверждает, что данные, полученные в течение недавних лет, даже лучше! Он приписывает лучшее выполнение работы опыту в оценке вероятности, который синоптики приобрели за эти годы, а также тому факту, что эти данные были собраны на реальной ситуации на рабочем месте.

Ранние лабораторные исследования

В 1957 Адамс (Adams) сообщил о калибровке испытуемых, которые использовали 11-ти бальную шкалу уверенности: испытуемых “попросили выразить свою уверенность в терминах процента ответов, сделанных на определенном уровне уверенности, которые, как они ожидают, будут правильными. ... Из ответов, сделанных с уверенностью p , приблизительно $p\%$ должны быть правильными” (с. 432-433).

В задаче Адамса, каждое из 40 слов было представлено с помощью тахистоскопа 10 испытуемым 10 раз подряд, каждый раз с более ярким освещением. После каждого показа каждый испытуемый записывал, что по его мнению, он видел и давал оценку уверенности. Результирующая кривая калибровки показала, что пропорции, которые были правильными, значительно превышали оценки уверенности по полному диапазону ответов (кроме ответов 100). Интерпретация этих данных должна производиться очень осторожно: поскольку каждое слово показывали 10 раз, ответы сильно взаимозависимы. Неизвестно, какое влияние такая взаимозависимость оказывает на калибровку. Испытуемые, возможно, хотели “сдержаться” при первых показах слова, не желая давать высоких ответов, когда они знали, что это же слово будет представлено еще несколько раз.

В следующем году, Адамс и Адамс (Adams и Adams, 1958) сообщили об обучающем эксперименте, используя ту же шкалу ответов, но новую, трехальтернативную задачу с единственным ответом: Для каждой из 156 пар слов за показ, испытуемых спрашивали, были ли слова антонимами, синонимами или несвязанными. Средние калибровочные оценки (основанные на абсолютной разности $[r_i - c]$), у 14 испытуемых экспериментальной группы, которым были показаны числа и кривые калибровки после каждого из пяти показов, уменьшилась на 48% от первого серии к последней. Шесть испытуемых контрольной группы, единственной обратной связью которых был список их не засчитанных ответов, показали среднее увеличение на 36% в оценках несоответствия.

Адамс и Адамс (Adams и Adams, 1961) обсуждали много аспектов калибровки (используя термин: *реализм уверенности*), предвосхищая много работы, сделанной другими, в недавние годы, и представили большее количество сведений, включая чрезвычайно самонадеянную кривую калибровки пизофреника, который полагал, что он Иисус Христос. В задаче изучения бессмысленных слогов они обнаружили большую самоуверенность в первом испытании и улучшение после 16 испытаний. Они также кратко описали процедуру эксперимента: в первый день испытуемые приняли 108 решений по поводу процента синих точек в последовательности синих и красных точек. На второй и четвертый день испытуемые определяли истинность или ложность 250 утверждений на общеобразовательные темы. На третий день они с завязанными глазами поднимали предметы, чтобы определить их вес. На пятый день они приняли 256 решений по поводу отношения между парами слов (синонимия, антонимия или отсутствие отношения). Восемь испытуемых экспериментальной группы, которым предлагали обратную связь калибровки по истечению каждого из первых четырех дней, на пятый день показывали среднее абсолютное значение несоответствия, которое было значимо ниже, чем у 8 испытуемых контрольной группы (без обратной связи), предполагающее некоторый трансфер обучения. Наконец, авторы сообщили, что для 56 испытуемых, сдающих экзамен по элементарной психологии с множественным выбором, более плохая калибровка была связана с большим страхом неудачи ($r=0.36$). К страху неудачи не имели отношения ни наличие знаний, ни самонадеянность.

Оскамп (1962) предъявлял испытуемым 200 профилей ММРІ² в качестве стимула. Половина этих профилей принадлежала людям, принятым в больницу управления ветеранов с психическими болезнями; остальные попали туда по чисто медицинским причинам. Задачей испытуемых было определить, для каждого профиля, имел ли пациент психическое или медицинское заболевание, и указать вероятность того, что их решение было правильным. Каждый профиль был охарактеризован как трудный (61), средний (88) или легкий (51) на основе статистически полученной классификации, которая различала 57%, 69% и 92% трудных, средних и легких профилей, соответственно.

Все 200 профилей были оценены тремя группами испытуемых: 28 старшекурсников, специализирующихся в психологии, 23 стажера-психолога в клинике и 21 опытный клинический психолог. 28 неопытных испытуемых были позже разбиты на две согласованные группы, и им снова дали те же самые 200 профилей. Половину из них обучили в течение второго испытания, как улучшить точность; остальных обучили, как улучшить калибровку.

Оскамп (Oskamp) использовал три измерения выполнения работы испытуемых: точность (исправление пропорции), уверенность (средний ответ

² Профиль ММРІ (Minnesota Multiphasic Personality Inventory) – представленные в виде графика данные обследования по Миннесотскому Многофакторному Личностному опроснику.

вероятности), и правильность уверенности (счет калибровки):

$$\frac{1}{N} \sum_t n_t |r_t - c_t|.$$

Все три группы имели тенденцию быть самоуверенными, особенно старшекурсники в первом испытании (точность 70%, уверенность 0.78). Однако все три группы были неуверенными в себе в легких профилях (точность 87%, уверенность 0.83).

Испытуемые, обученные улучшению точности, увеличили свою точность от 67% до 73%, приблизившись к их уровню уверенности 0.78, который не изменился в результате обучения.³ Субъекты, обученные калибровке, понизили свою уверенность от 0.78 до 0.74, приблизив ее к своей точности 68%, которая осталась неизменной. Как и ожидалось от этих изменений, оценка калибровки обеих групп улучшилась.

Исследование обнаружения сигнала

На ранних стадиях исследования обнаружения сигнала, исследователи изучили возможность использования скорее оценок уверенности, чем ответов да-нет, чтобы уменьшить количество данных, требуемых для определения стабильных кривых операционных характеристик приемника (ОХП). Свит, Таннер и Бедсалл (Swets, Tanner, и Birdsall, 1961) попросили, чтобы четыре наблюдателя указали свою уверенность в том, что они слышали сигнал плюс шум, а не просто шум, для каждого из 1.200 испытаний. Хотя трое из четырех испытуемых были сильно калиброваны, все кривые очень различались. Один испытуемый показал серьезную тенденцию назначать слишком маленькие вероятности (например, сигнал присутствовал более, чем в 70% раз, когда этот субъект использовал категорию ответа "0.05-0.19").

Кларк (Clarke, 1960) предъявил одно из пяти различных слов, смешанных с шумом, слушателям через наушники. Слушатели выбрали слово, которое, как они думали, они слышали и затем оценивали свою уверенность, указывая одну из 5 категорий, которые разделяли шкалу вероятности на пять диапазонов. После каждого из 12 практических испытаний 75 пунктов, слушатели записывали свои собственные результаты и отмечали процент правильных идентификаций в каждой категории оценки, таким образом, они могли менять стратегию на следующем испытании. Кларк (Clarke) обнаружил, что, хотя все пять слушателей оказались хорошо калиброванными, когда данные были усреднены по пяти словам стимула, анализ отдельных слов показал, что слушатели имели тенденцию быть самоуверен-

³ Любители ММРІ могли бы обратить внимание, что с этим минимальным обучением старшекурсники показали столь же высокую точность как лучшие эксперты или лучшие системы прогнозирования.

ными для плохо понятных слов и неуверенными для хорошо понятных слов.

Поллак и Декер (Pollack и Decker, 1958) использовали устно определенную 6-ти бальную шкалу уверенности, от “Я уверен, что я получил сообщение правильно” до “Я уверен, что я получил сообщение неправильно.” С этой шкалой оценки невозможно определить, хорошо ли калиброван индивидум, но возможно увидеть изменения в калибровке во всех условиях. Кривые калибровки для легких слов обычно лежат выше кривых калибровки для трудных слов, независимо от отношения “сигнал-шум”, и кривые для высоких значений отношения “сигнал-шум” лежат выше кривых для низких значений отношения “сигнал-шум”, независимо от трудности слова.

В большинстве этих исследований, калибровка мало интересовала исследователей; важным вопросом было то, не дадут ли оценки уверенности те же самые кривые ОХП что процедуры “да-нет”. К 1966, Грин и Свитс (Green & Swets) заключили, что шкалы оценок и процедуры “да-нет” дают почти идентичные кривые ОХП. С тех пор, исследования калибровки исчезли из литературы по обнаружению сигнала.

Недавнее лабораторное исследование

Чрезмерная уверенность. Наиболее важное открытие в недавнем исследовании – то, что люди являются самонадеянными в общеобразовательных вопросах умеренной или повышенной трудности. Некоторые типичные результаты, показывающие самонадеянность представлены на рис. 2. Хазард и Петерсон (Hazard & Peterson, 1973) попросили 40 служащих вооруженных сил, обучающихся в Школе внешней разведки, определить вероятность или шансы для 50 вопросов по общей осведомленности, имеющим две альтернативы (например, “Какой журнал имел самый большой тираж в 1970, *Плейбой* или *Тайм*?”). Лихтенштейн (Lichtenstein, неопубликовано) получил подобные результаты, используя те же самые пункты, но только для 19 служащих научно-исследовательского института штата Орегон, также, как Филлипс и Райт (Phillips, Wright, 1977) с различными пунктами, используя Британских старшекурсников в качестве испытуемых.

Другие многочисленные исследования, использующие вопросы на общие темы, показали ту же самую самонадеянность (Fischhoff, Slovic и Lichtenstein, 1977; Koriat, Lichtenstein и Fischhoff, 1980; Lichtenstein и Fischhoff, 1977, 1980a, 1980b; Nickerson и McGoldrick, 1965). Кембридж и Шрекенгост (Cambridge, Shreckengost, 1978) обнаружили самонадеянность у аналитиков Центральной спецслужбы. Фишхофф и Словик (Fischhoff и Slovic, 1980) обнаружили серьезную самонадеянность, используя множество невозможных или почти невозможных задач (например, прогнозирование победителя в скачках на полтора километра или диагноз злокачественности опухоли). Питц (Pitz, 1974) сообщил о самонадеянности, используя метод полного диапазона.

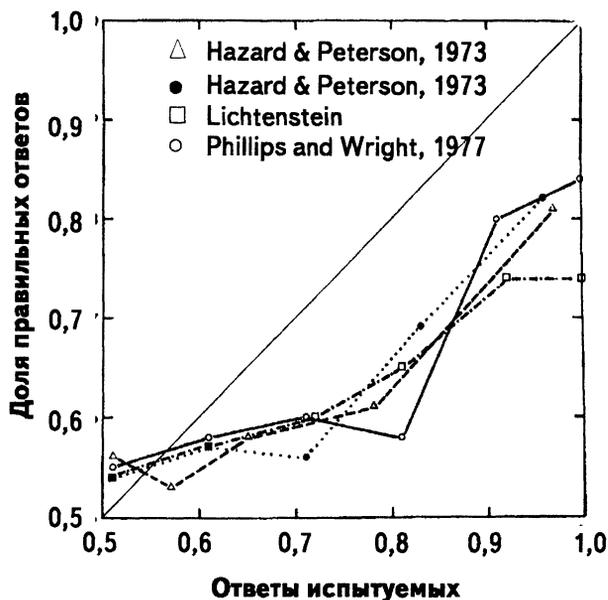


Рис. 2. Калибровка для половины диапазона для пунктов на общее знание

Фишхофф, Словик и Лихтенштейн (Fischhoff, Slovic и Lichtenstein, 1977) сосредоточились на правильности выражения уверенности. Используя разнообразные методы (без альтернатив, одна альтернатива и две альтернативы для половины диапазона и с полным диапазоном), они обнаружили, что только от 72% до 83% пунктов, на которые были даны ответы 1,0, были правильными. В задачах с полным диапазоном, пункты, которым был присан другой экстремальный ответ, нуль, были правильными от 20% до 30% раз. Использование вероятностного ответа не исправляло самонадеянность. Ответы, которым назначили вероятность 1.000:1 в том, что они правильные, только от 81% до 88% были правильными; для вероятности 1.000.000:1 правильная альтернатива была выбрана только от 90% до 96% раз. Испытуемые охотно использовали экстремумную вероятность; в одном из экспериментов почти четвертая часть ответов была 1.000:1 или больше. Далее исследования показали, что чрезвычайная самонадеянность распространялась не только на нескольких испытуемых или несколько пунктов.

Влияние трудности. Самонадеянность наибольшая в задачах большой трудности (Clarke, 1960; Nickerson и McGoldrick, 1965; Pitz, 1974). При рассмотрении невозможных задач (различить европейский и американский профиль, рисунки азиатских и европейских детей, рост и падение цен на

акции) кривые калибровки не повышались вообще; для всех оцененных вероятностей, доля правильно выбранных альтернатив была близка к 0.5 (Lichtenstein и Fischhoff, 1977). Испытуемые не отказывались использовать высокие вероятности в этих задачах; от 70% до 80% всех ответов были больше 0.5.

По мере того, как задачи становятся легче, самонадеянность уменьшается. Лихтенштейн и Фишхофф (Lichtenstein и Fischhoff, 1977) позволили одной группе испытуемых в задаче различения профиля изучить правильно помеченный набор типовых стимулов перед оцениванием вероятности. Этот опыт сделал задачу намного более легкой (71%, правильных ответов против 51% для контрольной группы) и экспериментальная группа была лишь немного самонадеянна. Лихтенштейн и Фишхофф (1977) выполнили последующий анализ влияния трудности на калибровку, используя две большие совокупности данных из двуальтернативных задач для половины диапазона на общие темы. Они отделили легкие пункты (для которых большинство испытуемых выбрало правильную альтернативу) от трудных, и хорошо осведомленных испытуемых (тех, которые выбрали наиболее правильные альтернативы) от менее осведомленных. Они обнаружили систематическое уменьшение самонадеянности с увеличением процента правильных ответов. Действительно, хорошо осведомленные испытуемые при ответе на самые легкие пункты были *неуверены* в себе (например, 90% правильных ответов при ответе с вероятностью 0.80). А находка была повторно обнаружена с двумя новыми группами испытуемых, которым были предложены наборы вопросов, выбранных, как трудные или легкие на основе ответов предыдущих групп. Результирующие кривые калибровки показаны на Рисунке 3, наряду с соответствующими кривыми калибровки последующего анализа.

В упомянутом исследовании трудность была определена на основе ответов испытуемых (Clarke, 1960; Lichtenstein и Fischhoff, 1977). Ранее, Лихтенштейн и Фишхофф (Lichtenstein и Fischhoff, 1980a), после Оскампа (Oskamp, 1962), разработали набор из 500 двух-альтернативных вопросов на общие темы, трудность которых могла быть определена независимо. Имелось три типа вопросов: какой из двух городов, штатов, стран или континентов является более густонаселенным (например, Лас Вегас или Майами), какой из двух городов находится дальше на расстоянии от третьего города (например, “Находится ли Мельбурн дальше от Рима или от Токио?”), и какое историческое событие произошло первым (например, подписание Великой хартии вольностей или рождение Мухаммеда). Таким образом, каждый вопрос связывал два числа (население, расстояние или прошедшее время с настоящим). Отношение большего к меньшему из этих чисел было принято как мера трудности: 250 вопросов с самыми большими отношениями были обозначены как *легкие*; оставшиеся — как *трудные*. Эта априорная классификация была весьма успешна; для более чем 35 испытуемых, процент правильных ответов составил 81 для легких вопросов и 58 — для

трудных. Эти результаты, также, показали самонадеянность для трудных вопросов и неуверенность испытуемых для легких.

Влияние легкости или трудности, кажется, является результатом неспособности экспертов понять, насколько вопрос трудный или легкий.

Филлипс и Чу (Phillips, Chew, не опубликовано) не обнаружили корреляции между процентом правильных ответов и оценками трудности по 11-ти бальной шкале. Однако испытуемые дают различные распределения ответов для различных задач; Лихтенштейн и Фишхофф (Lichtenstein и Fischhoff, 1977) отметили корреляцию в 0.91 между процентом правильных ответов и средним ответом для 16 различных наборов данных. Но разница в распределении ответов – меньше чем, она должна быть: по тем же 16 наборам данных, доля правильных ответов варьировалась от 0.43 до 0.92, в то время как средний ответ, варьировался только от 0.65 до 0.86.

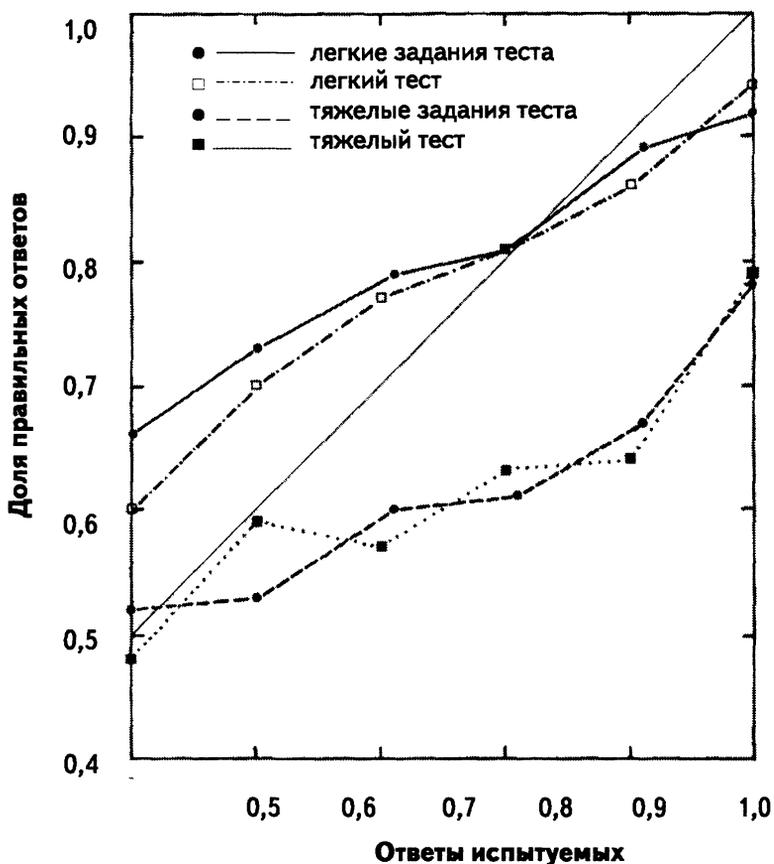


Рис. 3. Калибровка для трудных и легких тестов и для трудных и легких заданий тестов

Феррелл и МакГоу (Ferrell и McGoeu, 1980) недавно разработали модель для калибровки дискретных оценок вероятности, которая учитывает эффект легкости – трудности. Модель, основанная на теории обнаружения сигнала, предполагает, что эксперты преобразовывают свое чувство субъективной неопределенности в переменную решения, X , которая разделена на секции с границами $\{x_i\}$. Оценивающий сообщает о вероятности r_i , всякий раз, когда X находится между x_{i-1} и x_i . Феррелл и МакГоу (Ferrell и McGoeu) предполагают, что, при отсутствии обратной связи о действии калибровки, оценивающий не будет изменять набор значений пределов, $\{x_i\}$, по мере того, как изменяется трудность задачи. Это предположение ведет к прогнозированию самонадеянности в трудных вопросах и неуверенности в легких. Применение модели к большому количеству данных из работ Лихтенштейна и Фишхоффа (1977) показало соответствие как по кривым калибровки, так и по распределению ответов согласно предположению, что значения пределов остались постоянными по мере изменения трудности. Таким образом, влияние трудности или легкости рассматривается как неспособность изменить пределы, вовлеченные в преобразование от чувства определенности к вероятностным ответам.

Влияние базового значения. Одно-альтернативные (истина-ложь) задачи могут быть охарактеризованы долей истинных утверждений в наборе вопросов. Чтобы быть хорошо калиброванным на некотором наборе вопросов, нужно принимать во внимание информацию базового значения. Модель обнаружения сигнала Феррелла и МакГоу (Ferrell и McGoeu, 1980) предполагает, что на калибровку воздействуют независимо (а) доля истинных утверждений и (б) способность оценивающего различать истинные утверждения от ложных. Допуская, что значения границ $\{x_i\}$, остаются постоянными, модель прогнозирует различные воздействия на калибровку в зависимости от изменения доли истинных утверждений (оставив различимость постоянной) в отличие от изменения различимости (оставляя постоянной пропорцию истинных утверждений). Феррелл и МакГоу (Ferrell и McGoeu) представили данные, подтверждающие их модель. Студенты на трех курсах, изучающие технические науки оценили вероятность, что ответы, которые они написали на экзаменах, будут оценены экзаменатором как правильные. Последующий анализ, распределяющий испытуемых на четыре группы (высокий или низкий процент правильных ответов, высокая или низкая различимость), показал различия калибровки, спрогнозированные моделью. Неопубликованные данные, собранные Фишхоффом и Лихтенштейном, показанные на Рисунке 4, также подтверждают правильность модели. Четыре группы испытуемых получили 25 одно-альтернативных вопросов на общие темы (например, “Написана ли Энеида Гомером?”), различающиеся по количеству истинных утверждений: 8%, 20%, 50% и 71%. Группы показали сильно различающиеся кривые калибровки, приблизительно той же формы как спрогнозировали Феррелл и МакГоу для изменения базового значения, при постоянной различимости.

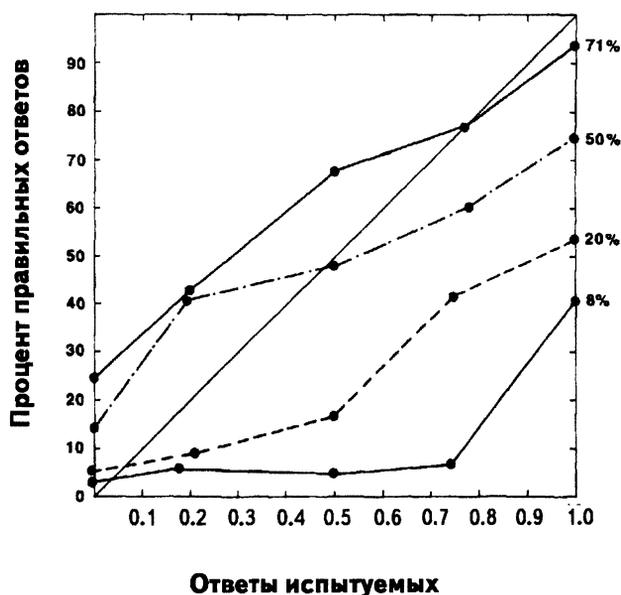


Рис. 4. Влияние на калибровку изменений в проценте истинных утверждений (Источник: Fischhoff и Lichtenstein, неопубликовано)

Индивидуальные различия. Неквалифицированные утверждения, что один человек лучше калиброван, чем другой, трудно сделать по двум причинам. Во-первых, необходимо, по крайней мере, несколько сотен ответов, чтобы получить стабильное измерение калибровки. Во-вторых, оказывается, что калибровка сильно зависит от задачи, особенно от ее трудности. Действительно, Лихтенштейн и Фишхофф (1980а) предполагали, что для каждого человека может существовать “идеальная” задача (то есть, задача, уровень трудности которой не приводит ни к самонадеянности, ни к неуверенности в себе, и таким образом, задача, в которой человек будет наилучшим образом калиброван). Однако, уровень трудности “идеальной” задачи может быть разным для разных людей. Таким образом, даже, когда один человек лучше, чем другой в некотором наборе вопросов, для более трудного или легкого набора может быть истинным обратное.

Сравнения между различными группами субъектов показали мало различий, когда трудность управлялась. Аспиранты-психологи, которые, предположительно, более осведомлены, чем обычные испытуемые (старшекурсники, которые ответили на объявление в газете колледжа), не показали никакой разницы в калибровке (Lichtenstein и Fischhoff, 1977). И при этом мы не обнаружили различия в калибровке или самонадеянности между мужчинами и женщинами (Lichtenstein и Fischhoff, 1981).

Райт и Филлипс (1976) исследовали отношения среди нескольких измерений качеств личности (авторитаризм, консерватизм, догматизм и нетерпи-

мость к двусмысленности), вербальные выражения неопределенности (например, количество слов таких, как *маловероятно*, использованных в коротких письменных ответах на 45 вопросов), и несколько измерений калибровки. Единственные связи, которые они обнаружили между шестью шкалами личности и семью измерениями калибровки, были две скромные корреляции (0.41 и 0.34) со шкалой авторитаризма. Калибровка ответов определенности (то есть, ответы 1.0) не коррелировала с калибровкой ответов неопределенности (< 1.0). Измерения вербальной неопределенности были некоррелированы с любым из числовых измерений калибровки. Авторы заключили, что вероятностное мышление не является ни единственным фактором, ни сильно связанным с индивидуальными различиями качеств личности.

Райт и другие (1978) исследовали межкультурные различия в калибровке. Калибровка их британской выборки показана на рис. 2 (обозначенная как Phillips и Wright, 1977). Их другими выборками были студенты из Гонконга, Индонезии и Малайзии. Азиатские группы показали по существу плоские кривые калибровки. Авторы решили, что азиатские философы, интересующиеся вопросами судьбы, могли бы объяснить эти различия.

Попытки коррекции. Фишхофф и Словик (Fischhoff и Slovic, 1980) пробовали предотвратить появление самонадеянности на задаче различения рисунков азиатских от европейских детей, используя явно обескураживающие инструкции:

Все рисунки были взяты из Детской художественной коллекции доктора Роды Келлогг (Rhoda Kellogg), главного сторонника теории, что рисунки детей из разных стран и культур очень похожи.... Помните, что может быть совершенно невозможно различить эти рисунки. Сделайте все, что в ваших силах. Но если, в крайнем случае, Вы чувствуете полную неопределенность относительно происхождения этих рисунков, не смущайтесь отвечать 0.5 для каждого из них. (с. 792)

Эти инструкции понизили средний ответ примерно до 0.05, но тем не менее была обнаружена существенная самонадеянность.

Улучшит ли калибровку усиление мотивации? Сибер (Sieber, 1974) сравнил калибровку двух групп студентов на связанных с курсом набором вопросов с четырьмя альтернативами. Одной группе сказали, что они сдают экзамен в конце семестра. Другой группе было представлено не как экзамен в конце семестра, а как подготовка к нему. Эти две группы не отличались по числу выбранных правильных альтернатив, но предположительно более мотивированная группа (та, в которой ответы определяют их уровень) показала значимо худшую калибровку (большую самонадеянность).

Обученные эксперты, с обратной связью о калибровке, показали смешанные результаты. Как упомянуто, Адамс и Адамс (1958) обнаружили незначительное улучшения качества калибровки после пяти сессий обучения и, в более позднем исследовании (1961) наблюдалась некоторая генерализация обучения. Чу (Choo, 1976), используя только одну сессию обучения с 75 двух-альтернативными вопросами на общие темы, обнаружил малое улучшение и никакой генерализации.

Лихтенштейн и Фишхофф (1980b) обучили две группы испытуемых путем предоставления обширной, персонифицированной обратной связи калибровки после каждой 2-й или 10-й серии вопросов, составленных из 200 двух-альтернативных вопросов на общие темы. Они обнаружили заметное улучшение калибровки, которая произошла между первой и второй сериями. Скромная генерализация произошла для задач с различными уровнями трудности, содержанием и способом ответа (четыре, а не две альтернативы), не было обнаружено никакого улучшения для задачи оценки квантиля (описанной в следующей главе) или различения европейского от американского образцов профиля.

Другой подход к улучшению калибровки – это реструктурировать задачу таким способом, который препятствует появлению самонадеянности. В исследовании Кориата, Лихтенштейна и Фишхоффа (Koriat, Lichtenstein и Fischhoff, 1980), испытуемые сначала отвечали на 30 двух-альтернативных вопросов на общие темы обычным способом. Затем они получили 10 дополнительных вопросов. Для каждого вопроса они записали все причины, которые они могли придумать, в поддержку или опровержение каждому из двух возможных ответов, а затем сделали обычный выбор и оценки вероятности. Эта процедура значительно улучшила их калибровку. Дополнительное исследование помогло точно определить эффективный компонент этого метода. После ответа на начальный набор из 30 вопросов, испытуемым предлагали еще 30 вопросов. Для каждого, они сначала выбрали предпочтительный ответ, затем писали (а) причину, поддерживающую их выбранный ответ, (b) причину, противоречащую выбранному ответу или (c) две причины, одну в поддержку и одну в опровержение. Затем они оценили вероятность, что выбранный ответ был правилен. Только та группа, которую попросили написать опровержения, показала улучшенную калибровку. Этот результат, также как исследования корреляции, основываясь на данных от первого исследования, говорит о том, что эффективным средством от самонадеянности является поиск причины, почему человек *не прав*.

Экспертиза. Студенты, изучающие в колледже определенный предмет, являются экспертами, по крайней мере, временно, в материале темы предмета. Сибер (Sieber, 1974) сообщил о превосходной калибровке у студентов, сдающих пробный экзамен (то есть, у группы студентов, которым сказали, что это *не* экзамен). Более чем 98% их ответов с вероятностью 1.0 и только 0.5% из их ответов с вероятностью 0 были правильными. Пиз (Pitz, 1974) попросил студентов сделать прогноз своих оценок на его курсе; они также были хорошо калиброваны.

Были ли бы эти испытуемые так же калиброваны на вопросах эквивалентной трудности, которые не были в их области изучения? Лихтенштейн и Фишхофф (1977) попросили аспирантов-психологов ответить на 50 двух-альтернативных вопросов на общие темы и на 50 вопросов, охватывающих знания по психологии (например, “тест Ипсихара – это (а) перцептивный тест,

(b) тест социального беспокойства”). Два субтеста имели равную трудность, и калибровка была одинаковой для этих двух задач.

Кристенсен-Сзалански и Бушпихед (Christensen-Szalanski и Bushyhead, 1981) сообщили об оценках вероятности пневмонии девяти врачей для 1.531 пациентов, которые были обследованы из-за жалоб на кашель. Их калибровка была ужасна; кривая повышалась настолько медленно, что для самого высокого уровня уверенности (приблизительно 0.88) доля пациентов, фактически больных пневмонией, была меньше чем 0.20. О подобных результатах было сообщено Ластедом (Lusted, 1977) для диагнозов перелома черепа и пневмонии, и ДеСметом, Фрайбеком и Торнбери и для диагнозов перелома черепа (DeSmet, Fryback, и Thornbury, 1979). Результаты этих полевых исследований с врачами четко противоречат превосходной калибровке прогнозов погоды метеорологами. Мы считаем, что несколько факторов благоприятствуют синоптикам. Во-первых, они делают вероятностные прогнозы в течение многих лет. Во-вторых, задача постоянно повторяется; вопрос, на который нужно ответить (Будет ли дождь?) – всегда тот же самый. Напротив, лечащий врач постоянно рассматривает множество возможностей (Действительно ли это перелом черепа? У нее фарингит? Он нуждается в дальнейшей госпитализации?). Наконец, обратная связь исхода для синоптиков хорошо определена и быстро получена. Это не всегда так для врачей; пациенты могут не возвращаться или обратиться в другое место, или диагнозы остаются неопределенными.

Люди, которые держат пари или устанавливают вероятность на скачках, могли бы также считаться экспертами. При тотализаторном методе, окончательная вероятность определена количеством денег, поставленных на каждую лошадь, что позволяет вычислить своего рода кривую калибровки. Такие кривые (Fabricand, 1965; Hoerl и Fallin, 1974) показывают превосходную калибровку людей, с небольшой тенденцией ставить на кон слишком много на длинных дистанциях. Однако такие данные только теоретически связаны с оценкой вероятности. Более уместные результаты калибровки, сообщенные Дауи (Dowie, 1976), который изучал предполагаемые ставки, напечатанные ежедневно в спортивной газете в Британии. Эти прогнозы, в форме вероятностей, сделаны одним человеком для всех лошадей в данной скачке; приблизительно восемь человек делали прогнозы в течение исследуемого года. Калибровка прогнозов для 29.307 лошадей показала небольшую неуверенность для вероятностей большее чем 0.4 и превосходную калибровку для вероятностей меньше чем 0.4 (что включило 98% данных).

Развивающееся исследование относительно калибровки привело к развитию нового вида экспертизы: эксперты калибровки, которые знают об общих ошибках, которые люди делают в оценке вероятностей. Лихтенштейн и Фишхофф (1980a) сравнили калибровку 8 таких экспертов с 12 неопытными испытуемыми и 15 испытуемыми, которые предварительно были обучены, чтобы быть хорошо калиброванными. Нормативные эксперты не только преодолели самонадеянность, типично показанную неопытными испытуемыми, но и очевидно чрезмерно ее компенсировали, поскольку они были неуверенны-

ми в себе. Эксперты были более чувствительны к трудности вопросов, чем две другие группы.

Будущие события. Райт и Вишхухда (Wright & Wishudha, 1979) предположили, что калибровка для будущих событий может отличаться от калибровки для вопросов на выявление уровня общих знаний. Если это верно, это ограничило бы экстраполяцию от исследования с вопросами по общей осведомленности к прогнозированию будущих событий. К сожалению, вопросы Райта и Вишхухда на общие темы были более трудными, чем их будущие события, что может объяснить лучшую калибровку последних.

Фишхофф и Бейт (Fischhoff и Beyth, 1975) попросил 150 израильских студентов оценить вероятность 15 будущих событий, возможные исходы широко освещенной в средствах массовой информации поездки Президента Никсона в Китай и Россию (например, “Президент Никсон встретится с Мао, по крайней мере, один раз”). Результирующая кривая калибровки была весьма близка к линии идентификации. Однако, Фишхофф и Лихтенштейн (не опубликовано) недавно обнаружили, что калибровка будущих событий показала ту же высокую самонадеянность, как и для вопросов на общие темы сопоставимой трудности. Филлипс и Чу (не опубликовано) получили кривые калибровки для трех наборов вопросов: общие темы, будущие и прошлые события (например, “авиалайн-гигант потерпел крушение, и погибло более 100 человек за прошедший месяц”). Все три кривые показали самонадеянность. Калибровка для будущих и прошлых событий была идентична, и несколько лучше, чем для вопросов на общие темы. Уровни трудностей трех наборов вопросов не могли объяснить эти результаты.

Джек Доуи (Dowie) и его коллеги теперь собирают данные калибровки в Заочном университете в Милтоне Кейнесе, Англия, от нескольких сотен студентов на курсах по изучению риска, используя вопросы, связанные с курсом, на общие темы и о событиях будущего. Студенты получили общее представление о концепции калибровки, и им предложили обратную связь относительно их результатов и калибровки. Предварительные результаты (Dowie, 1980) говорят, что они были умеренно самонадеянны. Калибровка была лучшей в вопросах на общие темы и самая плохая в вопросах, связанных курсом, но значимость и происхождение этих различий еще нужно исследовать.

Непрерывные предположения: неопределенные величины

Метод квантиля

Неопределенность значения неопределенной непрерывной величины (например, какая доля студентов предпочитает скотч бурбону? Каково самое короткое расстояние от Англии до Австралии?) может быть выражена как функция плотности вероятности для возможных значений этой величины. Однако экспертов обычно не просят строить всю функцию. Вместо этого,

обычная процедура – некоторая разновидность метода квантиля. В этом методе оценивающий задает значения неопределенной величины, которые связаны с маленьким количеством предопределенных квантилей распределения. Для медианы или квантиля 0.50, например, оценивающий задает такое значение величины, что истинное значение, с равной вероятностью будет попадать выше или ниже заданного значения; квантиль 0.01 – это такое значение, что существует только 1 шанс из 100, что истинное значение является меньшим, чем заданное. Обычно оцениваются три или пять квантилей, включая медиану. В варианте под названием метод *тертилей*, оценивающий задает две величины (0.33 и 0.67 квантили) такие, что полный диапазон разделен на три одинаково вероятных секции.

Обычно регистрируются два измерения калибровки. *Межквартильный* индекс – процент вопросов, для которых истинное значение попадает в межквартильный диапазон (то есть, между квантилями 0.25 и 0.75). Полностью калиброванный человек будет, в конечном счете, иметь межквартильный индекс 0.50. *Индекс неожиданности* – это процент истинных значений, которые не попадают в крайние оцененные квантили. Когда наиболее крайние квантили оценены как 0.01 и 0.99, полностью калиброванный человек будет обладать индексом неожиданности 2. Большой индекс неожиданности показывает, что границы уверенности экспертов были слишком узкими, чтобы охватить достаточно истинных значений, и таким образом, указывает на самонадеянность (или гиперточность; Pitz, 1974). Неуверенность будет обозначена межквартильным индексом больше 50 и низким индексом неожиданности; ни о каких подобных данных не сообщалось в литературе.

Импульс для исследования калибровки функций плотности вероятности появился в 1969 в книге Альперта и Рейфа (Alpert и Raiffa, 1969, 21). Альперт и Рейфа работали со студентами бизнес факультета Гарварда, которые были знакомы с анализом принятия решений. В группе 1, все испытуемые оценивали пять квантилей, три из которых были 0.25, 0.50 и 0.75. Крайние квантили различались для четырех подгрупп; 0.01 и 0.99 (группа А); 0.001 и 0.999 (группа В); “минимальная возможная величина” и “максимальная возможная величина” (группа С); и “удивительно низкая величина” и “удивительно высокая величина” (группа D). Межквартильные индексы и индексы неожиданности для этих четырех подгрупп показаны в табл. 1. Разочарованные большим количеством случаев неожиданности, Альперт и Рейфа провели эксперимент с тремя дополнительными группами (2, 3 и 4) которые, после оценки 10 неопределенных величин, получили обратную связь в форме расширенного сообщения и объяснения результатов, наряду с призывом “Расширьте ваши крайние квантили!” Затем испытуемые ответили на 10 новых неопределенных величин. Результаты до и после получения обратной связи показаны в табл. 1. Результаты улучшились, но испытуемые все еще показывали значительную самонадеянность.

Хессиан и МакКарти (Hession и McCarthy, 1974) собрали данные, сопоставимые с первым экспериментом Альперта и Рейффы, используя 55 неопределенных величин и 36 аспирантов в качестве испытуемых. Их инструкции вынудили испытуемых убедиться, что интервал между квантилем 0.25 и квантилем 0.75 действительно захватывает половину вероятности. "Болезненно позднее обсуждение с отдельными испытуемыми прояснило, что эта проверка последовательности закончилась в большинстве случаев реорганизацией, уменьшая первоначально оцененный межквартильный диапазон" (с. 7) таким образом, все только ухудшилось! Этот акцент на обучение, не используемый Альпертом и Рейффом, может объяснять, почему испытуемые Хессиан и МакКарти были так плохо калиброваны, как показано в Таблице 1.

Таблица 1. Резюме калибровки для непрерывных вопросов: Процент истинных значений, попадающих в межквартильный диапазон и не попадающих в экстремумные квантили

	N	Наблюдаемые межквартильные индексы ^a	Индекс неожиданности	
			наблюдаемый	идеальный
<i>Альперт и Рейффа (1969)</i>				
Группа 1-А (0.01, 0.99)	880		46	2
Группа 1-В (0.001, 0.999)	500	33	40	0.2
Группа 1-С ("мин" и "макс")	700		47	?
Группа 1-Д ("удивительно высокая/низкая величина")	700		38	?
<i>Группы 2, 3 и 4</i>				
До обучения	2.270	34	34	2
После обучения	2.270	44	19	2
<i>Хессиан и МакКарти (1974)</i>	2.035	25	47	2
<i>Селвидж (Selvidge, 1975)</i>				
Пять квантилей	400	56	10	2
Семь квантилей (вкл. 0.1 и 0.9)	520	50	7	2
<i>Московитц и Буллерс (Moskowitz & Bullers, 1978)</i>				
<i>Пропорции</i>				
Три квантиля	120	—	27	2
Пять квантилей	145	32	42	2
<i>Доу-Джонс (Dow-Jones)</i>				
Три квантиля	210	—	38	2

Пикхарт и Воллейс (Pickhardt & Wallace, 1974)

Группа 1

Первое испытание	?	39	32	2
Пятое испытание	?	49	20	2

Группа 2

Первое испытание	?	30	46	2
Шестое испытание	?	45	24	2

<i>Т. А. Браун (Т.А. Brown, 1973)</i>	414	29	42	2
---------------------------------------	-----	----	----	---

Лихтенштейн и Фишхофф (Lichtenstein & Fischhoff, 1980b)

До испытания	924	32	41	2
После испытания	924	37	40	2

Сивер, фон Винтерфельдт и Эдвардз (Seaver, von Winterfeldt & Edwards, 1978)

Квантили	160	42	34	2
Вероятностные квантили	160	53	24	2
Вероятности	180	57	5	2
Шансы	180	47	5	2
Протокол шансов	140	31	20	2

Шефер и Борчердинг (Schaefer u Borcharding, 1973)

Первый день, квантили	396	23	39	2
Четвертый день, квантили	396	38	12	2
Первый день, гипотетическая выборка	396	16	50	2
Четвертый день, гипотетическая выборка	396	48	6	2

Ларсон и Ринан (Larson & Reenan, 1979)

"Вполне уверен"	450	—	42	?
-----------------	-----	---	----	---

Претт (Pratt, 1975)

"Удивительно высокие/ низкие" значения	175	37	5	?
--	-----	----	---	---

Продолжение табл. 1

<i>Мерфи и Винклер (Murphy и Winkler, 1974)</i>				
Экстремумы 0.125 и 0.875	132	45	27	25
<i>Мерфи и Винклер (Murphy и Winkler, 1977b)</i>				
Экстремумы 0.125 и 0.875	432	54	21	25
Штаел фон Хольстен (Stael von Holstem, 1971a)	1,269	27	30	2

Примечание: N - общее количество оцененных распределений.

^a Идеальный процент событий, попадающих в пределы межквартильного диапазона 50, для всех экспериментов кроме Брауна (1973). Он выявлял квантили 0.30 и 0.70, так что идеал – 40 %.

Хессиан и МакКарти также предложили испытуемым множество индивидуальных черт характера: авторитаризм, догматизм, жесткость, шкала ширины категории Петтигрю и интеллект. Корреляции экзаменационных отметок испытуемых с их межквартильными индексами и индексами неожиданности были весьма низкими, хотя шкала авторитаризма соотносилась как -0.31 с межквартильной оценкой и $+0.47$ с оценкой неожиданности ($N = 28$). Это соответствует выводу Райта и Филлипса (1976), что авторитаризм был мало связан с калибровкой.

Селвидж (Selvidge, 1975) расширил исследование Альперта и Рейффы, сначала задав испытуемым четыре вопроса относительно их самих (например, “Вы предпочитаете скотч или бурбон?”). Их ответы определили истинный ответ для долей, *произведенных группой* (например, какая доля испытуемых, ответивших на анкетный опрос, предпочитала скотч бурбону?). Одна группа дала пять квантилей, 0.01, 0.25, 0.50, 0.75 и 0.99. Другая группа дала эти же пять плюс два других: 0.1 и 0.9. Как показано в Таблице 1, группа с семью квантилями показала немного лучшие результаты. Результаты с пятью квантилями не столь отличны от результатов Альперта и Рейффы, как кажется. Три неопределенные величины Альперта и Рейффы были долями, произведенными группой, подобно вопросам Селвиджа. В этих трех вопросах, Альперт и Рейффа обнаружили 57% в межквартильном диапазоне и 20% неожиданностей. Наконец, для одного из вопросов, половину испытуемых в группе с пятью квантилями просили сначала дать 0.25, 0.50 и 0.75, а затем 0.01 и 0.99, в то время как другую половину попросили сначала оценить экстремумы. Селвидж обнаружил меньше случаев неожиданности для прежнего порядка (8%), чем для последнего (16%).

Московитц и Буллерс (Moskowitz и Bullers, 1978) также использовали пропорции, произведенные группой, но обнаружили намного больше случаев

неожиданности, чем Селвидж. Одна группа дала те же самые пять квантилей, которые использовал Селвидж (в порядке 0.5, 0.25, 0.75, 0.01, 0.99). Другую группу попросили сделать только три оценки (способ распределения и квантили 0.01 и 0.99). Перед оцениванием, группе с тремя квантилями представили некоторые типичные базовые события (например, “Рассмотрим лотерею, в которой участвуют 100 человек. Ваш шанс получить выигрышный билет равняется 1 из 100”), чтобы обеспечить лучшее понимание вероятности 0.01. Как показано в Таблице 1, у группы с тремя квантилями было меньшее количество случаев неожиданности, чем у группы с пятью квантилями. В другом эксперименте, использующем те же самые два метода, Москович и Буллерс попросили 44 старшекурсника, специализирующихся в коммерции, оценить среднее значение индустриального индекса Доу-Джонса (Dow-Jones) за 1977, 1974, 1965, 1960 и 1950 годы. Каждый испытуемый дал оценки до и после участия в обсуждениях с тремя людьми. Так как никаких систематических различий из-за участия в обсуждении не было обнаружено, данные были объединены в Таблице 1. Опять же, группа с тремя квантилями (в которой испытуемые получили представление о значении 0.01) имела меньшее количество случаев неожиданности, чем группа с пятью квантилями. Результаты группы с пятью квантилями были чрезвычайно плохими.

Пикхарт и Воллейс (Pickhardt и Wallace, 1974) повторил работу Альперта и Рейффы с изменениями. Для нескольких групп они сообщили об обнаружении от 38% до 48% случаев неожиданностей до получения обратной связи и не меньше, чем 30% неожиданностей после получения обратной связи. Две разновидности, используя или не используя зачет по предмету как награду за хорошую калибровку и используя или не используя расчетные правила обратной связи, не привели к каким-либо различиям в количестве случаев неожиданностей. Пикхарт и Воллейс также изучали влияние расширенного обучения: две группы из 18 и 30 испытуемых (о количестве неопределенных величин не сообщается) прошли пять и шесть серий испытаний с калибровочной обратной связью после каждого испытания. Было обнаружено небольшое улучшение, как показано в табл. 1.

Наконец, Пикхарт и Воллейс (1974) изучали влияние увеличивающегося знания на калибровку в контексте игры моделирования производства PROSIM. Тридцать два аспиранта произвели 51 оценку в течение смоделированных 17 “дней” планирования производства. Каждая оценка касалась события, которое произойдет через 1, 2 или 3 “дня”. Чем ближе время оценивания ко времени события, тем больше испытуемый знал о событии. Самонадеянность уменьшилась с увеличением информации: оказалось 32% неожиданностей с 3-дневными задержками, 24% с 2-дневными задержками, и 7% с 1-дневными задержками. Улучшения не наблюдалось в течение 17 “дней” стимуляции.

Т.А. Браун (1973) попросил 31 испытуемого оценить семь квантилей (0.01, 0.10, 0.30, 0.50, 0.70, 0.90, 0.99) для 14 неопределенных величин. Результаты, показанные в табл. 1, являются особенно обескураживающими.

ми, потому что каждый вопрос сопровождался обширными историческими данными (например, для вопроса “На какой отметке будет находиться индекс потребительских цен в декабре 1970?” испытуемым предлагали индексы потребительских цен в течение каждого квартала между мартом 1962 и июнем 1970). Для 11 вопросов, независимо от того, дал ли испытуемый исторический минимум как квантиль 0.01 и исторический максимум как квантиль 0.99, у них не было неожиданностей вообще. Другие 3 вопроса показали строгое увеличение или строгое уменьшение историй, и истинная величина была близка к простому приближению исторической тенденции. Испытуемые, должно быть, сильно полагались на свои собственные ошибочные знания, чтобы дать распределения, настолько плотные, чтобы показать 42% случаев неожиданности.

Лихтенштейн и Фишхофф (1980b) выявили пять квантилей (0.01, 0.25, 0.5, 0.75, 0.99) от 12 испытуемых на 77 неопределенных величинах, как до, так и после того, как испытуемые получили обширное обучение калибровке на дискретных двух-альтернативных вопросах. Как показано в Таблице 1, испытуемые не улучшили значимо калибровку по неопределенным величинам.

Другие методы

Сивер, фон Винтерфельдт и Эдвардз (Seiver, von Winterfeldt и Edwards, 1978) изучали влияние пяти различных способов ответа на калибровку. Две группы использовали метод одного квантиля, либо пяти квантилей (0.01, 0.25, 0.50, 0.75, 0.99), либо эквиваленты вероятности этих квантилей (1:99, 1:3, 1:1, 3:1, 99:1). Три других группы отвечали вероятностями, шансами или заполняли протоколы шансов на одно-альтернативные вопросы, которые определяли некоторое значение неопределенной величины (например, “Какова вероятность того, что население Канады в 1973 превысило 25 миллионов?”). Пять таких установленных значений давались для каждой неопределенной величины, и, исходя из ответов, экспериментаторы оценивали межквартильные индексы и индексы неожиданности. Для каждого метода, семь из девяти студентов отвечали на 20 неопределенных величин. Как показано в Таблице 1, группы, дающие вероятностные ответы и ответы шансов, показали лучшие индексы неожиданности, чем испытуемые, использующие метод квантиля. Неясно, появляется ли это превосходство из-за информации, сообщенной значениями, которые выбрал экспериментатор. Способ ответа на протоколы шансов не функционировал должным образом.

Шефер и Борчердинг (Schaefer и Borcharding, 1973) попросили 22 студентов оценить 18 пропорций, сделанных группой, в каждом из четырех испытаний. Каждый испытуемый использовал два метода оценки: (a) метод квантиля (0.01, 0.125, 0.25, 0.5, 0.75, 0.875, 0.99), и (b) метод гипотетических выборок. В последнем методе, оценивающий задает размер, n , и число успехов, r , гипотетической выборки, которая лучше всего отражает знание

экспертов неопределенной величины (то есть, я чувствую себя столь же уверенным относительно истинного значения пропорции, как я чувствовал бы себя, если бы я наблюдал выборку из n событий с r успехами). Большие значения n отражают большую уверенность относительно истинного значения пропорции. Отношение r/n отражает среднее значение функции плотности вероятности. Испытуемые испытывали большие трудности в этом методе, несмотря на инструкции, которые включили примеры бета распределений, лежащих в основе этого метода. После каждого испытания, испытуемым предлагали обширную обратную связь, с акцентом на их собственной калибровке и калибровке группы. Результаты первого и последнего испытания показаны в табл. 1. Улучшение было обнаружено для обоих методов. Результаты метода гипотетических выборок, в начале давал результаты хуже (50% неожиданностей и только 16% в межквартильном диапазоне), а в конце лучше (6% неожиданностей и 48% в межквартильном диапазоне), чем метод квантиля.

Барклей и Петерсон (Barclay и Peterson, 1973) сравнили метод тертилей (то есть, квантили 0.33 и 0.67) с “точечным” методом, в котором оценивающего просят дать модальное значение неопределенной величины, а затем два значения, одно выше и одно ниже моды, каждое из которых может произойти также вероятно, как и модальное значение (то есть, точки, для которых функция плотности вероятности является наполовину настолько же высока, как и в моде). При использовании 10 вопросов как неопределенных величин и 70 студентов Факультета внешней разведки в межгрупповом проекте, для метода тертилей они обнаружили, что 29% (а не 33%) истинных ответов попадают в центральный интервал. Для точечного метода, только 39% попадают между двумя равновероятными точками, принимая во внимание, что, для большинства распределений, приблизительно 75% плотности попадает между этими точками.

Питц (Pitz, 1974) сообщил о нескольких результатах, использующих метод тертилей. Для 19 испытуемых, экспертов населения 23 стран, он обнаружил только 16% истинных значений, попадающих в центральную треть распределений. В другом эксперименте он изменил вопросы согласно глубине и богатству знаний, которыми, по его предположению, обладали испытуемые. Для населения стран (низкое знание) он обнаружил 23% истинных значений в центральной трети; для высот известных зданий (посредственное знание) 27%; а для возрастов известных людей (высокое знание) 47%, причем последнее означение было намного больше ожидаемых 33%. В другом исследовании, он попросил 6 испытуемых оценить тертили и несколькими днями позже, сделать выбор среди ставок, основанных на их собственных значениях тертилей. Он обнаружил сильное предпочтение ставкам, из центральной области, как раз противоположное к чему должно было привести плотное расположение интервалов.

Испытуемые Ларсона и Ринана (Larson и Reenan, 1979) сначала угадывали истинный ответ (то есть, моду) и затем еще два значения, которые определяли интервал, в пределах которого, они были “вполне уверены”, находился пра-

вильный ответ. Сорок два процента истинных значений, лежали вне этой области. Обратите внимание, насколько похож этот индекс неожиданности на индексы испытуемых Альперта и Рейффы, которым дали вербальные формулировки “минимум / максимум” (47%) и “удивительно высоко / низко” (38%).

Реальные задачи с экспертами

Претт (1975) попросил некоторого эксперта дать прогноз посещаемости для 175 кинофильмов или двойных сеансов, показанных в двух местных кинотеатрах в течение более, чем одного года. Эксперт оценил медиану, квартили “удивительно высокие” и “удивительно низкие” значения. Как показано в табл. 1, межквартильный диапазон имел тенденцию быть слишком маленьким. Даже притом, что эксперт получал обратную связь исхода в течение эксперимента, единственное свидетельство улучшения калибровки через какое-то время пришло в первые несколько дней.

Три эксперимента использовали синоптиков в качестве испытуемых. В двух экспериментах Мерфи и Винклер (Murphy и Winkler, 1974, 1977b) просил синоптиков давать пять квантилей (0.125, 0.25, 0.5, 0.75, 0.875) для высокой температуры завтра. Результаты, показанные в табл. 1, указывают превосходную калибровку. Эти испытуемые имели меньшее количество случаев неожиданности в экстремуме распределения 25% чем, большинство испытуемых Альперта и Рейффы на экстремуме 2%! Мерфи и Винклер обнаружили, что пять испытуемых в двух экспериментах, которые использовали метод квантиля, были лучше калиброваны, чем четыре других испытуемых, которые использовали метод фиксированного диапазона (fixed-width method). Для метода фиксированного диапазона, синоптики сначала оценили температуру медианы (то есть, высокой температуры, для которой, как они полагали, существует вероятность 0.5, того, что она будет превышена). Затем они задали вероятность, что температура будет падать с интервалами 5°F и 9°F сосредоточенными в медиане. Эти синоптики были самонадеянные; вероятность, связанная с температурой, попадающей в интервал, имела тенденцию быть слишком большой. Превосходство метода квантиля над методом фиксированного диапазона в отличие от вывода Сивера, фон Винтерфельдта, и Эдвардза, о предпочтительности методов фиксированных значений, возможно, обусловлено тем, что установленные интервалы, используемые Мерфи и Винклером (5°F и 9°F) были неинформативны.

Штаел фон Хольстен (Stael von Holstein, 1971a) использовал три задачи фиксированных оценок: (а) средняя температура на завтра и на следующий день (с разделением полного диапазона ответа на восемь категорий), (б) средняя температура через 4 и 5 дней (восемь категорий), и (с) общая количество осадков в следующие 5 дней (четыре категории). От каждого набора ответов (четыре или восемь вероятностей, в сумме дающие 1.0) он оценил основную совокупную функцию плотности. Затем он объединил 1.269 функций, данные 28 участниками. Исходя из групповой совокупной функции плотнос-

ти, показанной в его статье, мы оценили индексы неожиданности и межквартильные индексы (см. табл. 1). В отличие от других синоптиков, эти испытуемые были весьма плохо калиброваны, возможно, потому что задачи были менее знакомы.

Резюме о калибровке с неопределенными величинами

подавляющее большинство свидетельств из исследования, использующего квантили для оценки неопределенных величин, говорит о том, что распределения вероятности имеют тенденцию быть слишком плотными. Оценка крайних квантилей особенно склонна смещаться. Обучение несколько улучшает калибровку. Результаты экспертов иногда высоки (Murphy и Winkler, 1974, 1977b), иногда нет (Pratt, 1975; Stael von Holstein, 1971a). Имеется некоторое свидетельство, что трудность связана с калибровкой для непрерывных предположений. Питц, Ларсон и Ринан (Pitz, 1974 и Larson и Reenan, 1979) обнаружили это влияние, Пикхарт и Воллейс (Pickhardt и Wallace, 1974) обнаружили, что 1-дневные задержки приводили к меньшему количеству неожиданностей, чем 3-дневные задержки в их игре моделирования. Несколько исследований (например, Barclay и Peterson, 1973; Murphy и Winkler, 1974) сообщили о корреляции между распространением оцененного распределения и абсолютной разностью между оцененной медианой и истинным ответом, указывая, что испытуемые частично чувствительны к тому, насколько, сколько они знают или не знают. Эта находка проводит параллель с корреляцией между процентом правильных ответов и средним ответом с дискретными предложениями.

Обсуждение

Зачем нужна хорошая калибровка?

Почему оценивающий вероятность должен волноваться относительно того, хорошо ли он калибровался? Фон Винтерфельдт и Эдвардз (1973) показали, что в большинстве жизненных проблем принятия решения с непрерывными выборами решения (например, вложите X долларов) довольно серьезные ошибки оценки имеют относительно небольшое значение для ожидаемого результата. Однако некоторые соображения приводят доводы против этой точки зрения. Во-первых, в двух-альтернативной ситуации, функция вознаграждения может быть весьма крута в критической области. Предположим, что ваш доктор должен оценить вероятность, что Вы имеете состояние А, и должны получить лечение А, против того, что у Вас состояние В и лечение В. Предположим, что ситуация такова, что лечение А является лучше, если вероятность, что Вы имеете состояние А большая или равна 0.4; иначе, лучше лечение В. Если доктор оценивает вероятность, что Вы имеете А как $p(A) = 0.45$, но плохо калиброван, так, что соответствующая вероят-

ность 0.25, тогда доктор использовал бы скорее лечение А, чем лечение В, и Вы будете терять много ожидаемой полезности. Решение функции полезности реальной жизни именно этого типа показаны Фрайбеком (Fryback, 1974).

Кроме того, когда вознаграждения очень большие, когда ошибки очень большие, или когда эти ошибки сложные, ожидаемые потери кажутся большими. Например, в исследовании радиоактивной безопасности (Американская Комиссия по Атомному Регулированию, 1975) “на каждом уровне анализа нормальное логарифмическое распределение данных пропорции неудачи было принято с определенными процентильными пределами 5 и 95” (Weatherwax, 1975, с. 31). Исследование, приведенное здесь, показывает, что распределения, построенные из оценок квантилей 0.05 и 0.95 могут быть чрезвычайно смещены. Если такие оценки сделаны на нескольких уровнях анализа, с каждым оцененным распределением, являющимся слишком узким, ошибки не будут отменять друг друга, а будут дополнять. И потому что цена сбоя на атомной электростанции велика, ожидаемая потеря от таких ошибок могла быть огромна.

Если хорошая калибровка важна, как ее достичь? Кокс (Cox, 1958) считал, что можно внешне рекалибровать оценки людей, подгоняя модель к набору оценок для вопросов с известными ответами. Тогда модель используется, чтобы исправлять или регулировать ответы, данные экспертом. Технические трудности, возникающие перед внешней перекалибровкой существенны. При выявлении оценок, которые будут смоделированы, нужно быть осторожными, чтобы не дать оценивающему больше обратной связи, чем они обычно получают, из страха того, что они изменят свою калибровку, когда она измеряется. Как указал Сэвадж (Savage, 1971), “Вы могли бы обнаружить на опыте, что ваш эксперт в некотором отношении оптимистичен или пессимистичен, и поэтому смягчать его суждения. Если он заподозрит Вас в этом, однако, и Вы, и он вполне можете приблизиться к гибели” (с. 796). Так как исследование показало, что тип наблюдаемого расбаланса калибровки зависит от уровня трудности задачи, возможно, пришлось бы также полагать, что будущее будет соответствовать трудности событий, используемых для перекалибровки.

Теоретические возражения внешней перекалибровке могут быть даже более серьезны, чем практические. Числа, произведенные процессом перекалибровки не будут следовать аксиомам теории вероятности (например, числа, связанные со взаимно исключающими и исчерпывающими событиями будут не всегда в сумме давать единицу, и при этом вообще не будет истинно что $P(A) \cdot P(B) = P(A, B)$ для независимых событий); следовательно, эти новые числа не могут называться вероятностями.

Более плодотворный подход состоял бы в том, чтобы обучить экспертов самокалибровке. При каких условиях можно ожидать, что эксперты достигнут этой цели? Нельзя ожидать, что эксперты будут хорошо калиброваны, если явная или неявная награда для оценок не мотивирует их быть честными в их оценках. Как крайний пример можно рассмотреть оценивающего, нахо-

дящегося под угрозой казни, если произойдет событие, вероятность которого была оценена в < 0.25 , который будет иметь серьезное основание не быть хорошо калиброванным с оценками 0.20 . Хотя этот пример кажется абсурдным, более тонкие давления такие как “не выглядеть глупо” или “произвести впечатление на босса” могли бы также обеспечивать сильные стимулы для плохой калибровки. Любое вознаграждение либо за размышления, либо за опровержение могло бы также смещать оценки.

При получении обратной связи исхода после каждой оценки – лучшее условие для успешного обучения. Дейвид (Dawid) показал, что при таких условиях эксперты, которые являются честными и последовательными субъективистами, будут хорошо калиброваны независимо от взаимозависимости среди оцениваемых вопросов. Напротив, Кадан (Kadane, 1980) показал, что в отсутствии обратной связи исхода “испытание за испытанием”, честный, последовательный субъективист будет, как ожидается, хорошо калиброван, только тогда, когда все оцениваемые вопросы независимы. Эта теорема накладывает сильные ограничения на ситуации, при которых было бы разумно ожидать, что эксперты будут учиться быть хорошо калиброванными. Даже если процесс обучения мог проводиться с использованием только событий, которые, как считали эксперты, независимы, может иметься серьезное основание, чтобы сомневаться относительно независимости задач реальной жизни, к которым эксперты применили бы их обучение. Важные будущие события могут быть взаимозависимыми либо потому что они находятся под влиянием общей основной причины, либо потому что оценивающий рассматривает их как идущие из основного запаса знаний. В таких обстоятельствах, можно не хотеть или не ожидать хорошей калибровки.

Возможность, что предубеждения людей изменяются как функция трудности задачи, препятствует обучению калибровке при отсутствии немедленной обратной связи исхода. Уровень трудности будущих задач может быть невозможно предсказывать, таким образом, делая обучение неэффективным.

Калибровка как когнитивная психология

Эксперименты по калибровке могут использоваться, чтобы понять, как люди думают. Даже если непосредственное практическое значение каждого исследования ограничено, оно может все-таки обеспечить понимание того, как люди развивают и выражают чувства определенности и неопределенности. Однако, поразительный аспект литературы, рассмотренный здесь – ее “сухой эмпиризм”. Психологическая теория часто отсутствует, либо как побуждение для исследования, либо как объяснение результатов.

Не все авторы избежали теоретизирования. Словик (Slovic, 1972a) и Тверски и Канеман (Tversky и Kahneman, 1974, 1) считали, что, в результате ограниченных способностей обработки информации, люди принимают упрощенные правила или эвристики. Хотя в целом полезные, эти эвристики

могут приводить к серьезным и систематическим ошибкам. Например, тенденция людей, давать плотные распределения при оценке неопределенных величин могла отражать эвристику, называемую “привязка и регулирование.” Когда людей спрашивают относительно неопределенной величины, они естественно думают сначала о точечной оценке, такой как медиана. Это значение тогда служит как привязка. Чтобы давать 25-ый или 75-ый процентиль, человек регулируется вниз или вверх от значения привязки. Но закрепитель имеет такое доминирующее влияние, что регулирование является недостаточным; следовательно, квантили находятся слишком близко вместе, выдавая самонадеянность.

Питц (Pitz, 1974) также допустил, что способность людей в обработке информации и вместимость рабочей памяти ограничены. Он предполагал, что люди занимаются сложными проблемами последовательно, прорабатывая часть за частью. Чтобы уменьшить когнитивное напряжение, люди игнорируют неопределенность в своих решениях ранних частей проблемы, чтобы уменьшить сложность вычислений в более поздних частях. Это может привести также к слишком плотным распределениям и самонадеянности. Питц также предполагал, что один путь, которым люди оценивают свою собственную неуверенность, – это увидеть, сколько существует различных путей, которыми они могут достигнуть ответа, то есть сколько различных последовательных решений они могут построить. Если найдено много путей, люди признают собственную неопределенность; если найдено мало, они не признают. Чем богаче база знаний, из которых можно строить альтернативные структуры, тем меньше тенденции к самонадеянности.

Филлипс и Райт (1977) представили последовательную модель с тремя стадиями. Их модель разделяет людей, которые естественно склонны думать о неопределенности вероятностным способом, от тех, которые выбирают из двух альтернатив. Их работа о культурных и индивидуальных различиях (Wright и Phillips, 1976, Wright и другие, 1978) была попыткой, с частичным успехом, идентифицировать различные когнитивные стили в обработке этого типа информации.

Кориат и другие (Koriat, 1980) также принимали подход обработки информации. Они обсуждали три стадии оценки вероятностей. Сначала человек ищет в памяти соответствующее свидетельство. Затем он оценивает это свидетельство, чтобы придти к чувству определенности или сомнения. Наконец, он переводит чувство определенности в число. Манипуляции, используемые Кориатом и другими, были предназначены, чтобы изменить первые две стадии, вынуждая людей искать и обращать внимание на противоречащее свидетельство, понижая свою уверенность.

Модель Феррелла и МакГои (Ferrell & McGoeu, 1980), с другой стороны, имеет дело полностью с третьей стадией, переводом чувства определенности в числовые ответы. Допуская то, что без обратной связи люди не могут изменить свои стратегии перевода по мере того, как либо трудность вопроса, либо основное значение событий изменяется, модель обеспечивает силь-

ные прогнозы, которые получили подтверждение от данных калибровки.

Начинают появляться теории структуры и процесса оценки вероятности; и мы надеемся, что дальнейшее развитие таких теорий будет служить, чтобы объединить это довольно специализированное поле деятельности в более широкую область когнитивной психологии.

23. Обреченным необходимо изучать прошлое: эвристики и предубеждения в ретроспективе*

Барух Фишхофф

Бенсон (Benson, 1972) идентифицировал четыре причины для изучения прошлого: развлекать, создавать групповую (или национальную) идентичность, показать степень человеческих возможностей и развивать систематическое знание относительно нашего мира, знание, которое может, в конечном счете, улучшить нашу способность прогнозировать и управлять. На сознательном уровне, по крайней мере, мы, ученые-бихевиористы, ограничиваемся последним мотивом. Преследуя его, мы исследуем случаи, оцениваем программы и делаем литературные обзоры. Мы даже проводим эксперименты, создавая искусственные истории, на которых мы можем проверить свои предположения.

Три основных вопроса, по-видимому, возникают в наших размышлениях о прошлом: (а) существуют ли модели, на которые мы можем опираться, чтобы в будущем стать мудрее? (в) существуют ли случаи непредусмотрительности, в которых мы можем увидеть ошибки, которых следует избегать; (с) мы действительно обречены повторять прошлое, если мы не изучаем его; то есть действительно ли мы приобретаем опыт, оглядываясь назад.

Неважно, какой вопрос мы задаем, предполагается, что прошлое с готовностью подскажет ответы. С точки зрения ретроспективы и дальновидности, последнее оказывается неприятной перспективой. Можно объяснить и понять любое событие прошлого, если применить соответствующее усилие. Прогнозирование считается более сложным. Данное эссе исследует это предположение, рассматривая некоторые первоначальные попытки выявить прошлое. Возможно, наиболее общее заключение – то, что мы должны уделять больше внимания прошлому, если мы пытаемся выявить его тайны. И

* Это - пересмотренная версия книги "Обреченным необходимо изучать прошлое: размышления об исторических оценках," в книге под редакцией R. A. Shweder и D. W. Fiske. Новые направления в методологии науки поведения: ошибочные суждения в исследовании поведения. Сан-Франциско: Jossey-Bass, 1980. Переиздано в соответствии с разрешением.

хотя прошлое развлекает, облагораживает и легко расширяется, оно просвещает только с помощью тонких уговоров.

В поисках мудрости

Хотя прошлое никогда точно не повторяется, в нем часто можно обнаружить повторяющиеся элементы. Люди принимают одинаковые решения, решают одинаковые задачи и терпят одинаковые неудачи достаточно часто, чтобы ученые-бихевиористы могли полагать, что они могут обнаружить повторяющиеся образцы поведения. Такое убеждение заставляет психометристов изучать диагностические тайны клинических психологов-асов, клиницистов — искать корреляты ненормального поведения, брокеров — замечать признаки роста цен, а диктаторов — обдумывать революционные ситуации. Их поиск обычно имеет логику, параллельную логике многократного регресса или корреляции. Собран набор соответствующих случаев, и каждый член характеризуется с помощью разнообразных измерений. В получившейся матрице ищут существенные связи, которые могли бы помогать нам в прогнозировании будущего....

Формальное моделирование

Ежедневный бланк бегов (Daily Racing), например, предлагает серьезному игроку приблизительно 100 отрезков информации относительно каждой лошади в любом забеге. Игрок, склонный к обработке данных, мог бы ввести в память компьютера содержание одного издания и пробовать получить формулу, предсказывающую скорость как оцененную сумму множества в различных измерениях. Например:

$$\tilde{y} = b_1x_1 + b_2x_2 + b_3x_3 \quad (1)$$

где \tilde{y} — наша оценка скорости лошади, x_1 , является процентом ее побед в предыдущих гонках, x_2 — процент побед в гонках ее жокея, и x_3 — вес, который ей предстоит нести в данной гонке. Предположив, что используются стандартизированные записи результатов¹, оценки (b_i), отражают важность различных факторов. Если $b_1 = 2b_2$, то данное изменение в проценте побед лошади воздействует на наше прогнозирование скорости вдвое больше, чем эквивалентное изменение в проценте побед жокея, потому что прошлые действия оказались вдвое более чувствительными к x_1 , чем к x_2 .

¹ Чтобы стандартизировать значения определенной переменной, нужно вычесть среднее из каждого значения и затем разделить на стандартное отклонение. Результат — это набор очков со средним, равным 0 и стандартным отклонением, равным 1.

Звучит просто, но здесь кроится тысяча ловушек. Одна из них появляется, когда прогнозирующие (x_i), являются коррелированными, что может происходить (и фактически происходит), когда из выигрыша лошадей следует выигрыш жокея (или наоборот). В таких случаях мультиколлинеарности, каждая переменная имеет некоторую независимую способность объяснять прошлое, а у двух есть некоторая общая особенность. Когда оценки определены, эта общая объяснительная способность, будет так или иначе разделена между этими двумя переменными. Как правило, это разделение порождает (b_i), не поддающееся толкованию с любой степенью точности. Таким образом, уравнение регрессии нельзя рассматривать как теорию скачек, показывая важность различных факторов.

Более скромная теоретическая цель состояла бы в том, чтобы определить, какие факторы важны, а какие – нет, на основе того, насколько каждый из них помогает нам понять y . Здесь применяется логика пошаговой регрессии, в уравнение добавляются дополнительные переменные, пока они что-либо добавляют к его прогнозирующей (или объясняющей) способности. Все же даже эта минималистская стратегия может привести к путанице мультиколлинеарности. Если отражения некоторого фактора (например, различные аспекты воспитания) включены, то их общая объяснительная способность может быть разделена на такие маленькие части, что ни один аспект не внесет “значимого” вклада.

Конечно, эти нюансы могут мало интересовать игроков на скачках, пока формула работает достаточно хорошо, чтобы приносить им выигрыш. Мы же, ученые, хотим не только эффективных методик, но и мудрости. Нам тяжело не интерпретировать оценки. Регрессивные процедуры не только выражают, но и порождают понимание (или, по крайней мере, результаты) механистическим, повторяющимся образом. Не удивительно, что им упорно следовали, несмотря на их ограничения. Один из лучших документально подтвержденных примеров был обнаружен в исследовании оценивания в клинике. Оценивающим был рентгенолог, который сортирует рентгеновские снимки опухолей на “доброкачественные” и “злокачественные”, кадровый работник, который отбирает подходящих кандидатов из списка, или сотрудник кризисной службы, который решает, звонки каких людей, грозящих покончить с собой, серьезны. В каждом из этих примеров диагноз включает принятие решения на основе некоторых признаков. Когда, как в этих примерах, принятие решений повторяется, и все случаи могут быть охарактеризованы тем же набором признаков, можно статистически смоделировать стратегию принятия решений оценивающего. Человек собирает набор случаев, для которых эксперт делает суммарную оценку (например, доброкачественное, серьезное), а затем выводит уравнение регрессии, такое как Уравнение 1, оценки которого показывают важность, которую оценивающий придал каждому признаку.

В течение двух десятилетий изучение такой стратегии постоянно приводило к двум беспокоящим выводам: (а) простые линейные модели, исполь-

зующие оцененную сумму признаков, отлично прогнозируют решения испытуемых, хотя (b) оценивающие заявили, что они использовали гораздо более сложные стратегии (Goldberg, 1969, 1970; Slovic & Lichtenstein, 1971). Общепринятая форма сложности называется конфигуральной оценкой, в которой диагностическое значение признака зависит от значения других признаков (например “интонация его голоса не кажется мне характерной для человека, решившего покончить жизнь самоубийством, если звонок не происходит рано утром”).

Две причины для конфликта между измеренным и сообщенным суждением возникли из последующего исследования, и каждая обладает негативными последствиями относительно полезности регрессивного моделирования для “черпания” мудрости решений прошлого. Одной из них было растущее понимание того, что комбинирование огромного количества информации в голове, как того требуют такие формулы, превышает вычислительную способность любого человека, кроме человека, проявляющего незаурядные способности в узкой области знаний. Оценивающий, который пытается применить сложную стратегию, просто не сможет сделать этого последовательно. Действительно, очень трудно выучить и использовать даже неконфигуральное правило принятия решений с оцененной суммой, когда существует много признаков или необычных отношений между признаками и прогнозируемыми переменными (Slovic, 1974).

Вторая вещь, которая была осознана с помощью исследования случаев из клинической практики (см. гл. 28) это то, что простые линейные модели дают очень сильные прогнозы. Простая субстантивная теория, показывающая, на какие переменные люди обращают внимание при принятии решений, – это, может быть, все, что нужно человеку для создания верного прогноза их поведения. Если некоторые признаки говорят в пользу некоторого решения, а другие – против него, то простое сопоставление количества одобряющих и опровергающих признаков даст возможность точного прогноза поведения человека. Как бы там ни было, результатом будет более скромная теория, чем теория, которую человек может вывести путем яркого моделирования регрессии (Fischhoff, Goitein & Shapira, в печати). Таким образом, в то время как кажется, что мы вот-вот пойдем прошлое, наши стандартные статистические процедуры не всегда говорят нам то, что мы хотим узнать. И если неверно их использовать, они могут завести нас не в ту сторону, сделав нас еще менее мудрыми, чем мы были в самом начале. Мы хотим охватить крайне сложные теории во всей их полноте, не понимая, что их сила исходит из очень простых основополагающих понятий, а не из понимания сущности прошлого.

Изучение недалёковидности

В центре внимания неудача

Поиск мудрости в исторических событиях требует веры – вера в существование повторения, ожидающего, чтобы его обнаружили. Поиск мудрости в поведении исторических личностей требует несколько другого рода веры – уверенности, что наши предшественники знали вещи, которых мы не знаем. Первая из этих вер основана на философии; она отличает тех, кто рассматривает историю как социальную науку, а не идеографическое исследование неповторимых событий. Вторая вера основана на милосердии и скромности. Она отличает тех, кто надеется видеть дальше, стоя на плечах у предшественников от тех, кто удовлетворен тем, что стоит на собственных ногах. Афоризмов как – “те, кто не изучают прошлое, обречены повторить его”, – наличие таких предполагает, что вера в мудрость наших предшественников относительно редка.

Активное исследование недалёковидности, конечно, имеет свои преимущества. Людям, у которых не все получается, приходится много объяснять, но такие объяснения очень важны для получения знаний из опыта. Осознав, почему дела пошли не так, как надо, мы надеемся заставить их улучшиться в будущем. Поиск причин неудач, чтобы объяснить их совсем не труден. Детектив, журналист и историк все вовлечены в хаос. Поездка на машине в магазин без несчастного случая, или правление без войн, депрессий, или землетрясений кажется им небогатыми на события.

Хотя концентрация на неудаче имеет приемлемые цели, она, вероятно, введет нас в заблуждение, создав искаженное представление о распространенности неудачи. Кажущаяся вероятность событий частично определена легкостью, с которой, они приходят на ум и вспоминаются (Tversky и Kahneman, 1973,11). Часто повторяющиеся неудачи должны, поэтому, непропорционально увеличить воспринятую частоту в прошлом (и возможно будущем).

Это, также, вероятно продвинет несбалансированную оценку поведения наших предшественников. Охотника до сенсаций, живущего в каждом из нас, тянет к историям благосостоятельных жуликов или “сверх-регулируемую” некоторых возможностей окружающей среды, (как например, печально известный стандарт на проект сиденья унитаза для производственных помещений, разработанный Министерством Здравоохранения и Охраны труда. Мы склонны забывать, тем не менее, что любая ошибочная, но не дьявольская, система принятия решения производит ошибки обоих видов. Для каждого мошенника, получающего незаслуженные выгоды, существует одна или несколько жертв, которым некая несовершенная система отказала в правах. Фактически, два ошибочных определения связаны отчасти интуитивным образом, зависящим от точности оценки и всех доступных ресурсов, например, процент бедноты, имеющей права, или риски, которое

можно принимать (Einhorn, 1978). Перед тем, как начать критиковать систему обеспечения за то, что она позволяет существование нескольких мошенников, мы должны рассмотреть, много ли существует страшных историй этого типа, учитывая отношение ошибок исполнения к ошибкам неисполнения.

Вообще, имеется большая вероятность того, что человек будет введен в заблуждение, если мы исследуем в изоляции решения, которые “вырабатываются” только на процентном основании.

В чем же проблема?

Существуют другие контексты, в которых ошибки в малом могут выглядеть другими, когда рассматривается более широкий контекст. Например, нас учат, что научные теории должны опровергаться, если представлено некое противоречащее им доказательство. В результате, мы быстро осуждаем ученых, которые упорствуют в своих теориях, несмотря на то, что их неправильность была доказана. Кун (Kuhn, 1962), однако, утверждает, что такая локальная недалновидность может быть совместимой с более глобальной мудростью в поиске научного знания. Другие (например, Feyerabend, 1975; Lakatos, 1970), фактически, превознесли роль дисциплинированной анархии в росте понимания и сомневались относительно возможности появления мудрости из организованной приверженности к любому одобренному методу исследования. Они доказывают, что нежелание принимать во внимание противоречащие данные или отказаться от очевидного опровергнутых теорий часто необходимо для научного прогресса.

Штраф в 125 миллионов долларов, наложенный на автомобильную Компанию “Форд” в случае с моделью “Пинто” (Pinto), сделал решение компании сэкономить несколько долларов на дизайне топливного бака автомобиля очень недалновидным. Все же в чисто экономических терминах, гарантируемая экономия, скажем, 15 долларов на каждом из 10 миллионов “Пинто” делает риск нескольких больших исков похожим на более разумную азартную игру. Так как в приговоре в этом хорошо освещенном в средствах массовой информации иске сумма была уменьшена до 6 миллионов долларов после апелляции, компания фактически выгадала в строгих экономических терминах, несмотря на то, что ситуация ухудшилась. В чем был промах компании, так это в рассмотрении одного контекста (количество автомобилей, на которых можно было сэкономить деньги), а не другого (неэкономические последствия ее решения). Компания, видимо, не поняла влияния, которое неблагоприятное общественное мнение оказало бы на образ “Форда” как производителя автомобилей, заботящегося о надежности, или на цены за используемый “Пинто” (хотя эту цену “Пинто” платил владелец, а не производитель).

Если осуждение – это название игры, ошибка – это ошибка. Все же, если вы заинтересованы в получении знаний из опыта других, важно определить то, какую проблему вы пытались решить. После осторожной проверки, мно-

гие очевидные ошибки, оказывается, представляют свежее решение неправильной проблемы. Например, если это может критиковаться вообще, компания Форд могла бы быть объявлена виновной в тактической мудрости и стратегической недалекости (или возможно в предпочтении здоровья социальному благосостоянию).

Это различие важно не только для оценки прошлого, но также и для знания, какие корректирующие меры должны быть приняты в будущем. Обычно, тактические ошибки легче исправить, чем стратегические недоразумения. Как только мы должным образом охарактеризовали ситуацию, мы можем завести “книгу”, в которую будем записывать обычную мудрость как накопленную с эмпирическим опытом, или, по крайней мере, формулы для оптимального объединения информации, имеющейся в нашем распоряжении (Hexter, 1971). Менеджеры в бейсболе либо знают, что лучше заменить отбивающего бегуном на первой игре и не заменять в игре на равных, либо применяют статистику, чтобы вычислить, как “получить проценты”. Эти руководящие принципы, однако, бесполезны или вводят в заблуждение, если реальная проблема, которая должна быть решена – это утверждение морали (бегун имеет шанс вести лигу на занятых базах) или помощь билетной кассе (болельщики должны видеть некоторую интригу). Исследования неожиданных нападений в международных отношениях показывают, что нации, которых заставляли врасплох, часто хорошо играли по своей книге, но неправильно определяли поле, на котором они играли (Ben Zvi, 1976; Lanir, 1978). В некотором смысле, они читали неправильную книгу; чем лучше они читали, тем быстрее они проигрывали.

Одна причина для трудности, поставленной стратегическими проблемами – то, что их нужно “продумывать” аналитически, не используя преимуществ совокупного (статистического) опыта. Второе ограничение – это то, что неправильные представления часто широко разделяются в пределах принимающей решения группы или сообщества. С человеком консультируются относительно решения только после того, как он закончил отвечать на вопросы в книге. Рекомендуемые советы для учреждений, заинтересованных в избегании неожиданностей: (а) основывайте несколько отдельных аналитических групп, чтобы обеспечить многократные, независимые взгляды на проблему, или (б) назначьте человека, который служил бы “адвокатом дьявола” для непопулярных точек зрения (Janis, 1972). Практически, первая стратегия может потерпеть неудачу, потому что распространенные неправильные представления делают группы очень похожими друг на друга, создавая скорее избыточность, чем плюрализм (Chan, 1979). Второй пункт терпит неудачу, потому что адвокаты либо склоняются под давлением группы, либо подвергаются гонениям, если они принимают свои непопулярные позиции всерьез, даже, когда эти “крайние” позиции радикально не противоречат мнениям группы.

Неспособность отличить тактические решения от стратегических могут также создать неоправданную иллюзию мудрости. Банки и страховые компании обычно считаются чрезвычайно рациональными и ловкими в процессах принятия решений. Все же более пристальное рассмотрение показывает, что эта

репутация происходит из их успеха в создании часто повторяющихся, тактических решений, в которых они почти не могут проигрывать. Ссуды под залог недвижимости и страхование жизни рассматриваются на основе консервативных интерпретаций статистических таблиц, приобретенных и отрегулированных с помощью массивного опыта проб и ошибок. Попытки этих учреждений принимать спекулятивные решения, требующие аналитических, стратегических решений, говорят о том, что они – не умнее, чем все мы. Коммерческие банки потеряли большие суммы денег в 60-ых, благодаря необдуманым инвестициям в тресты недвижимости; подобный небольшой процент их всех решений в 70-ых приковал экономику США к будущему полукредитоспособных стран Третьего мира, которым были сделаны большие займы (более \$ 60 миллиардов). (Хотя эта связь может быть оправдана пользой для всего человечества, которая вовсе не была проблемой, которую решали банки.) Медленный и беспорядочный ответ страховых компаний на изменения в экономике страхования несчастных случаев; и их почти случайные, неаналитические методы для многих необычных рисков должны заставить нас почувствовать себя не такими недалекими по сравнению с этими превозносимыми учреждениями.

Ретроспектива: Размышление над прошлым?

Если мы знаем то, что случилось и какую проблему человек пробовал решать, мы должны применить мудрость нашей собственной ретроспективы в объяснении и оценке его поведения. После более тщательной экспертизы, однако, преимущества знания того, как идут дела, могут быть переоценены (Fischhoff, 1975). В ретроспективе, люди последовательно преувеличивают то, что могло ожидать в будущем. Они не только имеют тенденцию рассматривать то, что случилось как неизбежное, но также и рассматривать это как оказавшееся “относительно неизбежным” прежде, чем это случалось. Люди полагают, что другие должны прогнозировать события намного лучше, чем это есть на самом деле. Они даже неправильно запоминают свои собственные прогнозы, чтобы преувеличивать в ретроспективе, что они знали заранее (Fischhoff и Beyth, 1975).

Как описано историком Георгием Флоровским (Georges Florovsky, 1969):

Тенденция к детерминизму, так или иначе, подразумевается в самом методе ретроспективы. Ретроспективно, мы, кажется, чувствуем логику событий, которые разворачиваются регулярным или линейным образом согласно распознаваемому образцу с предполагаемой внутренней потребностью. Так, чтобы мы получили впечатление, что это событие действительно не могло произойти по-другому, (стр. 369)

Подходящим названием для этой тенденции рассматривать исходы, о которых было сообщено, как об относительно неизбежных, могло бы быть “*пресмыкающийся детерминизм*”, в отличие от философского детерминизма, сознательного убеждения, что все, что происходит, должно случиться.

Одна из тенденций – сократить пропорцию исторических процессов, увеличивая скорость, с которой “неизбежные” изменения осуществляются (Fischer, 1970). Например, люди способны указать на момент, когда жители Латифундии были обречены, не понимая, что им потребовалось два с половиной столетия, чтобы исчезнуть. Другая тенденция состоит в том, чтобы помнить людей, более похожими на себя настоящих, чем фактически это было так (Yarrow, Campbell, и Burton, 1970). Третья тенденция может быть замечена в критическом анализе историографии идеологических корней нацизма Барраклау (Barraclough, 1972). Оглядываясь назад от времен Третьего Рейха, можно проследить его корни в работах многих авторов, из чьих произведений нельзя было вывести нацизм. Четвертая тенденция – это думать, что участники исторической ситуации полностью осознавали ее возможную важность (“Дорогой дневник, сегодня началась Столетняя война”, Fischer, 1970). Пятая тенденция – это миф критического эксперимента, недвусмысленно решающего конфликт между двумя теориями или устанавливая валидность одной. Фактически, “критический эксперимент считается критическим только десятилетия спустя. Теории не отменяются просто так, так как несколько аномалий всегда позволяют. Действительно, очень трудно опровергнуть программу исследования, которую поддерживают талантливые ученые с развитым воображением” (Lakatos, 1970, с. 157-158).

Неспособность игнорировать знание исхода имеет существенные выгоды. Весьма лестно считать, или заставлять других полагать, что мы знали бы все, что мы могли бы только знать со знанием исхода, то есть, что мы обладаем ретроспективным предвидением. В конечном счете, однако, обнаруженный пресмыкающийся детерминизм может серьезно повредить нашей способности оценивать прошлое или получать из него знания.

Рассмотрим людей принимающих решения, которые оказались неготовыми к некоторому повороту событий и которые пробуют увидеть, где они повели себя не так, как надо, воссоздавая состояние знания пред-исхода. Если ретроспективно событие кажется относительно вероятным, они не делают ничего, кроме как, будут ругать себя за то, что они не совершили действия, которое их знание, кажется, диктовало. Они могли бы добавить сожаление об ущербе, причиненном самим событием. Когда прошлое пересматривается наблюдателем, их неудача кажется проявлением некомпетентности, недалекости, или еще хуже.

В ситуациях, когда информация ограничена и неопределенна, случайные неожиданности и конечные неудачи неизбежны. Наказывать людей принимающих решения, которые допустили ошибку в подверженных ошибкам системах, не признавая этого и ничего не делая, чтобы улучшить систему, является несправедливым и пагубным. Согласно историку Роберте Вольштеттер (Roberta Wohlstetter, 1962), урок, который получен из неожиданного нападения в Перл Харборе – то, что мы должны “принять факт неопределенности и учиться жить с ним. Так как никакое волшебство не обеспечивает определенность, наши планы должны работать без нее” (с. 401).

Когда мы пытаемся понимать прошлые события, мы неявно проверяем гипотезы или правила, которые мы используем, чтобы интерпретировать и понимать мир вокруг нас. Если, в ретроспективе, мы систематически недооцениваем случаи неожиданности, которые были и есть у прошлого для нас, мы подвергаем эти гипотезы очень слабым тестам и, возможно, находим мало причин изменить их. Таким образом, само знание исхода, которое дает нам чувство того, что мы понимаем то, что прошлое хотело нам сообщить, может помешать нам действительно вынести из него какой-либо урок.

Защита от этого предубеждения требует некоторого понимания психологических процессов, задействованных в его создании. Кажется, что когда мы получаем знание исхода, мы немедленно понимаем его, интегрируя в то, что мы уже знаем о предмете. После такой переинтерпретации, сообщенный исход кажется более или менее неизбежным продуктом переинтерпретированной ситуации. "Понимание" того, что нам говорят о прошлом, в свою очередь, настолько естественно, что мы можем не знать, что знание исхода оказывало какое-либо влияние на нас. Даже если мы знаем, что влияние было, мы можем все еще точно не сознавать, каким оно было. В попытке восстановить наше ретроспективное состояние, мы остаемся закрепленными в нашей ретроспективе, оставляя рассматриваемый исход слишком вероятным.

В результате, просто предупреждение людей относительно опасностей предубеждения ретроспективы не имеет влияния (Fischhoff, 1977b). Более эффективная манипуляция должна вынудить человека приводить доводы против неизбежности сообщенных исходов, то есть попытаться убедить, что это могло бы получиться иначе. Вопрос о валидности причин, которые вам потребовались для объяснения его неизбежности, мог бы быть хорошим началом (Koriat, Lichtenstein, и Fischhoff, 1980; Slovic и Fischhoff, 1977). Так как даже этот необычный шаг, кажется не полностью адекватным, можно было далее пробовать проследить часть неопределенности, окружающей прошлые события в их первоначальной форме. Существуют ли расшифровки стенограммы командованию Перл Харбора до 7 часов утра 7 декабря? Существует ли записная книжка, показывающая акции, которые Вы рассмотрели до покупки акций Волтем Индастриз? Существуют ли дневники, показывающие взгляд Чемберлена на Гитлера в 1939? Интересным вариантом было намерение Дугласа Фримана не знать ничего о последующих событиях на любом отрезке времени при его работе над биографией Роберта Е. Ли. Хотя это и замечательно, эта стратегия все-таки требует некоторых предположений о распространенности знания относительно того, кто сдался в Аппоматоксе.

Взгляд на все сразу

Зачем вообще смотреть?

Изучение прошлого основано на убеждении, что, если мы смотрим, мы будем способны разглядеть некоторые поддающиеся интерпретации модели. Обширное исследование полагает, что это убеждение хорошо обосновано. Люди, кажется, имеют замечательную способность находить некоторый порядок или значение даже в случайно произведенных данных. Один из наиболее знакомых примеров – ошибка игрока в казино. Мы считаем, что при подбрасывании монеты, четыре последовательных “орла” будут сопровождаться “решкой” (Lindman и Edwards, 1961). Таким образом, в наших умах даже случайные процессы вынуждены иметь внутренние свойства. Канеман и Тверски (1972b, 3) предположили, что из 32 возможных последовательностей шести бинарных событий только 1 практически рассматривается как “случайная”.

Хотя ошибка игрока в казино обычно упоминается в контексте пикантных, но тривиальных примеров, она может также быть найдена в более серьезных попытках объяснить исторические события. Например, после отчетливой демонстрации того, что вакансии в Верховном Суде распределяются более или менее случайно (согласно процессу Пуассона), с вероятностью 0.39, по крайней мере, одной вакансии в год, Моррисон (Morrison, 1977) потребовал, чтобы

(Президент) Рузвельт объявил свой план созыва Суда в феврале 1937, вскоре после начала своего пятого года правления в Белом доме. 1937 был также годом, в котором он сделал свое первое назначение в Суд. То, что он имел эту возможность в 1937, не должно было бы показаться удивительным, потому что вероятность, что он в течение пяти лет не назначал бы одного или более судей, была всего лишь 0.08, или один к двенадцати. Другими словами, когда Рузвельт решил изменить Суд, создав дополнительные места, вероятность была уже *одиннадцать к одному в его пользу*, что он сможет назвать одного или более судей традиционным способом именно в этом году. (с.143-144)

Однако, если вакансии появляются наугад, то это рассуждение неправильно. Оно предполагает, что вероятностный процесс, создающий вакансии, подобно монете, имеет память и чувство справедливости, как будто он знает, что он переходит в пятый год правления президента Рузвельта и что он “должен” Ф.Д.Рузвельту вакансию. Однако, 1 января 1937 прошлые четыре года были историей, и вероятность, по крайней мере, одной вакансии в наступающем году была все еще 0.39 (Fischhoff, 1978).

Феллер (Feller, 1968) предложил следующий рассказ, вовлекающий даже более высокие ставки: жители Лондона в течение бомбардировок потратили значительные усилия на интерпретации зон немецкой бомбежки, развитие сложных теорий того, куда немцы целились (и когда снимать маскировку). Однако, когда Лондон был разделен на маленькие, смежные географические области, распределение частоты попаданий бомб в область было почти

точным приближением к пуассоновскому распределению. Кейтс (Kates, 1962) предполагает, что природные бедствия составили другую категорию важных событий, где (напуганные) обыватели, видят порядок, когда эксперты видят случайность.

Одна загадка устойчивости таких убеждений – неспособность продолжать использовать достаточно полные записи, чтобы противостоять ошибкам. Историки подтверждают, что роль отсутствующего свидетельства в облегчении объяснений с комментариями, такими, как “история Викторианской Эпохи никогда не будет написана. Мы знаем слишком много относительно ее. Так как невежество – первая необходимая вещь для историка – невежество, которое упрощает и разъясняет, которое выбирает и пренебрегает, со спокойным совершенством, недостижимым самым высоким искусством” (Strachey, 1918, предисловие).

Даже, если записи доступны и неизбежны, мы, кажется, имеем необычную способность объяснять или обеспечивать причинную интерпретацию для всего, что мы видим. Когда события обусловлены вероятностными процессами с неосязаемыми свойствами, случайное изменение не может даже казаться нам потенциальной гипотезой. Например, тот факт, что атлеты, наказываемые за плохое выступление, имеют тенденцию добиваться большего успеха в следующий раз, удовлетворяет нашим наивным теориям о поощрении и наказании. Это удобное объяснение ослепляет нас по отношению к вероятности того, что улучшение наоборот, происходит благодаря регрессу к среднему в выступлении атлетов (Furby, 1973; Kahneman и Tversky, 1973, 4).

Фама (Fama, 1965) действительно считал, что колебания цен на фондовой бирже лучшие поняты как отражение случайного процесса. Случайный процесс, однако, имеет даже более неинтуитивные свойства, чем двоичный процесс, с которым он формально связан (Carlsson, 1972). В результате, мы обнаруживаем, что рыночные аналитики имеют объяснение каждого изменения в цене, преследующее некоторую цель или нет. Некоторые объяснения, подобно показанным на рис.1, являются противоречивыми²; другие, кажется, отрицают возможность случайного компонента, например, полностью выдуманный фактор “техническое регулирование.”

Псевдовласть наших объяснений может быть проиллюстрирована по аналогии с анализом регресса. Учитывая набор событий и достаточно большой или богатый набор возможных объясняющих факторов, можно всегда получать постдикции или объяснения для любой желаемой степени плотности. В терминах регресса, расширяя набор независимых переменных можно всегда найти нескольких прогнозирующих с любой желаемой корреляцией с независимой переменной. Цена, которую приходится платить за чрезмерную совместимость,

² Один из моих любимых контрастов – это, когда рыночные цены повышаются после хороших экономических новостей, то это считается ответом на новости; если же они падают, это объясняется, тем, что хорошие новости уже были обесценены.

конечно, сокращение, неспособность полученного правила работать на новой выборке событий. Частота и сила методологических предупреждений против чрезмерной совместимости предполагают, что корреляционное применение избыточной мощности – это предубеждение, которое является весьма стойким даже к расширенному профессиональному обучению (для ссылок, см. Fischhoff и Slovic, 1980).

Теория чрезмерной совместимости подобна костюму, скроенному точно для одного человека в одной специфической позе, который не подойдет кому-либо еще или даже, тому же самому человеку в будущем или даже в настоящем, если у него происходит некоторое заметное изменение (например, если он делает выдох, чтобы показать живот). Историк, который построил гермети-



Рис. 1. Два примера признаков, используемых в выделении предшественников прошлых изменений в ценах акций: формирование сопротивления и формирование поддержки. Однако можно видеть, что до серьезных изменений, эти два образца были по существу идентичны. В этом свете, волнообразный отрезок ни прогнозирует, ни объясняет ничего в этих данных

ческий случай, объясняющий все доступные свидетельства, как могли бы победившие большевики оказаться в грустном положении, если бы СССР рассекретил документы, показывающие, что меньшевики были более серьезные противники, чем думали раньше. Цена, которую инвестиционные аналитики платят за чрезмерное совпадение – их неспособность в конце концов предсказать ничего лучше, чем средние рыночные числа (Dreman, 1979) – хотя циник мог бы сказать что они фактически выживают через поколение надежды (и комиссии).³

Чрезмерное совпадение происходит из-за капитализации на случайных колебаниях. Если измерение достаточно точное, два случая, отличающиеся одной переменной, также будут отличаться любой другой переменной, которую мы хотим назвать. В результате, можно вычислить отличную от нуля (фактически, в этом случае, совершенную) корреляцию между этими двумя переменными и получить “интересную” самостоятельную теорию. Процессы, аналогичные этому двух-размерному случаю, работают с любыми m наблюдениями в n -мерном пространстве, определенном нашим набором возможных объясняющих концепций.

В этих примерах, данные заданы и бесспорны, в то время как набор возможных объяснений относительно неограничен; человек ищет до тех пор, пока не находит объяснение, которое его удовлетворяет. Другая популярная форма капитализации на случае оставляет набор установленных объяснений (обычно для одного кандидата) и перебирает данные до тех пор, пока подтверждающее свидетельство не найдено. Хотя более грубая форма этой процедуры хорошо известна, другие формы являются более тонкими и даже несколько неоднозначными в своей характеристике. Например, Вы проводите эксперимент и не получаете ожидаемый результат. Думая об этом, Вы обращаете внимание на элемент вашей процедуры, который мог уменьшить влияние зависимой переменной. Вы исправляете это; снова нет результата, снова возникает возможная проблема. Наконец, Вы (или ваши испытуемые) получаете все, как надо, и ожидаемый эффект получен. Теперь, правильно ли выполнить ваше статистическое испытание на этой энной выборке (для которой оно значимо) или на всех выборках? Если бы Вы сделали правильный эксперимент сначала, вопрос даже не будет возникать. Или, как токсиколог, Вы “уверены”, что воздействие химического вещества X является вредным для здоровья, так что Вы сравниваете рабочих, которые работают и не работают с ним на некотором заводе на наличие рака мочевого пузыря, но все еще не получаете никакого результата. Так что Вы проверяете их на рак кишечника, эмфизему, головокружение, и так далее, пока Вы наконец не получаете существенное различие в наличии рака кожи. Это различие

³ Мой друг однажды проходил курс по чтению форм диаграмм в сфере маклерства. Каждое занятие включало обучение 10-12 новым признакам. Когда курс закончился, через пять занятий и 57 признаков, у преподавателя осталось еще много в запасе.

значимое? Конечно, способ проверить эти объяснения или теории – испытание с новыми выборками. Этот шаг, к сожалению, редко предпринимается и часто не возможен по техническим или этическим причинам (Tukey, 1977).

Связанные осложнения могут возникать даже с установленными теориями и наборами данных. Диаконис (Diaconis, 1978, с. 132) отмечает трудность оценки количества неожиданностей в результатах экстрасенсорного восприятия, даже в редких случаях, в которых они были получены в контролируемой обстановке, потому что определения нужного события продолжают изменяться. “Главный ключ к успеху Б.Д. было то, что он не определял заранее результат, который нужно рассматривать как неожиданный. Разногласия против *некоторого* совпадения существенно меньше, чем против любого заранее *определенного* совпадения”.⁴

Туфт и Сан (Tuft и Sun, 1975) обнаружили, что существование или отсутствие пределов зависит от творческого потенциала и гибкости, допускаемой в определении события (для какого офиса? на каких выборах? насколько хорошо? пределы, которые отсутствуют, должны быть включены?). Они, как полагают, существуют, потому что мы имеем удивительную способность обнаружить сигнал даже в полном шуме.

Мы уже достаточно увидели?

Учитывая, что мы почти уверены относительно обнаружения чего-то подающегося толкованию, когда мы смотрим на прошлое, наш следующий вопрос такой, “Поняли мы его?” Исследование ретроспективы, описанное ранее, предполагает, что мы не только быстро находим порядок, но также и чувствуем, что мы знали это все некоторым способом или были бы способны спрогнозировать результат, если нас вовремя попросили бы. Действительно, легкость, с которой мы обесцениваем информативность того, что нам говорят, делает удивительным то, что мы когда-либо задаем прошлому, или любому другому источнику, много вопросов. Эта тенденция усугубляется за счет (а) не понимания того, как мало мы знаем или того, что нам говорят, оставляя нас не сознающими то, какие вопросы мы должны задать в поиске неожиданных ответов (Fischhoff, Slovic и Lichtenstein, 1977, 1978) и (б) порождения далеко идущих заключений даже из малого количества ненадежных данных (Kahneman и Tversky, 1973, 4; Tversky и Kahneman, 1971, 2).

Любая склонность не смотреть дальше поощряется нормой сообщения истории, как хорошей истории, со всеми уместными подробностями, аккуратно объясненными и неопределенностью, окружающей событие до его завершения,

⁴ Диаконис продолжает, “Чтобы далее усложнить любой анализ, несколько таких неточно определенных экспериментов часто проводились одновременно, взаимодействуя друг с другом. Молодой исполнитель держал аудиторию в напряжении. Его часто полностью неверные предположения вообще расценивались скорее с симпатией, чем сомнением; и для большинства наблюдателей они, казалось, только подтверждали наличие сверхъестественных способностей у Б.Д.”

наряду с любым сметением, который автор, возможно, чувствовал (Gallie, 1964; Nowell-Smith, 1970). Только одна из тайн этого явления раскрыта Тауни (Tawney, 1961); “Историки приписывают появление неизбежности существующему порядку, выдвигая на первый план силы, которые одержали победу и отодвигая на задний план те, которые они проглотили” (с. 177).⁵

Хотя это и интуитивно привлекательная цель, построение последовательных рассказов подвергает читателя некоторым интересным предубеждениям. Законченный рассказ состоит из ряда независимых звеньев, каждое довольно хорошо определено. Правда рассказа зависит от правды звеньев. Вообще, чем большее количество звеньев появляется, и большее количество деталей находится в каждом звене, тем менее вероятно, что история должна быть правильной в ее полноте. Однако Словик, Фишхофф и Лихтенштейн (Slovic, Fischhoff и Lichtenstein, 1976) нашли, что добавление детали к описанию события может увеличивать его кажущуюся вероятность возникновения, очевидно, увеличив его тематическое единство. Бар-Хиллел (Bar-Hillel, 1973) обнаружила, что люди последовательно преувеличивают вероятность конъюнкции ряда вероятных событий. Например, ее испытуемые вообще предпочитали ситуацию, в которой они получают приз, если семь независимых событий, каждое с вероятностью 0.90 должны произойти, ситуации, в которой они получают тот же самый приз, если при подбрасывании монеты выпал “орел”. Вероятность сложного события меньше чем 0.50, принимая во внимание, что вероятность единственного события – 0.50. Другими словами, неопределенность, кажется, нарастает в медленном темпе.

Что случается, если последовательность включает одно или несколько слабых или маловероятных звеньев? Вероятность его самого слабого звена должна установить верхний предел вероятности полного рассказа. Последовательные суждения, однако, могут быть компенсационными, с последовательностью сильных звеньев “выравнивающих” бессвязность слабых связей. Этот эффект используется адвокатами, которые помещают самое слабое звено в их аргументах рядом с началом их выводов и заканчивают поток убедительных, неоспоримых аргументов.

Коулс (Coles, 1973) предоставил восхитительный пример полной последовательности истории, скрывающей неправдоподобность ее звеньев: наиболее серьезная попытка Фрейда в психоистории была составить биографию Леонардо да Винчи. В течение многих лет, Фрейд искал тайну, позволяющую понять Леонардо, чье детство и юность были в основном неизвестны. Наконец, он обнаружил ссылку Леонардо на воспоминание о стервятнике, касающемся его губ, в то время как он был в колыбели. Отмечая идентичность египетских иероглифов для “стервятника” и “матери” и другого свидетельства, Фрейд продолжал строить внушительный и последовательный анализ Лео-

⁵ Такие стратегии могут затрагивать дух, также как и разум, субъективно увеличивая силу и стабильность статуса-кво и сокращая его очевидную способность к изменению (Markovic, 1970).

нардо. При составлении окончательного издания работ Фрейда, однако, редактор обнаружил, что немецкий перевод воспоминания Леонардо (первоначально по-итальянски) который использовал Фрейд, был ошибочным, и что это был бумажный змей, а не стервятник, который гладил его губы. Несмотря на наличие ключа к разрушению анализа Фрейда, редакторы решили, что остающийся каркас был достаточно прочен, чтобы выстоять это. Как Хекстер заметил (Hexter, 1971, с. 59), “Частично, потому что написать плохую историю довольно легко, хорошие истории пишутся очень редко”.

Заключение

Какие общие уроки мы можем извлечь из изучения прошлого, за исключением факта, что понимание является более неуловимым, чем это может быть часто подтверждено?

Презентизм (приверженность настоящему)

Неизбежно, мы все пленники нашей существующей личной перспективы. Мы знаем вещи, которые наши предшественники не знали. Мы используем аналитические категории (например, феодализм, Столетняя война), которые являются значимыми только ретроспективно (E. A. R. Brown, 1974). Мы обладаем собственными точками зрения при интерпретации прошлого, чтобы доказать, что оно является неоднозначным, чтобы избежать наложения нашей идеологической перспективы (Degler, 1976). Историки действительно “используют новые фокусы на мертвых в каждом поколении” (Becker, 1935).

Не существует никакого доказанного противоядия для презентизма. Некоторые частичные средства могут быть обобщены из обсуждения того, как избежать предубеждения ретроспективы при пересмотре прошлого. Другие появляются в почти любом тексте, посвященном обучению историков. Возможно, наиболее общими выводами, являются, (а), знание нас самих и настоящего настолько хорошо, насколько возможно; “историк, который больше всего ощущает собственную ситуацию, также наиболее способен к ее трансцендированию” (Benedetto Croce, цитируемый Carr, 1961, с. 44); и (b) быть настолько благожелательными, насколько возможно, по отношению к нашим предшественникам; “историк – не судья, еще меньше вешающий судья” (Knowles, цитируемый Marwick, 1970, с. 101).

Педантичность

В дополнение к неизбежному заключению в рамках нашего времени, мы часто еще больше ограничиваем нашу перспективу, добровольно надевая шоры, которые сопровождают строгую приверженность единственному научному методу. Даже когда он используется рассудительно, ни один метод

не является адекватным для ответа на многие вопросы, которые мы задаем прошлому. Каждый сообщает нам что-либо и несколько вводит в заблуждение. Когда мы не знаем, как получить правильный ответ на вопрос, необходима альтернатива эпистемологии: используйте по возможности широкий диапазон методов или перспектив, каждый из которых позволяет вам избежать некоторых видов ошибок. Это означает своего рода междисциплинарное сотрудничество и отношение, отличное от того, с которым люди сталкиваются в большинстве попыток смешать два подхода. Совпадения или не совпадения, подобно психоистории, слишком часто предпринимаются защитниками, нечувствительными к ловушкам в их принятых областях (Fischhoff, в печати - b). Хекстер (Hexter, 1971, с. 110) описывает историков, вовлеченных в некоторые такие приключения как “крысы, прыгающие с борта интеллектуально тонущего корабля”.

Обсуждение

Возвращаясь к Бенсону (Benson, 1972), если мы хотим, чтобы прошлое служило будущему, мы не можем обращаться с ним в изоляции. Правила, которые мы используем, чтобы объяснить прошлое, должны также быть теми, которые мы используем, чтобы прогнозировать будущее. Мы должны накопить опыт, обращая внимание на все уместные испытания наших гипотез. Один аспект выполнения этого компилирует отчеты, которые могут быть подвергнуты систематическому статистическому анализу; второй – хранит след обсуждения, предшествующего нашим собственным решениям, понимая, что настоящее будет скоро прошлым и что хорошо сохраненное описание является лучшим средством предотвращения предубеждения ретроспективы; третий делает прогнозы, которые могут быть оценены; один тревожащий урок из ядерной катастрофы на острове Три Майлс – то, что не в полне ясно, что этот якобы диагностический случай показал нам относительно валидности Исследования атомной безопасности (Американская Комиссия по атомному урегулированию, 1975), которое попыталось оценить угрозы ядерной энергии: четвертый аспект получает лучшую идею относительно валидности наших собственных чувств достоверности, поскольку уверенность в существующем знании управляет нашей жадной новой информации и интерпретаций (Fischhoff, Slovic и Lichtenstein, 1977). Таким образом, мы должны структурировать нашу жизнь, чтобы облегчить обучение.

Неопределенность

В конце, тем не менее, не может быть никаких ответов на многие из вопросов, которые мы излагаем. Некоторые плохо сформулированы. На другие нельзя ответить с помощью существующих или возможных инструментов. Столько, сколько мы хотели бы знать “как профессионалы делают это”, не существует статистического способа смоделировать образ суждения экспер-

тов до желательной степени точности с реальными стимулами. Наши теории часто “настолько сложные, что никакая отдельная количественная работа не могла даже начать проверять их валидность” (O’Leary и другие, 1974, с. 228). Когда группы, которые мы хотим сравнить по одной переменной, также отличаются по другой, не существует логической процедуры для приравнивания их на этой мешающей переменной (Meehl, 1970). Когда мы попробовали применить много возможных объяснений к одному набору данных, не существует твердого способа узнать, сколько степеней свободы мы использовали, и как мы капитализировали случайности (Champbell, 1975). Когда мы используем множественные подходы, знание, которое они производят, никогда точно не конвергирует. В конце концов, можно попытаться применить философскую перспективу Тревельяна, что “несколько несовершенных подходов к истории – это лучше, чем ни одного” (по Марвик, 1970, стр. 57).

Часть VII
Многоступенчатая оценка

24. Оценка составных вероятностей в последовательном выборе*

Джон Коэн, Е. Я. Чесник и Д. Харан

Часто бывают ситуации, в которых успешный исход зависит от человека, делающего правильный выбор на каждой из нескольких более или менее независимых стадий. В качестве примера можно привести выбор транспорта на различных стадиях совершения поездки. Похожие затруднительные моменты встречаются в профессиональной, административной, политической, военной жизни и в коммуникационных сетях вообще. Временной порядок этих нескольких выборов – не всегда жизненно важный фактор. Кроме того, ситуация в целом может иметь стохастический характер, где вероятность правильного выбора может изменяться от стадии к стадии.

Мы опишем эксперимент, моделирующий этот общий тип ситуации, которая ясно требует, для успешного исхода, умножения вероятностей. Исследования предпочтения определения местонахождения цели в матрице $m \times n$, или в дисплее, разделенном на концентрические зоны, указывают, что ячейки в матрице или концентрических зонах субъективно не равновероятны как местоположения для цели (Cohen, Boyle и Chesnick, 1969). Испытуемые, кажется, не наугад определяют неизвестное местоположение цели. Их стратегия поиска скорее имеет характер “предсказания”, где экспериментатор скрыл цель, или расположил там, где они ожидали ее увидеть. Другие эксперименты (Cohen и Hansel, 1958) предполагают, что многие люди при столкновении с ситуацией, вовлекающей сложные вероятности, склонны складывать, а не умножать, возможности на различных стадиях.

Мы можем соответственно спрогнозировать, что в матрице $m \times n$, с неизвестной целью в каждом из m рядов, оценка испытуемого его вероятности угадать все цели будет правильно преувеличена, как оцененная сложной

* Эта глава первоначально появилась в *Nature*, 1971, 32, 414-416. Авторское право © 1971 Macmillan Journals, Ltd. Переиздано в соответствии с разрешением.

вероятностью, n^m , расположения цели. Наш эксперимент предназначен для проверки этого прогноза и объяснения явлений, которые возникнут в ситуации.

Таблица 1. Психологические вероятности (ψ), основанные на выборе в лотерее

Количество альтернатив в стадии	Количество стадий в массове					
	2	3	4	5	8	M
2	0.45	0.51	0.36	0.38	0.37	0.41
3	0.41	0.21	0.21	0.21	0.17	0.24
4	0.31	0.23	0.21	0.12	0.11	0.19
5	0.30	0.22	0.17	0.18	0.08	0.19
8	0.07	0.16	0.11	0.13	0.05	0.11
M	0.31	0.27	0.21	0.20	0.16	

Примечание: Пункты в таблице представляют собой средние величины десяти наблюдений.

Аппарат состоял из доски, на которой был набор из m рядов, каждый из которых содержал n гнезд, где m и n принимали значение 2, 3, 4, 5 или 8. Все гнезда были пустыми кроме одного в каждом ряду, в который помещался билет. Количество отдельных стадий поэтому было от 2 до 8, с вероятностью правильно угадать в каждой стадии в пределах от 2^{-1} до 2^{-3} .

Испытуемыми были 50 учащихся средней школы в возрасте 14-15 лет, «интеллект» которых был, по крайней мере, столь же хорошим как у среднего взрослого. Испытуемому объяснили, что для того, чтобы выиграть приз, он должен был угадать правильное местоположение билета в каждом ряду, и позволили делать только одно предположение на 1 ряд. Его задача состояла в том, чтобы приравнять то, что он считал вероятностью выигрыша приза, с одним из наборов лотереи.

Всего проводилось четырнадцать лотерей. Десять из них имели 100 билетов каждая, а вероятность вытянуть выигрышный билет, основанный на количестве выигрышных билетов в данной лотерее, было 0.01, 0.1, 0.2, 0.3..., 0.9. Оставшиеся четыре лотереи имели 500, 1.000, 5.000 и 10.000 билетов, с соответствующим шансом вытянуть счастливый билет 0.002, 0.001, 0.0002 и 0.0001.

Диапазон значений, принимаемых рядами, и диапазон значений гнезд учитывал двадцать пять различных ситуаций. Каждый из пятидесяти испытуемых был приписан наугад одной из пяти ситуаций посредством случайных латинских квадратов. Десять испытуемых, таким образом, оценивали $m \times n$ ситуаций каждый.

Количество информации, которая фактически требовалась, чтобы определять местонахождение цели – $m \log_2 n$ “бит”, где m - количество рядов, которые мы теперь назовем “стадиями”, и n количество гнезд, которые будут определяться как “альтернативы”. Вероятность вытянуть выигрышный билет в лотерее, выбранной испытуемым, мы будем рассматривать как косвенную оценку его психологической вероятности ψ – получения приза.

Результаты, представленные в Таблице 1, показывают что все значения ψ – являются оцененными слишком высоко, при оценке сложными вероятностями, p , таким образом, подтверждая наш прогноз. Величина переоценки обозначена отношениями ψ к p в табл. 2, что предполагает, что реализм оценок нарушает очевидная тенденция меньше принимать во внимание количество стадий, чем количество альтернатив в стадии, хотя дисперсионный анализ показывает, что относительно остаточной дисперсии, как дисперсия между стадиями, так и между альтернативами – значимо различаются ($P < 0.01$).

Таблица 2. Отношения ψ/p психологической и сложной вероятности

<i>Количество альтернатив в стадии</i>	<i>Количество стадий в матрице</i>				
	2	3	4	5	8
2	1,8	4,0	5,7	12,2	95,0
3	3,8	5,9	17,8	53,2	10^3
4	4,9	14,8	54,0	118,0	7×10^3
5	7,5	27,8	106,0	563,0	3×10^4
8	4,3	82,4	465,0	4×10^3	9×10^5

Этот отличительный эффект становится ясным, если мы составим, как на Рисунке 1, логарифм относительной переоценки, $\log_{10} \psi/p$, против количества информации в битах, $m \log_2 n$, необходимые для определения местонахождения цели. Отношение – линейное и задается уравнением:

$$\log_{10} \psi / p = 0.26 m \log_2 n - 0.31 \quad (1)$$

Оно может быть переписано как:

$$\psi / p = e^{(0.86 m \log_2 n - 0.72)} \quad (2)$$

из которого следует, что если мы оставляем количество стадий постоянным, ψ/p прямо пропорционально показателю степени количества альтернатив в

стадии, и, если мы оставляем количество альтернатив постоянным, ψ / p изменяется по экспоненте с количеством стадий.

Если у нас есть m и n , мы можем предсказать среднее значение ψ — потому что

$$p = n^m \quad (3)$$

Сложная вероятность p получена путем умножения соответствующей вероятности правильного угадывания в каждой из m стадий, и каждая из этих вероятностей, равна $1/n$.

Поэтому, из (2) мы получаем, что

$$\psi = e^{-0.72} n^{-0.14m} \quad (4)$$

$$(e^{-0.72} = 0.49)$$

В основном, поэтому относительная переоценка ψ/p следует из объективного ослабления мультипликативного фактора m количества стадий. Его ослабление имеет порядок шести седьмых от количества стадий, в пределах наших экспериментальных ограничений. Это приводит к тому, что мы выводим то, что мультипликативный элемент в сложной вероятности является далеким от того, чтобы быть “простым” или интуитивным, что может помочь объяснить особые трудности, с которыми люди сталкиваются в изучении статистики.

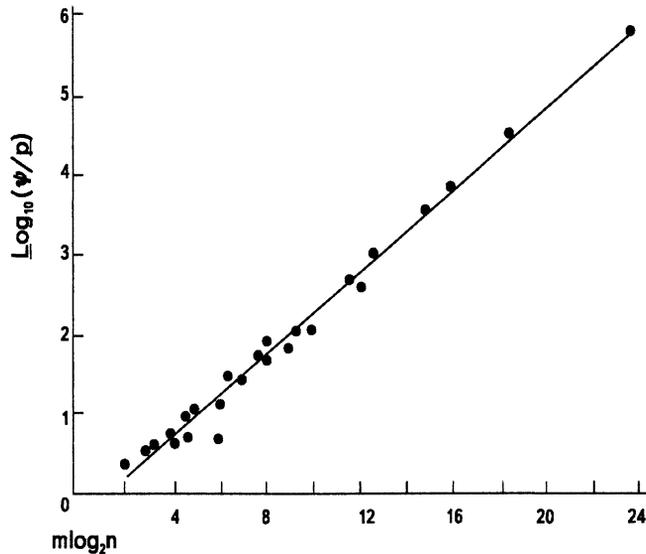


Рисунок 1. Переоценка, относящаяся к логарифму, против количества битов ($\log_{10}(\psi/p)$ против $m \log_2 n$). p — сложная вероятность; m — количество стадий; n — количество альтернатив

Наша интерпретация поддерживается подобными экспериментами с испытуемыми в возрасте 9+ и 10+ лет, у которых, кроме относительной переоценки u – никакие тенденции не являются заметными с вариациями в значениях m и n . Это также предполагает, что мультипликативный элемент не примитивен.

Метод, который мы использовали, вовлекает косвенную оценку ψ . Более прямая оценка может быть получена, если попросить испытуемого выбрать между различными типами матриц. Полезность выбора, однако, могла бы тогда стать важным фактором.

Этот эксперимент объясняет очевидную тенденцию для принимающего решения, в разнообразии ситуаций многоступенчатого выбора недооценивать вероятность его успеха, и поэтому использовать несоответствующую стратегию, о которой он будет позже сожалеть.

С исторической точки зрения интерес представляет тот факт, что самые искушенные мыслители древней Греции, хотя они и очень интересовались идеей возможного, особенно в стоической философии, никогда не применяли комбинаторный анализ, который появился только в шестнадцатом столетии. Сам Аристотель, очевидно, недооценивал понятия вероятности. Любая интуиция испытуемого, которая была у него и у других, была нагружена долгими принятыми привычками мышления.

Относительная переоценка составных сложных вероятностей, которую показал эксперимент, может быть проявлением значительной общности в принятии решений и выборе. Если так, это заслуживает специального обозначения. Мы предлагаем назвать это “инерционным ψ – эффектом”.

25. Консерватизм в процессе обработки информации человеком*

Вард Эдвардс

...Большое количество исследований показало, что люди являются консервативными в обработке неверной информации. Такие эксперименты сравнивают человеческое поведение с результатами теоремы Байеса, формально оптимального правила о том, как мнения (то есть вероятности) должны быть пересмотрены на основе новой информации. Оказывается, что изменение мнения обычно упорядочено прямо пропорционально числам, полученным из теоремы Байеса – но их недостаточно. Первое близкое приближение к данным показало бы, что требуется где-то от двух до пяти наблюдений, чтобы сделать одно наблюдение стоящим работы для побуждения испытуемого изменить свое мнение. Множество экспериментов было нацелено на объяснение этого явления. Они показывают, что одна из главных, а вероятно, и главная причина консерватизма – неверное накопление данных у людей. То есть люди точно чувствуют каждую данную величину и хорошо знают о ее индивидуальном диагностическом значении, но неспособны хорошо скомбинировать ее диагностическое значение с диагностическим значением других данных при пересмотре мнений....

Вероятности определяют степень неопределенности. Вероятность, как согласно Байесу, так и нашей интуиции, составляет просто число между нулем и тем, что представляет степень, для которой несколько идеализированный человек считает, что утверждение верно. Причина, по которой человек несколько идеализирован, состоит в том, что сумма его вероятностей для двух взаимно исключаящих событий должна равняться его вероятности того, что произойдет любое из этих событий. Свойство аддитивности имеет такие последствия, что мало реальных людей могут соответствовать им всем. Так как

* Выдержки из статьи, которая появилась в В. Kleinmuntz (Ed.), *Formal representation of Human Judgement*. Нью-Йорк: John Wiley and Sons, Inc, 1968. Переиздано в соответствии с разрешением.

такие вероятности описывают человека, который поддерживает мнение больше, чем событие, о котором это мнение, их называют индивидуальными вероятностями (см. Savage, 1954).

Теорема Байеса – это тривиальное последствие свойства аддитивности, бесспорное и согласованное для всех сторонников вероятностей, как Байеса, так и других. Один их способов написать это следующий. Если $P(H_A | D)$ – последующая вероятность того, что гипотеза A была после того, как данная величина D наблюдалась, $P(H_A)$ – его априорная вероятность до того, как наблюдалась данная величина D , $P(D | H_A)$ – вероятность того, что данная величина D будет наблюдаться, если верно H_A , а $P(D)$ – безусловная вероятность данной величины D , то

$$P(H_A | D) = \frac{P(D | H_A)P(H_A)}{P(D)} \quad (1)$$

$P(D)$ лучше всего рассматривать как нормализующую константу, заставляющую апостериорные вероятности составить в целом единицу по исчерпывающему набору взаимно исключающих гипотез, которые рассматриваются. Если ее необходимо подсчитать, она может быть такой:

$$P(D) = \sum_i P(D | H_i) \cdot P(H_i)$$

Но более часто $P(D)$ устраняется, а не подсчитывается. Удобный способ устранять ее состоит в том, чтобы преобразовать теорему Байеса в форму отношения вероятность-шансы. Рассмотрим другую гипотезу, H_B , взаимно исключающую H_A , и изменим мнение о ней на основе той же самой данной величины, которая изменила ваше мнение о H_A . Теорема Байеса говорит, что

$$P(H_B | D) = \frac{P(D | H_B)P(H_B)}{P(D)} \quad (2)$$

Теперь разделим Уравнение 1 на Уравнение 2; результат будет таким:

$$\frac{P(H_A | D)}{P(H_B | D)} = \frac{P(D | H_A)}{P(D | H_B)} \cdot \frac{P(H_A)}{P(H_B)}$$

или

$$\Omega_1 = L \cdot \Omega_0$$

где Ω_1 составляет апостериорные шансы в пользу H_A через H_B , Ω_0 априорные шансы, а L – количество, знакомое статистикам как отношение вероятности. Уравнение 3 – это такая же соответствующая версия теоремы Байеса как и Уравнение 1, и часто значительно более полезная особенно для экспериментов, с участием гипотез.

Сторонники Байеса утверждают, что теорема Байеса – формально оптимальное правило о том, как пересматривать мнения в свете новых данных. Этот пересмотр мнения в свете данных составляет именно то, из чего состоит статистический вывод, и поэтому он должен быть структурирован вокруг теоремы Байеса – со многими последовательными отличиями от классической статистической практики. Для элементарной демонстрации этих идей, написанных для психологов - экспериментаторов (см. Edwards, Lindman и Savage, 1963). Но мы – не статистики, или, во всяком случае, ни один из нас сегодня не носит «дурацкий колпак» статистика. Вместо этого, как психологи, мы интересуемся сравнением идеального поведения, определенного теоремой Байеса, с фактическим поведением людей.

Чтобы дать Вам некоторое представление о том, что это означает, давайте попробуем провести эксперимент с Вами как с испытуемым. Эта сумка содержит 1.000 покерных фишек. У меня две такие сумки, причем в одной 700 красных и 300 синих фишек, а в другой 300 красных и 700 синих. Я подбросил монету, чтобы определить, какую использовать. Таким образом, если наши мнения совпадают, ваша вероятность в настоящее время, что выпадет – сумка, в которой больше красных фишек – 0.5. Теперь, Вы наугад составляете выборку с возвращением после каждой фишки. В 12 выборках, Вы получаете 8 красных и 4 синих. Теперь, на основе всего, что Вы знаете, какова вероятность того, что выпала сумка, где больше красных? Ясно, что она выше, чем 0.5. Пожалуйста, не продолжайте читать, пока Вы не записали вашу оценку.

Если Вы похожи на типичного испытуемого, ваша оценка попала в диапазон от 0.7 до 0.8 – хотя утверждение, которое часто делалось в предыдущих параграфах, что люди консервативны в обработке информации, возможно, сместило ваш ответ вверх. Если бы мы проделали соответствующее вычисление, тем не менее, ответ был бы 0.97. Действительно очень редко человек, которому предварительно не продемонстрировали влияние консерватизма, приходит к такой высокой оценке, даже если он был относительно знаком с теоремой Байеса.

Где-то в 1960 Уильям Л. Хейс, аспирант по имени Лоренс Д. Филлипс и я заинтересовались обнаружением несоответствий между человеческим поведением и поведением, указанным теоремой Байеса. Простой пример предыдущего параграфа не приходил нам на ум; вместо этого мы были уверены, что мы должны будем использовать довольно сложную ситуацию, чтобы получить не Байесовское поведение. Так что, мы использовали гипотетическую компьютеризированную радарную систему. Было 12 возможных наблюдений, 4 возможные гипотезы и испытуемые должны были понять и использовать демонстрацию 48 различных значений $P(D|H)$. Испытуемые выполняли задание при двух условиях. В одной, испытуемый видел единственный стимул, точку в секторе дисплея радара; затем он пересматривал свои априорные вероятности по этим четырем гипотезам на основе данной величины, «урегулировав» четыре «рычага» к своим следую-

щим оценкам вероятности, затем повторно «устанавливал рычаги» в позицию 0 в подготовке к следующему стимулу. Вторым стимулом состоял из старой точки плюс новой; испытуемый «устанавливает рычаги», чтобы сообщить о совокупном воздействии обеих точек. И так далее, пока не накопились 15 точек. Во втором условии, стимулы были перетасованы, и испытуемый, в действительности начинал заново с каждым новым стимулом. К удивлению экспериментаторов предсказание теоремы Байеса, что это различие в условиях не должно произвести никакого различия для поведения, было подтверждено. Кроме того, было еще одно условие, в котором каждая новая точка была показана одна, но испытуемым позволяли сохранять оценки от одного стимула до следующего вместо того, чтобы повторно «регулировать рычаги» к нулевой позиции после каждой оценки. И опять, изменение в условиях произвело мало различий в поведении.

Положительных результатов эксперимента Филлипса-Хэйза-Эвардса было три. Во-первых, испытуемые были очень консервативны. Во-вторых, они были наименее консервативны на первой точке, становясь более консервативными с увеличением их количества. Наконец, суммы их оценок вероятности, которые не были ограничены, вообще составили в целом больше чем 1, и увеличивались, по мере прохождения испытуемых через последовательные стимулы в упорядоченной последовательности. Очевидно, для испытуемых было легче определить, какая гипотеза была одобрена стимулом, и таким образом усилить вероятность этой гипотезы, чем решать, от какой другой гипотезы должна получаться вероятность для того, чтобы придать ей легитимность.

Мы сильно медлили с публикацией этого первоначального эксперимента по консерватизму. Хотя данные были полными уже к 1962, книга Филлипса-Хэйза-Эвардса не выходила в печать до 1966 (Phillips и другие, 1966).

Масштаб и устойчивость проявления консерватизма поразили нас. Казалось уместным попробовать намного более простые задачи. Так, без большой веры, Филлипс и я пробовали предварительное испытание, подобное по характеру примеру с сумкой и покерными фишками, который Вы рассматривали выше. К нашему удивлению, это сработало очень хорошо. Большая часть текущего исследования, сравнивающего человеческое поведение с теоремой Байеса, может быть прослежено к этому предварительному испытанию и последующему эксперименту.

Если доля красных фишек в сумке — p , то вероятность получения r красных фишек и $(n - r)$ синих в n выборках с возвращением в некотором порядке — $p^r(1-p)^{n-r}$. Так, в типичном эксперименте с сумкой и покерными фишками, если H_A означает, что доля красных фишек составляет p_A и H_B — означает, что доля составляет p_B , тогда отношение вероятности

$$L = \frac{p_A^r(1-p_A)^{n-r}}{p_B^r(1-p_B)^{n-r}} \quad (4)$$

Обратите внимание, что, в то время как Уравнение 4 было получено из рассмотрения фактической последовательности красных и синих в выборке, оно, возможно, одинаково могло быть получено из рассмотрения r красных и $(n - r)$ синих в любом порядке; двучленный коэффициент, который представляет число различных способов, с помощью которых можно получить r красных в n испытаниях, появляется и в числителе, и в знаменателе, и, таким образом, сокращает отношения вероятности. Это иллюстрация принципа вероятности статистики Байеса (см. Edwards, Lindman и Seavage, 1963), который в действительности говорит, при применении формулы Байеса необходимо учитывать только вероятность фактического наблюдения, а не вероятности других наблюдений, которые он, возможно, сделал бы, но не сделал. Этот принцип имеет широкое воздействие на все статистические и нестатистические применения теоремы Байеса; это самый важный технический инструмент размышления Байеса.

В особом случае, в котором $p_A = 1 - p_B$ (симметрический двучленный случай), отношение вероятности сокращается до

$$L = \left(\frac{p_A}{1 - p_A} \right)^{2r - n} \quad (5)$$

Обратите внимание, что $2r - n = r - (n - r)$ — различие между количеством красных и синих в выборке; только это различие, а не общее количество наблюдений, составляет релевантным выводом в этом симметрическом случае. Статистическая традиция отмечает, что успехи различия (difference successes) минус неудачи, или $s - f$; это обычная независимая переменная экспериментов с сумкой и покерными фишками. Чтобы понимать объяснение для обычных зависимых переменных, замените Уравнение 5 на Уравнение 3, возьмите логарифмы и перестройте термины. Результат

$$\log L = (2r - n) \log \frac{p_A}{1 - p_A} = \log \Omega_1 - \log \Omega_0$$

Если испытуемый действует по Байесу, отношение вероятности логарифма, которое может быть выведено, вычитанием логарифма априорных шансов из логарифма апостериорных шансов, должно быть пропорционально $s - f$, независимой переменной. Уместно начертить выведенное отношение логарифма вероятности испытуемого, таким образом подсчитанное из его апостериорных шансов (которые в свою очередь были рассчитаны из его апостериорных вероятностей, если он оценивал вероятности) и объективно соответствующие априорные шансы, против $s - f$.

Большинство экспериментов с сумкой и фишками в Мичиганской лаборатории использовало генеральную совокупность из 48 пронумерованных пультов, состоящих из кнопки, красной и зеленой лампочек. Когда кнопка

на пульте нажата, одна из лампочек гаснет и снова загорается; испытуемым говорят, что это эквивалентно выборке с возвращением фишки соответствующего цвета. Испытуемым говорят, что программа, которая управляет лампочками, была написана, путем осуществления выборок из сумки. Фактически, для большинства экспериментов, программа скорее тщательно разработана так, чтобы показанная последовательность была репрезентативна выборке из сумки, и в особенности так, чтобы в каждом эксперименте выборки размера n одобряли несоответствующую гипотезу, часто используя значение p_A , для всех значений n .

Филлипс и я (1966) исследовали влияние p_A , используя последовательно 20 фишек и p_A значений 0.55, 0.7 и 0.85. Тестируемые оценили следующие вероятности, распределяя 100 белых деревянных дисков в два сосуда. Типичные результаты таких экспериментов представлены на рис. 1 для 0.7 мешков с различными априорными вероятностями. Три результата, проиллюстрированные на Рисунке 1, были обнаружены для всех испытуемых. Во-первых, выведенные отношения логарифма вероятности были приблизительно прямо пропорциональны $s-f$. Во-вторых, априорные вероятности использовались соответственно; то есть лучшая подходящая линия для всех

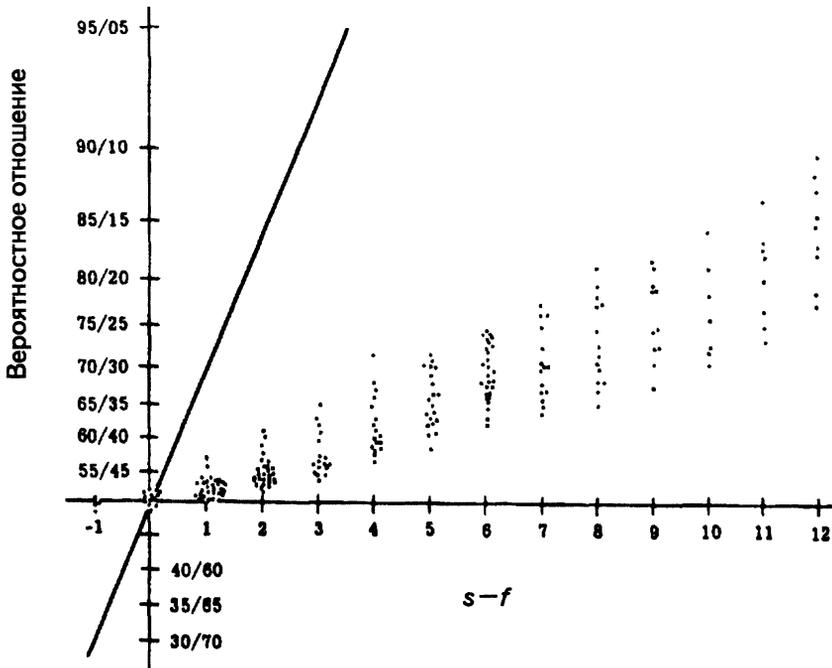


Рис. 1. Оценки отдельного испытуемого для p_A 0.7 выраженные во внешкальном логарифме вероятностных отношений как функции различия между количеством успехов и количеством неудач в выборке

точек данных проходит через изначальную. В-третьих, испытуемые были консервативны; лучшая подходящая линия была, более ровной, чем линия, представляющая оптимальную формулу Байеса. Обнаружение почти линейности выведенных коэффициентов логарифма вероятности для $s-f$ (или, что эквивалентно, с коэффициентами логарифма вероятности Байеса) предлагает еще одну зависимую переменную: отношение наклона лучше всего соответствующей линии для всех оценок испытуемых к наклону линии Байеса. Петерсон, Шнайдер и Миллер (Peterson, Schneider и Miller, 1965) назвали это отношение отношением точности; они также обнаружили, что оно более или менее постоянно для $s-f$.

Рисунок 2 показывает соотношения точности данных Филлипса-Эвардса для трех значений p_A . Для наименее диагностической информации, испытуемые показали более экстремумные результаты, чем теорема Байеса. (Dale обнаружил то же самое; см. W. Edwards, 1965.) Но для информации, имеющей разумно высокое диагностическое значение, испытуемые были консервативны, и соотношение точности было очень постоянным для $s-f$. Обратите внимание, что с увеличением диагностичности, консерватизм также увеличивается. Это стандартный вывод в таких экспериментах; любая процедура, которая увеличивает диагностичность индивидуального наблюдения (одной фишки или нескольких) также увеличивает консерватизм. (См. например Peterson, Schneider и Miller, 1965.)

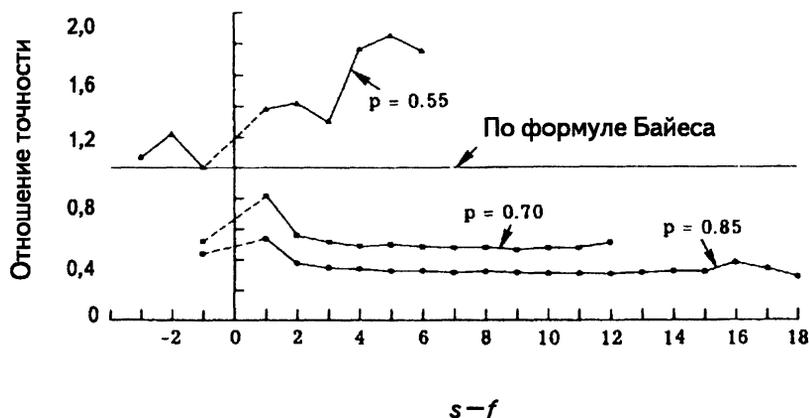


Рис. 2. Отношение точности для трех значений p_A для различных составов выборок

Филлипс и я, после получения этих результатов, пришли к выводу, что одна причина для консерватизма могла бы состоять в том, что испытуемые, зная, что шкала вероятности ограничена и, замечая, что доказательство могло продолжать подниматься, занижали свои оценки. Очевидное средство, в таком случае, должно использовать неограниченный способ ответов, таких как шансы. Так что мы провели исследование с четырьмя группами. Контрольная группа оценила вероятности, распределяя 100 дисков в два сосуда, как прежде. Группа, оценивающая шансы вербально, просто сделала устные оценки шансов; мы всегда принимаем шансы, как числа, равняющиеся единице или больше, и поэтому всегда сопровождаем утверждения шансов утверждениями, гипотезы которых поддерживаются шансами. Группа, оценивающая шансы с помощью логарифмических шкал, сделала оценки, перемещая указатель по шкале шансов, которая содержала четыре цикла логарифма, так, что могли быть, оценены шансы от 1:1 до 10.000:1. Четвертая группа также использовала шансы на логарифмической шкале, но числа, введенные напротив маркировки шкалы, были скорее вероятностями, чем шансами (таким образом, 0.5, а не 1:1, 0.67, а не 2:1, 0.80, а не 4:1, и т.д.). Это назвали вероятностью на группе логарифмической шкалы шансов. Было обнаружено, что все группы были весьма консервативны. Вероятностная группа была самой консервативной, группа на логарифмической шкале шансов была еще хуже, а две группы шансов были приблизительно сопоставимы, а шансы на логарифмическом масштабе были немного большими.

Это открытие просто подчеркивает факт, который стал все более и более ясным в ходе нашей работы Байеса. Вероятность – довольно плохое измерение неопределенности, кроме как в ситуациях, в которых перераспределение или другое прямое использование свойства аддитивности являются необходимыми. Что лучше шансы или логарифм шансов? Шансы наиболее интуитивно понятны для неопытных испытуемых, и могут наиболее легко быть связанными с простыми действиями (например, выбор среди ставок); тот факт, что сфера азартных игр структурирует свои утверждения и имеет дело с шансами, а не с вероятностями, является и признанием, и возможно причиной большей интуитивной значимости шансов. Логарифм шансов, уникальный среди более или менее общей метрики неопределенности, имеет такое свойство, которое в этом метрическом свидетельстве является аддитивным. Если мнение измерено с помощью логарифма шансов, количество изменений во мнении, произведенных частью свидетельства, независимо от того, с чего мнение должно было начинаться. Это изящное свойство делает логарифм шансов уникально удобными для экспериментов Байеса.

Данные Филлипса-Эвардса могут быть пригодны за счет простой модификации теоремы Байеса:

$$\Omega_1 = L^c \cdot \Omega_0$$

Константа C , степень, в которую каждое отношение вероятности возводится до обработки ее посредством теоремы Байеса, составляет отношением точности. К сожалению, она зависит от важных независимых переменных, включая диагностичность данных и ответную метрику. Однако, то, что такая простая наглядная модель так хорошо вписывается, можно объяснить в соответствии с любой теорией консерватизма....

... Вероятностная система обработки информации, или *ВСОИ* (или с англ. *PIP*)... – это идея о том, как проектировать человеко-машинные системы, которые должны обрабатывать информацию для достижения заключения о том, в каком состоянии находится мир. Примеры ситуаций, в которые такая обработка информации должна быть сделана, включают медицинский диагноз, военное командование (в котором командующий, возможно, должен определить, действительно ли он находится под угрозой нападения, и если так, каков план его противника), и управление бизнесом (например, в случае бизнесмена, решающего, необходимо ли производить новое изделие). Идея относительно *ВСОИ* слишком сложная, чтобы объяснять ее подробно здесь (см. Edwards, Lindman и Phillips 1965, или W. Edwards, 1966). Сущность *ВСОИ* в том, что задача диагностической обработки информации может быть разделена на два класса подзадач. Один класс подзадач состоит из оценки диагностического воздействия индивидуальной данной величины на отдельную гипотезу или пару гипотез. Для вербальных, качественных видов данных и для гипотез, которые характеризуют многие реальные диагностические установки, это, кажется, задача, которая обязательно должна выполняться человеком с большим опытом. Но второй класс подзадач – это соединение этих отдельных диагностических воздействий для всех данных и гипотез в картину того, как все гипотезы в настоящее время представлены в свете всех доступных данных. Эта задача соединения может быть легко механизирована посредством теоремы Байеса, если диагностические воздействия индивидуальных данных оценены в форме значений $P(D|H)$ или отношения вероятности. (В большинстве ситуаций, хотя не во всех, оценки отношений вероятности явно предпочтительны, по формальным причинам, для оценки $P(D|H)$.)

Мне и приблизительно пятнадцати сотрудникам было интересно, работает *ВСОИ* или нет. Мы разработали придуманный, но сложный мир 1975. В этом мире мы перечислили шесть гипотез, которые тестируемые должны были рассмотреть, определяющие три источника данных (Система дальнего обнаружения баллистических ракет, система спутникового зондирования и система разведки), которые обеспечивали данные, опирающиеся на эти гипотезы, и разработали четыре системы для обработки данных. Эти четыре системы называли PIP, POP, PEP и PUP. В PIP, тестируемые оценивали пять отношений вероятности для данной переменной. Одна из этих шести гипотез была “Будет продолжаться мир”, а другие пять были вероятностью различных войн; пять конъюнкций войны с миром определяли пять отношений вероятности, которые будут оценены. Другие три системы обра-

ботки информации, которые все подразумевали, что испытуемый оценивал апостериорные шансы или вероятности или подобные апостериорные величины; таким образом, в РІР компьютер соединил данные посредством теоремы Байеса, в то время как во всех трех других системах испытуемые должны были объединить данные в уме. Чтобы помочь им это сделать, тестируемые в РОР, РЕР и РІР производили оценивание, после того, как им становилась доступна n -ная данная величина, когда они рассматривали $(n+1)$ данную величину, так что им приходилось только модифицировать оценки, на которые повлияли данные величины.

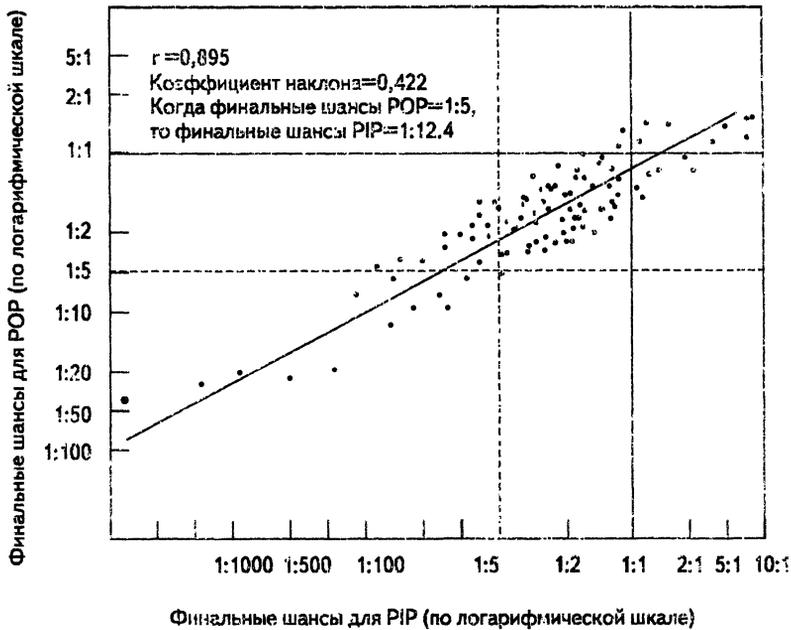


Рис. 3. Финальные шансы в пользу войны для ROP в сравнении с RIR по логарифмической шкале

Было всего 18 сценариев с 60 условиями данных в каждом сценарии. Все условия кроме информации о системе дальнего обнаружения баллистических ракет были представлены в форме коротких абзацев. 34 испытуемых были обучены характеристикам мира, гипотезам, этим трем источникам данных и системе обработки информации, которую каждый должен был использовать.

Так как RIR была явно лучшей, а ROP была следующей, я представлю только сравнение между ними. (RUR была третьей, а PER, самый близкий результат, к тому способу, которым обработка информации происходит сейчас, была худшей.) Рис. 3 показывает заключительные шансы, после 60-го условия данных в каждом сценарии, в пользу каждой войны по сравнению с миром для RIR и для ROP. Две самые важные вещи, которые нужно заме-

тять на графике – это то, что эти две группы очень хорошо согласуются качественно (корреляция между ними – 0.895), но они не согласуются количественно. PIP намного более подвержена влиянию данных, чем POP; тот же самый сценарий, который приведет PIP к тому, что он будет очень уверенным относительно мира или войны, приведет POP, к намного меньшей уверенности. Иными словами, PIP – менее консервативна, чем POP, возможно, потому что в POP, субъекты должны объединить данные, в то время как в PIP, испытуемые оценивают диагностическое воздействие каждой данной величины отдельно, и теорема Байеса делает их соединение.

Вы должны также обратить внимание, что обе оси на рис. 3 логарифмически разделены. Если Вы переводите различие в эффективности обратно в шансы, различие между PIP и POP становится очевидным. Например, вычисление, исходя из линии регресса, привел ли сценарий PIP к тому, чтобы дать 99:1 шансы в пользу войны над миром, POP дал бы только 4:1 (шансы) в пользу этой войны, а не мира.

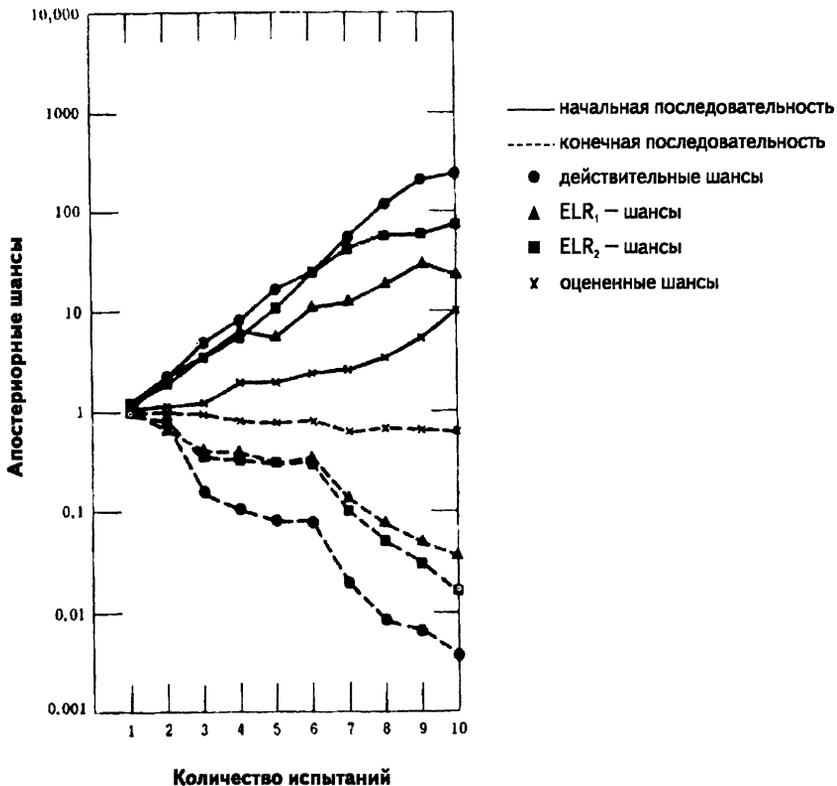


Рис. 4. Медианные апостериорные шансы для всех испытуемых в пользу начального множества как функции количества испытаний

Гипотеза неправильного восприятия не может объяснить это несоответствие между РІР и РОР, в РІР испытуемые оценивают диагностическое воздействие каждой данной величины отдельно; в РОР испытуемые должны объединить в уме – и сделать это весьма консервативно. Так как никакая модель процесса порождающего данные не доступна, невозможно сказать каковы правильные апостериорные шансы. Но различие между РІР и РОР явно вызвано различием в процессе агрегации.

Лэрри Филлипс, один из сотрудников в этом эксперименте, был заинтересован тем, что никакая модель процесса порождающего данные не была доступна и, так что нельзя было сказать с уверенностью, был ли РІР или РОР более правилен. Так в своей докторской диссертации он сравнил РІР с РОР в ситуации, в которой модель процесса порождающего данные была доступна, имело смысл спросить соотношение вероятности для отдельной данной величины, и процедура РОР произвела консервативные оценки. Его испытуемые были редакторами студенческой газеты Мичиганского Университета. Он взял статьи каждого редактора в течение семестра, посчитал две первые и последние буквы в каждом слове каждой статьи, и таким образом, для каждого редактора подготовил множество начальных биграмм и множество конечных биграмм. Для задачи РІР, он взял определенные биграммы, и попросил редактора оценить (только для его собственного множества) отношение вероятности, взятое с гипотезой множества начальных биграмм в числителе и гипотезой множества конечных биграмм в знаменателе, связанное с каждой биграммой. Для задачи РОР, он подготовил последовательность биграмм, выбранных из одного из множеств, и спросил редактора, когда он прорабатывал последовательность, оценить апостериорные шансы, что это было начальное, а не конечное множество. Много усилий было потрачено на предварительное обучение редакторов, и оценки отношения вероятности были собраны дважды, один раз до и один раз после апостериорных оценок шансов.

В анализе данных возникла проблема, потому что все оценки, и для РІР, и для РОР, были смещены в пользу начального множества. Это вероятно, потому что намного легче, например, думать о словах, которые начинаются на *ge*, чем о словах, которые заканчиваются на *ge*, даже при том, что *ge* скорее встречается как окончание, чем как начало слова; мы приучены к маркировке слов по их началам, а не окончаниям, когда мы, например, ищем их в словаре. Однако, возможно исправить такие предубеждения. Рисунок 4 показывает результаты после такого исправления. Действительные шансы, подсчитанные из фактического счета биграмм, являются наиболее экстремумными. Затем идут шансы, подсчитанные из второго набора оценок отношения вероятности. Затем идут шансы, подсчитанные из первого набора оценок отношения вероятности. И, наиболее близкими к середине и поэтому наиболее консервативными, являются непосредственно оцененные апостериорные шансы. Если мы верим этим данным (а я верю), хотя РІР значительно менее консервативен, чем РОР, он все-таки слишком консервативен, но оценки РІР улучшаются с практикой.

26. Гипотеза угадывания в многоступенчатом выводе*

*Чарльз Ф. Джеттис, Клинтон Келли III и
Камерон Р. Петерсон*

Многоступенчатый вывод состоит из ряда одноступенчатых выводов, где результат каждой предыдущей стадии становится входными данными к следующей стадии. В выводах с единственной стадией рассуждают, исходя из данных или однозначно наблюдаемого свидетельства к набору гипотез. Многоступенчатый вывод начинается с тех же самых однозначных данных или свидетельства в первой стадии; однако, входные данные для следующей стадии – это результат предыдущей стадии. Следующая стадия вывода поэтому базируется на вероятностях событий, а не на определенном знании, что некоторый случай является верным (Gettys и Wilike, 1969).

Например, предположим, что Вы хотели предсказать успех или неудачу большой вечеринки в саду. Предположим, что вечеринка, менее вероятно, будет успешной, если собравшиеся будут толпиться в закрытом помещении из-за дождя. Ваша переменная – присутствие темного облака на горизонте. Первая стадия вывода связала бы темное облако с наличием или отсутствием дождя в течение вечеринки. Предположим, что Вы оценили, что вероятность дождя была 0.70. Эта оценка стала бы входными данными к следующей стадии вывода. Если бы Вы знали с уверенностью, что будет дождь, то Вы могли бы вывести вероятность того, что вечеринка будет иметь успех. Но Вы полностью не уверены, что будет дождь; данные, которые Вы имеете, указывают на дождь с вероятностью 0.70, так что, что вам теперь делать?

Модифицированная теорема Байеса (МТБ) обеспечивает оптимальную модель для таких многоступенчатых выводов (Dodson, 1961; Gettys и Wilike, 1969). Множество исследований показало, что интуитивная работа в многоступенчатой задаче заканчивается *большой определенностью*, извлекаемой из данных, чем предсказано моделью МТБ. Например, в задаче оценки шан-

* Эта глава первоначально появилась в *Organizational Behavior and Human Performance*, 1973, 10, 364-373. Авторское право © 1973 Academic Press, Inc. Переиздано в соответствии с разрешением.

сов шансы испытуемых типично больше, чем рассчитанные МТБ. Этот результат весьма удивителен, потому что свидетельство указывает, что поведение людей в задаче вывода единственной стадии почти всегда консервативно; то есть, люди извлекают меньше определенности, чем гарантируют данные (например, W. Edwards, 1966). Парадокс, конечно, в том, что многоступенчатый вывод является совокупностью выводов. Если люди извлекают меньше определенности, чем гарантируют данные в одноступенчатых выводах, то в многоступенчатой ситуации можно было ожидать, что они станут все более консервативными с каждой последующей стадией, так как их отклонения от не оптимальности должны накапливаться от стадии к стадии. Фактически это изменение верно; они более уверены в конце двух стадий вывода, чем гарантировано оптимальной моделью МТБ. Из этого следует, что некоторый процесс происходит в "интерфейсе" одноступенчатого вывода, который является настолько экстремумным, что любой консерватизм побежден.

Задача вывода единственной стадии всегда базируется на данных, которые известны. Однако, даже при том, что многоступенчатая задача начинается с определенных данных, апостериорные стадии вывода имеют дело с неопределенными данными. Несколько моделей были сформулированы для объяснения, как необходимость иметь дело с вероятностями данных, а не с определенными данными могла бы создать чрезмерную определенность в многоступенчатом выводе. Одна неоптимальная модель, имеющая свойство предсказания чрезмерной определенности – модель "Как будто" (Gettys и Willke, 1969; Howell, Gettys, и Martin, 1971). Эта модель, разработанная для ситуаций, в которых люди имеют выбор, чтобы собрать больше данных, если они чувствуют, что они нужны, предполагает, что сбор данных продолжается в первой стадии вывода, пока человек принимающий решения не достаточно уверен в состоянии мира. Как только его уверенность превышает некоторое пороговое значение, он переходит к следующей стадии вывода, действуя "как будто" он полностью уверен относительно данных следующей стадии. Чтобы возвратиться к примеру вечеринки в саду, принимающий решение, увидев темное облако, мог бы получить текущее сообщение прогноза погоды. Предположим, в прогнозе погоды был предсказан серьезный шторм. Уверенность в вероятности дождя теперь превысила бы его пороговое значение, и он переходит ко второй стадии вывода, "как будто" он уверен в том, что будет дождь. Результат второй стадии вывода был бы его оценкой вероятности успеха или неудачи, основанной на его "как будто" предположении о дожде. *Его оцененная вероятность для неудачи должна теперь превысить фактическую вероятность (МТБ) неудачи, потому что, делая "как будто" предположение о дожде, он игнорирует вероятность, что этого дождя может не быть.* Если, фактически, его, "как будто" предположение неправильно, и дождь не идет, то вечеринка, вероятно, будет иметь успех. Оптимальная модель рассматривает обе возможности, дождь и отсутствие дождя, в приписывании вероятности к успеху или не-

удаче. Модель "как будто" рассматривает только возможность дождя, и по этой причине ведет к чрезмерной уверенности, что вечеринка не удастся.

Как человек мог бы вести себя, если бы его уверенность о входных данных ко второй стадии вывода была меньше, чем пороговое значение, требуемое для "как будто" предположения, и не было никакой надежды относительно увеличения его уверенности с получением дополнительных данных? Одна возможная гипотеза, которая является совместимой с возрастающей уверенностью, обнаруженной в предыдущих исследованиях – то, что он будет сначала делать предположение "как будто", которое является в лучшем случае угадыванием. Эта модель, названная "Угадывания" действует как усовершенствованная "как будто" модель и разделяет с ней идею того, что принимающий решения будет либо игнорировать, либо иметь тенденцию игнорировать значения других менее вероятных событий во второй стадии вывода, концентрируясь почти исключительно на наиболее вероятном событии. Для примера, если единственная информация, которую Вы имеете – темное облако на горизонте, Вы не захотите сделать *дисквалифицированное* "как будто" предположение, но Вы могли бы сначала предположить, что будет дождь и достигнуть субъективных шансов для успеха, основанного на этом предположении. Тогда, так как Вы не полностью уверены, что будет дождь, Вы могли бы несколько уменьшить ваши субъективные шансы, чтобы принять это во внимание. Эти субъективные шансы могли бы отличаться от рассчитанных с помощью МТВ, прежде всего потому, что Вы явно не рассмотрели вероятность того, что не будет никакого дождя.

Сэппер и Фрайбек (Snapper & Fryback, 1971) представили результаты, которые являются совместимыми с вышеупомянутым объяснением в эксперименте, касающемся надежности данных. Однако, их процедура не позволяла проводить прямое испытание модели "Угадывания"; что является целью следующего эксперимента.

Метод

Цель эксперимента требовала, по крайней мере, трех уровней переменных, построенных таким образом, что промежуточный уровень переменной содержит более двух событий. Далее требуется управление распределением вероятности для всех событий, кроме наиболее вероятного из промежуточных – управление, которое будет иметь результирующее воздействие на величину пересмотра оптимальной вероятности на верхнем уровне как результат возникновения события на более низком уровне.

Следовательно, три уровня приняли следующую форму. Переменная верхнего уровня состояла из двух мешочков, помеченных I и II, соответственно. Каждый мешочек был заполнен маленькими контейнерами, которые представляли собой события промежуточного уровня. К примеру, каждый мешочек содержал 18 маленьких коробочек (коробочки от 35-миллиметровой пленки) и каждая из них была помечена A, B, C или D. Наконец, в каждой

коробочке было 100 маленьких цветных картонных кружочков; каждый из которых был одного из четырех цветов: красный, желтый, зеленый, синий.

Состав каждого контейнера описан в Таблице 1. Часть А таблицы описывает состав мешочка относительно коробочек, и Часть В – относительно кружочков. Например, 8 коробочек, помеченных А, находятся в мешочке I, тогда как только I коробочка, помеченная А, находится в мешочке II. Как показано в части В, 80 кружочков находятся в коробочке А, 1 – в коробочке В, 1 – в коробочке С и 18 – в коробочке D.

Таблица 1. Числовой состав мешочков и коробочек из-под пленки

<i>Буква на коробочке из-под пленки</i>	<i>А. Состав мешочка</i>	
	<i>Мешочек 1</i>	<i>Мешочек 2</i>
А	8	1
В	3	6
С	6	3
D	1	8

<i>Цвет кружочка</i>	<i>В. Состав коробочки из-под пленки</i>			
	<i>Коробочка А</i>	<i>Коробочка В</i>	<i>Коробочка С</i>	<i>Коробочка D</i>
Красный	80	1	1	18
Зеленый	1	80	18	1
Желтый	18	1	80	1
Синий	1	18	1	80

Эксперимент проходил следующим образом. Выбирались наугад: один из двух мешочков, одна коробочка из этого мешочка, один кружочек из этой коробочки. Таким образом, выбор красного кружочка обеспечивает доказательство в пользу коробочки А, что в свою очередь обеспечивает доказательство в пользу мешочка 1. Заметьте, что, это только событие первого уровня, кружочек, который наблюдается непосредственно. Это наблюдение обеспечивает только частичное доказательство с учетом события промежуточного уровня, коробочка, которая, в свою очередь, обеспечивают частичное доказательство, о том, какое событие верхнего уровня было выбрано. Таким образом, первая стадия вывода имеет отношение к цвету кружочка, к букве на коробочке, а вторая стадия вывода имеет отношение, к букве на коробочке и номеру мешочка.

Стратегия действия, как будто наиболее вероятное событие верно на одном уровне, приведет к распределениям вероятности, которые являются экстремумными на следующем более высоком уровне. Таким образом, эта

стратегия совместима с эмпирическим результатом, согласно которому люди чрезмерно пересматривают вероятности верхнего уровня в многоступенчатой задаче.

Существует другая тестируемая гипотеза, которая может быть получена из стратегии "Угадывания". Если человек действует, как будто наиболее вероятное событие верно на любом промежуточном уровне, тогда он игнорирует распределение вероятности для всех других событий на этом уровне. Его пересмотр вероятности на верхнем уровне должно быть поэтому нечувствителен к изменениям в распределении вероятностей для всех событий, кроме наиболее вероятного на промежуточном уровне. Данный эксперимент был предназначен для проверки этой гипотезы.

План эксперимента

Были разработаны три задачи по типу представленных в Таблице 1. Частоты, показанные в части А, использовались во всех трех задачах. Матрица, показанная в части В, использовалась в одной задаче; в других двух задачах значение 80 в более низкой матрице было изменено или на 70, или 90, а значение 18 было изменено или на 28, или на 8, соответственно. Для последующего обсуждения эти три задачи будут определяться как задачи 70-28, 80-18, 90-8. Во всех трех задачах испытуемые оценили шансы мешочков, учитывая цвет одного кружочка, взятого из коробочки.

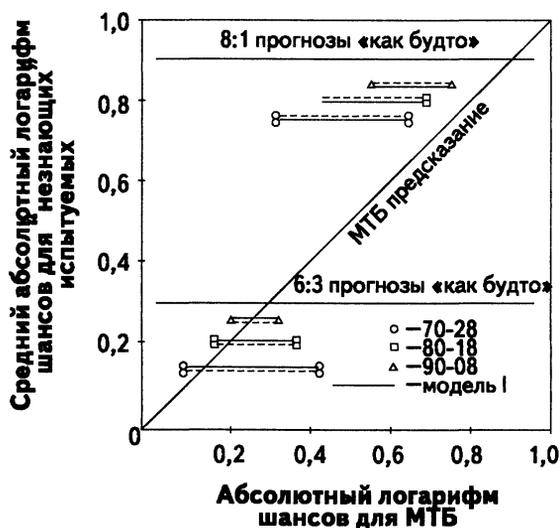


Рисунок 1. Модели "угадывания", "как будто" и МТБ как показатели поведения "незнающих" испытуемых

Испытуемые

25 испытуемых были студентами Мичиганского университета, которые ранее принимали участие в эксперименте с многоступенчатым выводом, продолжительностью около двух часов. В предыдущем эксперименте испытуемые были обучены необходимому способу ответа и произвели ряд оценок шансов в задаче с многоступенчатым выводом. Однако никогда не обсуждалась оптимальная модель, и никогда не использовалась обратная связь.

Инструкции испытуемым

Инструкции были краткими из-за предыдущего эксперимента. Испытуемым объяснили детали задачи. Их попросили представить, что мешочек был выбран наугад на основании подбрасывания монеты, что коробочка была наугад выбрана из мешочка, и что картонный кружочек был наугад выбран из коробочки. Их попросили предположить, что кружочек определенного цвета был выбран таким образом и оценить шансы мешочков на основании цвета этого кружочка.

Процедура

Получив инструкции, испытуемые оценили шансы мешочков в трех задачах. Матрицы, подобные указанным в Таблице 1, использовались, чтобы показать испытуемым относительные частоты коробочек и кружочков. Задачи были представлены в случайном порядке для каждой группы, состоящей от 4 до 6 человек. В каждой задаче четыре возможных цвета использовались в случайном порядке. Испытуемые оценивали шансы мешочков для всех возможных цветов перед тем, как приступить к следующей задаче. После 12 оценок (4 цвета в задаче \times 3 задачи) три задачи были повторены, используя другие случайные порядки для всех 24 оценок, по две для каждого цвета в каждой задаче.

Результаты и обсуждение

Обзор данных показал чрезвычайную бимодальность в оценках шансов. Для некоторых испытуемых теоретическое различие между синими и красными кружочками и между зелеными и желтыми кружочками не вызывало различия в оценках шансов. Другие испытуемые были более крайними в оценках шансов с синим кружочком, нежели с красным и более крайними с желтым кружочком, нежели с зеленым. Последние соответствовали МТБ, по крайней мере, в порядковом смысле. Оказалось, что некоторые испытуемые "не знали" о разнице синий-красный и желтый-зеленый, в то время, как другие испытуемые "знали" в той степени, что они отвечали, по крайней мере, в правильном отношении. Принимая это во внимание, все испы-

туемые, которые отвечали, по крайней мере, с одной оценкой шансов для синего, которая была на 2% больше, чем зеленого или оценка для желтого, на 2% выше, чем зеленого, были названы "знающими". Эти испытуемые были, по крайней мере, минимально "знающими", т.к. для одного или нескольких оцениваний их оценки шансов изменялись в паре синий-красный и паре желтый-зеленый в направлении, которое определяла МТБ. 10 из 25 испытуемых были названы "знающими" с помощью этого консервативного критерия.

Остальные 15 "незнающих" испытуемых, были не склонны отвечать по-разному на изменение в вероятностях менее вероятных событий. Они буквально игнорировали выводы менее вероятных коробок. Их ответы соответствуют крайней форме модели "угадывания". Медианы ответов "незнающих" испытуемых показаны на рис.1. Поскольку мешочек, по которому шансы предпочтительны иррелевантны, данные нанесены на шкалу абсолютного логарифма шансов. Средние ответы логарифма шансов на красные и синие кружочки соединены сплошной линией в верхней части рисунка для трех уровней неопределенности данных, медианы для желтых и зеленых кружочков также помечены в нижней части рисунка. На рис. 1 также обозначены прогнозы для МТБ (линия на диагонали), предсказания для модели "как будто" (две горизонтальные линии) и предсказания для версии модели угадывания, обозначенной на рисунке как модель I.

Предсказания модели I получены путем умножения вероятности наиболее вероятного события на апостериорные шансы, полученные, если событие было верным. Предположим, в задаче 80-18 был выбран красный кружочек. Вероятность коробочки А – 0.80, шансы 8:1, если действительно кружочек вынимался из А. Предсказания модели I будут $0.88/1=6.4$ или шансы 6.4:1. Модель "как будто" прогнозирует шансы 8:1 для синих и красных кружочков и 6:3 для желтых и зеленых, *при условии, что пороговая уверенность в типе коробочки преодолена*. Для МТБ оптимальные шансы для красного кружочка – 2.86:1 и могут быть вычислены в соответствии со следующей формулой для апостериорных шансов (изменено из формулы 5 в Gettys & Willke, 1969):

$$\frac{P(CI | \text{цвет})}{P(CII | \text{цвет})} = \frac{P(CI)}{P(CII)} \times \frac{\sum_i P(\text{цвет} | \text{коробка}_i) P(\text{коробка}_i | MI_i)}{\sum_i P(\text{цвет} | \text{коробка}_i) P(\text{коробка}_i | MII)}$$

где М – мешочек, и другие данные посчитаны из условных вероятностей, как показано в Таблице 1.

Данные на Рисунке 1 не подходят ни к прогнозам "как будто", ни к МТБ. Ответы испытуемых менее крайние, чем прогнозы "как будто" для верхних пар синий-красный, где прогноз "как будто" – 8:1 шансов, и также менее крайний, чем прогноз 6:3 шансов в нижней части рисунка. Однако крайние версии модели угадывания и модели I, соответствуют медианам Рисунок 1.

Горизонтальные пунктирные линии на Рисунке 1 – прогнозы модели I. Для всех задач прогнозы модели I расположены правее диагонали МТБ для желтых кружочков в нижней части Рисунка 1. Модель "как будто" и модель I не обязательно прогнозируют оценки шансов, которые являются более крайними, чем шансы МТБ. Эти вопросы поднимаются, например, в задаче 80-8, когда коробочка С является наиболее вероятной ($P=0.80$), а коробочка А ($P=0.18$) наименее вероятной. Наиболее вероятное событие дает шансы 6:3 для мешочка I, если оно верно и наименее вероятное событие дает шансы 8:1 для мешочка I. В этом случае, любая модель, которая игнорирует отношение 8:1, данное менее вероятным событием, будет консервативной по отношению к МТБ.

Если допустить это, испытуемые не приняли бы неоптимальную модель, если она была получена из субъективного чувства уверенности, и, возможно, важный результат – то, что они использовали модель I, потому что думали, что она правильная. Была определена величина оценки шансов моделью I, но в другой ситуации они могли использовать другое правило комбинации. Тот факт, что прогнозы модели I действительно соответствуют данным, предполагает, что испытуемые склонны сосредотачиваться на наиболее вероятной альтернативе и игнорировать выводы менее вероятной альтернативы.

Данные для 10 "знающих" испытуемых представлены на Рисунок 2. Как и на Рисунок 1, прогнозы модели "как будто" и модели МТБ показаны на рисунке, но прогнозы Модели I опущены, потому что они явно не соответствуют данным.

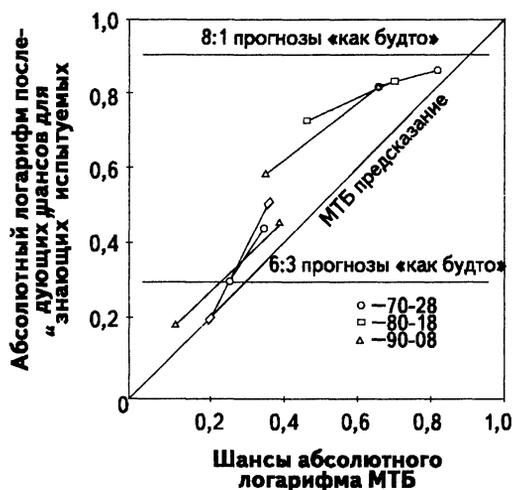


Рис. 2. Результаты "знающих" испытуемых в сравнении с прогнозами моделей МТБ и модели "как будто"

В общем, "знающие" испытуемые отвечали на те же переменные, что и МТБ, но количественное соответствие модели МТБ скромное. Как и МТБ, испытуемые менее уверены, чем подразумевалось моделью "как будто" для синих и красных кружочков. Также, как и в МТБ, их оценки желтому кружочку превосходят прогнозы "как будто". Это происходит, потому что наиболее вероятное событие имеет шансы 6:3 и менее вероятное событие имеет шансы 8:1. Если испытуемые осознают нюансы многоступенчатых ситуаций, они должны понимать, что шансы должны быть больше, чем 6:3. Единственное исключение к этой общей картине – это расположение данных 80-18 для желтых и зеленых кружочков. Апостериорные шансы, основанные на желтом цвете, должны увеличиться в то время, как вероятность наиболее вероятного события уменьшается, а шансы, основанные на зеленом, должны уменьшиться с увеличением вероятности. Ответы в условии 80-18 не следуют этой модели. В общем, "знающие" испытуемые используют правило комбинации, которое подобно МТБ, но оно несколько избыточно по отношению к МТБ.

Гипотеза тенденции угадывания в многоступенчатом выводе явно подержана "незнающими" испытуемыми. Очевидно, из-за сложности ситуации некоторые испытуемые были склонны сосредотачиваться почти полностью на наиболее вероятном событии в последующих стадиях вывода. Эффект угадывания в многоступенчатом выводе, как и консерватизм в выводе с одной стадией (W. Edwards, 1966), по-видимому, является другим примером общей неспособности сочетать сложную информацию. Так как большая часть процесса обработки информации людей является многоступенчатой и вероятностной по характеру, дальнейшим шагом в применении теоремы Байеса является нахождение способов, как помешать людям совершать ошибки, игнорируя все, кроме наиболее вероятных событий промежуточного уровня.

27. Заключение о личных характеристиках на основе информации, восстановленной из памяти*

Яков Троуп

Социальные оценки часто основываются на неточном вспоминании поведения других людей. Нас часто просят сделать выводы относительно характерных черт других людей, даже когда мы не уверены, что можем вспомнить, как они в действительности вели себя. При таких обстоятельствах нам приходится основывать нашу атрибуцию на неопределенном поведенческом доказательстве. Настоящее исследование посвящено изучению того, действительно ли люди включают этот источник неопределенности в их оценки других людей и как они это делают....

Модель Байеса для выводов из запомненного поведения

В настоящем исследовании испытуемые оценивали вероятности того, что некоторые субъекты обладают определенными признаками, на основе собственной ненадежной выборки поведения субъектов. Чтобы проверить гипотезу, что эти субъективные вероятности будут несправедливо высокими и недостаточно чувствительными к надежности, их сравнили с оптимальными вероятностями. Последние были получены из модели Байеса, которая была разработана исследователями многоступенчатого вывода вероятности (см. Peterson, 1973). Начинаясь с одноступенчатого вывода от действительного поведения B_j до характеристик личности A_i , теорема Байеса принимает следующую привычную форму

$$P(A_i / B_j) = P(A_i)P(B_j / A_i) / P(B_j) \quad (1)$$

где (A_i / B_j) – апостериорная вероятность A_i , учитывая B_j ; $P(A_i)$ – априорная вероятность A_i , $P(B_j)$ – полная вероятность B_j , $P(B_j / A_i)$ – условная вероятность B_j , при условии A_i . Последняя вероятность представляет *диагностическое значение* поведения, принимая во внимание характерные признаки актора. Апостериорная вероятность $P(A_i / B_j)$ может быть получена для каж-

* Данные впервые появились в The Journal of Personality and Social Psychology, 1978, 36, 93-106. Авторское право © 1978 принадлежит Американской психологической ассоциации. Перепечатано в соответствии с разрешением.

дого акта поведения, который может выбрать актер. Наблюдатели не знают, какой акт поведения был выбран в действительности. Они могут только вывести вероятности поведения из того, что они действительно помнят, V^* . Учитывая это заключение, теорема Байеса принимает вид

$$P(V_j / V^*) = P(V_j)P(V^* / V_j) / P(V^*) \quad (2)$$

Где $P(V_j / V^*)$ – апостериорная вероятность действительного поведения V_j , учитывая, что V^* было вспомнено; $P(V_j)$ – априорная вероятность V_j , $P(V^*)$ – полная вероятность того, что запомнили V^* , учитывая, что действительное поведение человека – V_j . $P(V_j / V^*)$ отражает *достоверность* – чем выше вероятность того, что выбрано действительное поведение, тем выше достоверность.

Модель Байеса двухступенчатого вывода сочетает значения $P(A_i / V_j)$, полученные для каждого действительного акта поведения через Уравнение 1, в оцененное среднее, с апостериорной вероятностью каждого акта поведения, $P(V_j / V^*)$, полученного из Уравнения 2, служащего как оценка. Таким образом, апостериорная вероятность характерного признака, учитывая, что было выбрано определенное поведение $P(V_j / V^*)$, может быть сформулировано:

$$P(A_i / V^*) = \sum_j P(A_i / V_j) P(V_j / V^*) \quad (3)$$

Уравнение 3 выражает количественно нормативные соображения, которые были обсуждены ранее. Так как $P(A_i / V_j)$ является возрастающей функцией $P(V_j / A_i)$ (диагностичность) и $P(V_j / V^*)$ является возрастающей функцией $P(V^* / V_j)$ (достоверность), $P(A_i / V^*)$ должна возрасти, как с достоверностью, так и с диагностичностью, и эффект одной переменной должен зависеть от уровня другой. Рассмотренное в терминах теории информационной интеграции Андерсена (1974), каждое выбранное поведение имеет определенный вес в оценивании вероятностей признаков. В свою очередь, вес рассматривается как мультипликативная функция диагностичности поведения и достоверности выборки. Правило интеграции, приписанное Уравнением 3, может быть применено как к одноступенчатому оптимальному выводу (т.е. значения $P(A_i / V_j)$ и $P(V_j / V^*)$, полученные из Уравнения 1 и 2, соответственно), так и к *субъективному* одноступенчатому выводу (т.е. собственные оценки испытуемых $P(A_i / V_j)$ и $P(V_j / V^*)$). В первом случае и одноступенчатые выводы, и их интеграция являются оптимальными (т.е. Байеса), а во втором случае оптимальна только интеграция. Настоящее исследование сравнивает наблюдаемые вероятности с двумя видами спрогнозированных вероятностей. Испытуемые делали выводы о зачислении студентов в аспирантуру и о политических взглядах студентов. Первые выводы

были основаны на вспоминании оценок студентов, а вторые – на вспоминании поведения студентов.

Эксперимент 1

Метод

Введение. Сначала испытуемым предложили оценки студентов на старшем курсе. Позднее им пришлось вспоминать оценки и оценивать вероятность того, что каждый студент был зачислен в аспирантуру школы бизнес-управления. Эта задача подразумевает 2 логических шага: (а) от вспомненной оценки (G^*) к реальной оценке студента (G); (б) от реальной оценки к тому, был ли студент зачислен или нет в аспирантуру (A). Неопределенностью последнего шага управляли посредством изменения диагностического значения оценок, принимая во внимание зачисление в школу бизнес-управления. Оценки низкой диагностичности были выпускными оценками по предмету «Введение в антропологию», а оценки высокой диагностичности были выпускными оценками по предмету «Введение в экономику». Надежность воспоминаний управлялась за счет изменения числа студентов, чьи оценки необходимо было вспомнить.

Процедура. Испытуемым сказали, что в определенный год было принято 50% претендентов на поступление в школу бизнес-управления. Таким образом, априорная вероятность быть принятым и не быть принятым, $P(A_y)$ и $P(A_n)$ была 0.50. Испытуемым сказали, что они будут оценивать вероятность зачисления для студентов, выбранных наугад. Им также сообщили, что оценки будут основаны на том, были ли оценки студента в данном предмете ниже 75 (G_1) или выше 75 (G_n) по шкале от 0 до 100.

Манипуляции с диагностическим значением. Два распределения вероятностей (каждое представлено в виде столбчатой диаграммы) соотнесли оценки и зачисление. Одна диаграмма показывала доли зачисленных студентов, которые получили G_1 или G_n . Эти доли представляли условные вероятности $P(G_1/A_y)$ и $P(G_n/A_y)$. Другая диаграмма показывала пропорцию не зачисленных студентов, которые получили эти оценки, то есть $P(G_1/A_n)$ и $P(G_n/A_n)$. Диаграмма для оценок высокой диагностичности (курс по экономике) показывала следующие доли: $P(G_1/A_n) = P(G_n/A_y) = 0.85$ и $P(G_1/A_y) = P(G_n/A_n) = 0.15$. Диаграмма для оценок низкой диагностичности (курс по антропологии) показывала следующие доли: $P(G_1/A_n) = P(G_n/A_y) = 0.55$ и $P(G_1/A_y) = P(G_n/A_n) = 0.45$. Каждый испытуемый оценивал две выборки студентов, для одной были предложены оценки высокой диагностичности, а для другой — оценки низкой диагностичности. Порядок представления был сбалансирован среди испытуемых.

Манипуляции с достоверностью. Обширное предварительное испытание показало, что определенный уровень достоверности может быть получен,

за счет изменения количества студентов, чьи оценки необходимо вспомнить. Испытуемым сообщили, что половина выборки получила низкие оценки, а другая половина – высокие, и что им предложат список студентов, получивших высокие оценки. Для того чтобы сделать минимальным различие между пунктами этого списка стимула, студенты были обозначены двузначными числами. Числа произносились магнитофоном со скоростью одно число в секунду. Высокий тон означал конец каждого списка стимула. Двумя секундами позже список испытания представлялся с помощью магнитофона. Список испытания состоял из 8 двузначных чисел, 4 из которых встречались в списке стимулов (т.е. числа G_h) и 4 не встречались в списке (т.е. числа G_l). Четыре числа G_h в списке испытания были взяты из полного ряда серийных позиций в списке стимула. Порядок чисел G_l и G_h был случайно определен для каждого списка испытания. Каждые 10 секунд проходило одно испытание. В промежутках испытуемым было необходимо указать, является ли оценка студента G_l или G_h и оценить вероятность того, что студент был зачислен в аспирантуру ($P(A_y / G^*)$), и вероятность, что он не был зачислен ($P(A_n / G^*)$). Предварительное испытание показало, что 10-секундного промежутка вполне достаточно, чтобы сделать оценки.

Использовались три вида списков: из 4, 7 и 14 пунктов. Данные, полученные из предварительного испытания, показывают что процент правильных опознаваний варьировался от 60% в длинном списке (условие низкой достоверности), 75% в среднем списке (условие средней достоверности) до 90% в коротком списке (условие высокой надежности). Каждый испытуемый отвечал на список высокой и низкой диагностичности. Наборы чисел в этих списках были уравновешены с диагностичностью. Двадцать испытуемых были распределены к каждому из трех условий надежности.

Условие полной достоверности. В этом условии одноступенчатого вывода группа из 46 испытуемых оценивала вероятность зачисления, исходя из оценок, которые были точно известны, $P(A_y / G)$ и $P(A_n / G)$. Испытуемым предложили буклет с диаграммами (т.е. информацию диагностичности) и оценки студентов. После ознакомления с оценками студентов, испытуемые отмечали оценки вероятности. Каждый испытуемый оценивал студентов с оценками низкой и высокой диагностичности. Для половины испытуемых оценки были низкими, а для другой половины – высокими.

Испытуемые. Всего в эксперименте принимало участие 106 человек (42 мужчины и 64 женщины), проходивших вводный курс по психологии в Еврейском университете в Иерусалиме. Участие в эксперименте было частичным выполнением требований курса.

Результаты и обсуждение

Доли верно опознанных низких оценок, $P(G^*_l / G_l)$, и высоких оценок, $P(G^*_h / G_h)$, служили мерами достоверности каждого испытуемого. Средние

величины $P(G^*_1 / G_1)$ и $P(G^*_h / G_h)$ показали ближайший случайный уровень выполнения в условиях низкой достоверности (список из 14 пунктов) ($M = 0.54$ и 0.60 соответственно), хорошее выполнение в условиях высокой достоверности (список из 4 пунктов) ($M = 0.93$ и 0.89 , соответственно), и среднее выполнение в условиях средней достоверности (список из 7 пунктов) ($M = 0.73$ и 0.77 , соответственно). Анализ изменения (Достоверность \times Диагностичность) этих долей выявил высоко значимые эффекты достоверности, $F(2,57) = 37.06$, $p < 0.001$, и $F(2,57) = 20.17$, $p < 0.001$ на $P(G^*_1 / G_1)$ и $P(G^*_h / G_h)$, соответственно. Другие источники изменений не были значимы.

Данные, полученные из условия полной достоверности, ясно показывают эффект влияния манипуляции с диагностичностью на *одноступенчатый* вывод из оценок, которые точно известны, до вероятности зачисления, $P(A/G)$. Испытуемые вывели более высокие $P(A_n/G_1)$ и $P(A_y/G_h)$ из оценок высокой диагностичности ($M = 0.81$ и 0.78 , соответственно), чем из оценок низкой диагностичности ($M = 0.57$ и 0.55 , соответственно), различие было значимым, как для $P(A_n/G_1)$, так и для $P(A_y/G_h)$, $t(22) = 8.08$, $p < 0.001$ и $t(22) = 8.59$, $p < 0.001$, соответственно.

Теперь обратимся к оптимальной апостериорной вероятности зачисления, учитывая вспомненную оценку, $P(A/G^*)$. Эти вероятности были определены для каждого испытуемого, объединив вероятности действительных оценок, учитывая вспомненные оценки, $P(G/G^*)$ и вероятность зачисления, учитывая действительные оценки, $P(A/G)$, в соответствии с Уравнением 3. Значения $P(G/G^*)$ были подсчитаны для каждого испытуемого с помощью Уравнения 2 из собственной достоверности при вспоминании оценок $P(G^*/G)$. Значения $P(A/G)$ были средними значениями оценок, полученных в условиях полной достоверности. Таким образом, оптимальная апостериорная вероятность зачисления, учитывая вспомненные оценки, $P(A/G^*)$, основывалась на субъективном одноступенчатом выводе от действительных оценок до вероятности зачисления, и на оптимальном одноступенчатом выводе (Уравнение 2) от вспомненных оценок к действительным оценкам. Пунктирные линии в левой части Рисунке 1 представляют средние значения $P(A_n/G^*_1)$, оптимальную вероятность того, что студент не был зачислен, учитывая, что его оценка была вспомнена как низкая; пунктирные линии в правой части Рисунке 1 представляют средние значения $P(A_y/G^*_h)$, оптимальную вероятность того, что студент зачислен, учитывая, что его оценка была вспомнена как высокая.¹ Ясно, что согласно модели Байеса, эти вероятности возрастут вместе с достоверностью и диагностичностью, и что эффект достоверности будет виден при высокой диагностичности, а не при низкой.

¹ Другой набор оптимальных значений $P(A/G^*)$ был подсчитан, исходя из оптимальных значений $P(A/G)$, полученных из Уравнения 1. Эти вероятности не рассматриваются отдельно, так как они были близки к оптимальным значениям $P(A/G^*)$, полученным из субъективных оценок $P(A/G)$.

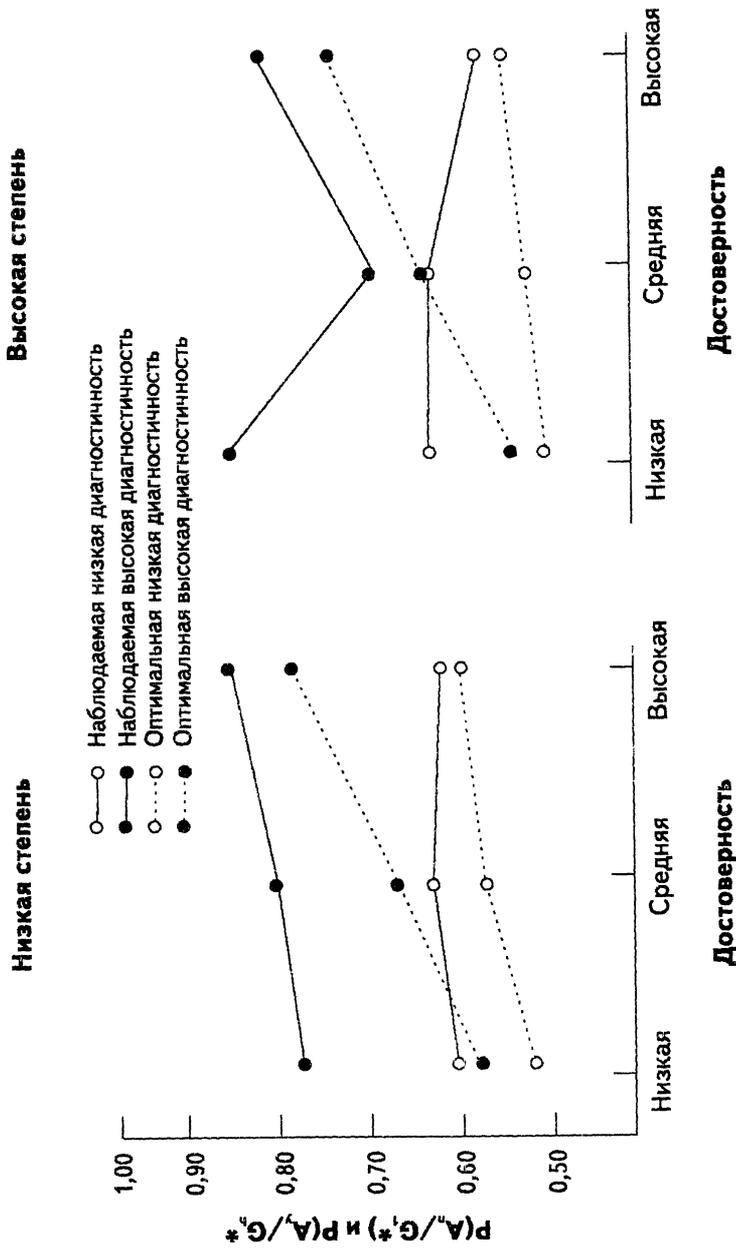


Рис. 1. Наблюдаемая и оптимальная вероятности ($P(A_n/G_1^*)$ и $P(A_n/G_n^*)$) как функция достоверности и диагностичности (Эксперимент 1)

Два наблюдаемых значения были определены для каждого испытуемого: первое – среднее значение $P(A_n/G_1^*)$, которое он оценил для студентов, которые, как он помнит, имели высокие оценки. Сплошные линии на Рисунке 1 показывают, что эта вероятность отличается как качественно, так и количественно от оптимальной вероятности. Диагностическое значение действительной оценки является единственным фактором, который имел постоянное влияние на наблюдаемые $P(A_n/G_1^*)$ и $P(A_y/G_h^*)$, $F(1,57) = 96.97$, $p < 0.001$ и $F(1,57) = 114,82$, $p < 0.001$, соответственно. Чем выше диагностическое значение действительных оценок, принимая во внимание зачисление, тем более крайними являются апостериорные вероятности. Достоверность, с которой были узнаны действительные оценки, имела очень маленькое и непоследовательное влияние на выводы ($F < 1$). Даже выводы на двух крайних уровнях достоверности в пределах каждого уровня диагностичности значительно не отличались. Другими словами, выводы не были менее крайними, когда степень правильных узнаваний была близка к случайному уровню, а не когда она была почти совершенной. Выводы испытуемых также не показывали взаимодействия Диагностичность – Достоверность, спрогнозированную моделью Байеса. Влияние диагностичности действительных оценок не уменьшалось с уменьшением достоверности. Согласованное влияние этих переменных было проверено на уровне выводов о студентах. Нормативно, изменение в уверенности в оценках различных студентов должно повлиять в большей степени на выводы об их зачислении, когда оценки обладают высоким диагностическим значением, чем когда их диагностическое значение – низкое. Следовательно, выводы о студентах, внесенных в список высокой диагностичности, должны больше различаться, чем выводы о студентах, внесенных в список низкой диагностичности. Изменение $P(A/G^*)$ для 8 студентов в пределах каждого из двух списков, было подсчитано для каждого испытуемого. В отличие от модели Байеса, эти изменения значимо не различались, $t(59) = 1.63$.

Рисунке 1 также показывает, что наблюдаемые вероятности были более крайними, чем оптимальные вероятности. Среднее различие между наблюдаемыми и оптимальными вероятностями было существенным как для $P(A_n/G_1^*)$, так и для $P(A_y/G_h^*)$, $F(1,57) = 33.38$, $p < 0.001$, и $F(1,57) = 84.67$, $p < 0.001$, соответственно. Для испытуемых были подсчитаны коэффициенты корреляции между оптимальными и наблюдаемыми вероятностями в пределах каждого из 6 условий. Корреляции были незначительными и непоследовательными, 5 корреляций были положительными и 7 — отрицательными. Две корреляции были значимыми, но отрицательными ($r = -0.50$, $p < 0.05$ и $r = -0.55$, $p < 0.01$ для $P(A_n/G_1^*)$ и $P(A_y/G_h^*)$, соответственно, в условиях низкой и высокой диагностичности).

Эти результаты показывают, что испытуемые не смогли учесть соображения достоверности в своих выводах вероятности. Испытуемые основывали свои выводы только на диагностическом значении действительных оценок. Вероятности, выведенные в экспериментальных условиях, были

несколько ниже, чем вероятности, выведенные в условии полной достоверности, показывая, что испытуемые не полностью усвоили стратегию “как будто” (т.е. делать выводы, как будто действительные оценки были точно известны). Однако уменьшение уверенности было маленьким и постоянно не изменялось с изменением достоверности. Как результат, в сравнении с вероятностями Байеса, выводы испытуемых были крайними. Заметим, что это было не так при одноступенчатом выводе от действительных оценок до зачисления, $P(A/G)$. Вероятности, выведенные при условии полной достоверности, были близки к оптимальным вероятностям, выведенным из модели одноступенчатого вывода Байеса (Уравнение 1). Этот результат предполагает, что неоптимальность (или самонадеянность) в условии двухступенчатого вывода, появилась благодаря введению дополнительного источника неопределенности – несовершенному вспоминанию действительных оценок...

Эксперимент 4

Метод

В этом эксперименте испытуемые оценивали вероятности того, что студенты мужского пола выступают за (A_p) или против (A_n) возвращения Западного берега (территории, занятой Израилем в войне 1967 года) арабам. Голосование за лейбористскую партию (относительно миролюбивую партию) или за партию Ликуд (Likud) (достаточно воинственную партию) служило как информация низкой диагностичности. Предварительный эксперимент показал, что короткие волосы и голосование за партию Ликуд (B_n) подразумевало отрицательное отношение к возвращению Западного берега, а длинные волосы и голосование за лейбористов (B_p) являлось показателем положительного отношения. Предварительный эксперимент также показал, что голосование студентов является более показательным, чем длина волос.

Задача на вспоминание и другие аспекты проекта и процедуры были такими, как и в Эксперименте 1. 20 испытуемых были распределены по каждому из трех условий достоверности.

Дополнительная группа, состоящая из 57 испытуемых, была распределена по условию полной достоверности. При этом условии испытуемые точно знали длину волос студентов или партию, за которую они голосовали. Каждый испытуемый оценивал вероятности отношения, исходя из ношения коротких волос, длинных волос, голосования за лейбористов и голосования за партию Ликуд.

В общей сложности в эксперименте 4 приняло участие 117 испытуемых (51 мужчина и 66 женщин), которые проходили вводный курс по психологии в Еврейском университете в Иерусалиме.

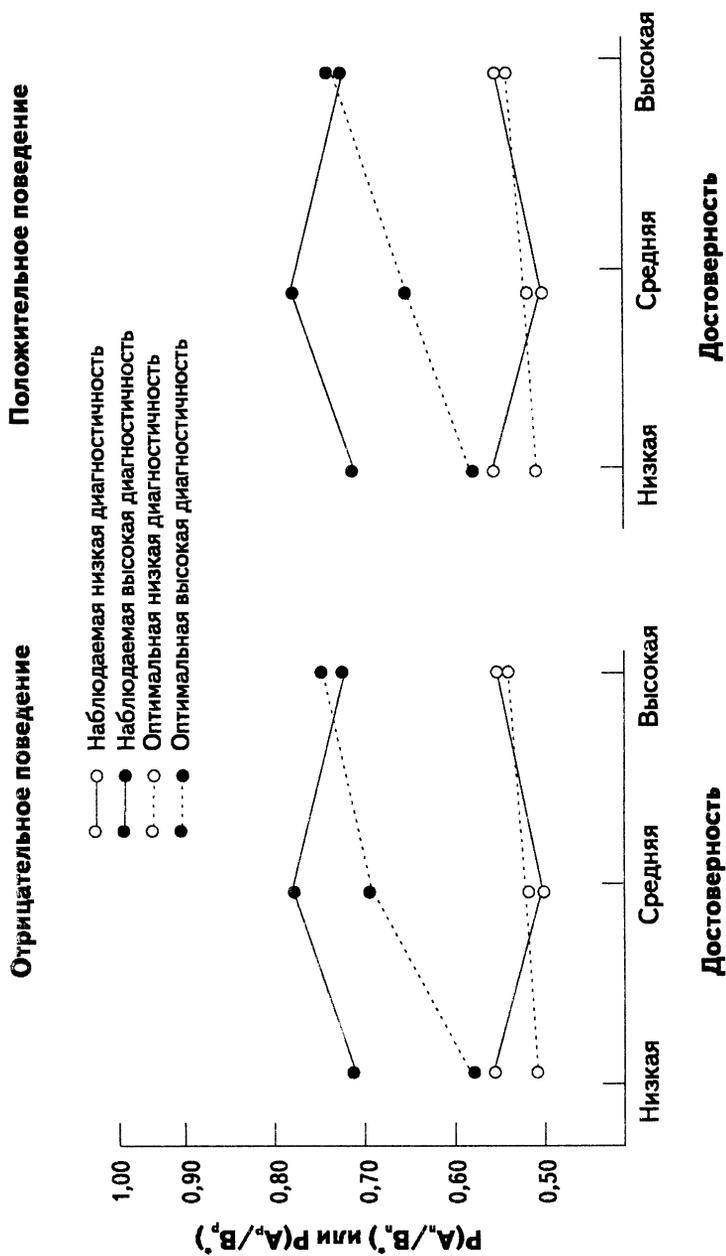


Рис. 2. Наблюдаемая и оптимальная $P(A_n/B_n^*)$ и $P(A_p/B_p^*)$ как функция достоверности и диагностичности (Эксперимент 4)

Результаты и обсуждение

Манипуляция с достоверностью имела сильное влияние на степень верных узнаваний как B_n (т.е. короткие волосы или голосование за партию Ликуд), так и B_p (т.е. длинные волосы и голосование за лейбористов), $F(2,57) = 55.45$, $p < 0.001$ и $F(2,57) = 22.04$, $p < 0.001$, соответственно. Средние значения $P(B_p^*/B_p)$ были 0.61, 0.82 и 0.91. Эффект диагностичности и его взаимодействие с достоверностью были незначимыми.

Как и ожидалось, значения $P(A/B)$, выведенные испытуемыми в условиях полной достоверности показывают, что субъективное диагностическое значение голосования было большим, чем длина волос. Таким образом, вероятности отношений $P(A_n/B_n)$, выведенных из того факта, что студенты, голосовавшие за партию Ликуд, ($M = 0.78$) были более крайними, чем те, которые были выведены из того факта, что студенты носили короткие волосы ($M = 0.52$), $t(56) = 13.61$, $p < 0.001$. Подобным образом, вероятность отношения $P(A_p/B_p)$, выведенная из факта, что студенты голосовали за лейбористов ($M = 0.75$) была более крайней, чем те, которые были выведены из факта, что студенты носили длинные волосы ($V = 0.53$), $t(56) = 9.89$, $p < 0.001$.

Оптимальные и наблюдаемые вероятности показаны на рис.2. Первые были получены от каждого испытуемого через Уравнение 3 от значений $P(B/V^*)$ (рассчитанные через Уравнение 1 от достоверности испытуемого в определении поведения студентов, $P(B^*/V)$) и из средних значений $P(A/B)$, полученных в условиях полной достоверности. Можно видеть, что только диагностичность поведения учитывалась и обеспечивала выводы вероятностей отношения $F(1,57) = 68.94$, $p < 0.001$, и $F(1,57) = 90.40$, $p < 0.001$ для $P(A_n/B_n^*)$ и $P(A_p/B_p^*)$, соответственно.² При нарушении модели Байеса основное влияние достоверности и её взаимодействие с диагностичностью были незначимы. Кроме того изменения наблюдаемых вероятностей в пределах списков высокой диагностичности были не больше, чем изменения в пределах списков низкой диагностичности, $t(59) = 0.59$. Как и в предыдущих экспериментах, наблюдаемые вероятности были более крайними, чем оптимальные вероятности $F(1,57) = 5.68$, $p < 0.05$ и $F(1,57) = 27.63$, $p < 0.001$ для $P(A_n/B_n^*)$ и $P(A_p/B_p^*)$, соответственно. Наконец, из двенадцати корреляций для испытуемых при условиях между наблюдаемыми и оптимальными вероятностями восемь были положительными, но только две были значимыми: $r = 0.436$, $p < 0.05$ для $P(A_n/B_n^*)$ при условии высокой диагностичности и высо-

² Необходимо отметить, что эти результаты не могут быть приписаны к тому факту, что диагностичностью управляли как внутригрупповой переменной, а достоверностью как межгрупповой переменной. Был проведен дисперсионный анализ по выводам первого списка студентов, чье поведение обладало низким диагностическим значением для половины испытуемых в каждом условии достоверности и высоким диагностическим значением для другой половины. В этом анализе диагностичность рассматривалась как межгрупповой фактор. Эти анализы показали значимое влияние диагностичности и незначимое влияние достоверности и обоих факторов вместе (диагностичность \times достоверность). Дисперсионный анализ данных предыдущих трёх экспериментов показал похожие результаты.

кой достоверности и $r = 0.402$, $p < 0.05$ для $P(A_p/B_p^*)$ при условии низкой диагностичности и высокой достоверности.

Эксперимент 4 обеспечивает дополнительную поддержку для обобщения наших выводов. Он показывает, что смещения, наблюдаемые в предыдущих экспериментах, также влияют на выводы о характеристиках личности за счет вспомненного поведения. Диагностичность действительного поведения субъектов являлась единственным фактором, определяющим выводы даже когда диагностичность должна основываться на собственных убеждениях испытуемых без получения свидетельства извне.

Общее обсуждение

Выводы, которые наши испытуемые сделали исходя из собственных ненадёжных воспоминаний, систематически нарушали нормативные правила вывода вероятности. Испытуемые были склонны полагаться на диагностичность информации, не принимая во внимание достоверность, не смотря на тот факт, что они знали о недостоверности своих воспоминаний.

Действительно, сравнения между выводами, исходя из верно вспомненной информации (например, выводы исходя из оценок, которые были вспомнены как низкие и которые действительно были низкими) и выводы, исходя из неверно вспомненной информации (например, выводы исходя из оценок, которые были вспомнены как низкие, но которые в действительности были высокими), показали незначительные различия во всех четырёх экспериментах. То есть последние выводы вероятности были не менее крайними, чем первые. Эти результаты соответствуют точки зрения, что люди склонны уменьшать сложность задач, используя простые правила эвристики. Недооценивание информации достоверности может отражать использование одной такой эвристики суждения — эвристики репрезентативности (Kahneman & Tversky, 1973, 4). Например, вспоминая партию, за которую голосовал субъект, испытуемые оценивали вероятность того, что он является сторонником возвращения западного берега арабам, по степени, в которой голосование за такую партию репрезентативно стереотипу субъекта, поддерживающего подобное мнение.

Субъект, голосующий за милитаристскую партию, олицетворяет больше черт этого стереотипа, чем субъект, который носит длинные волосы. (И в первых трёх экспериментах высокая оценка по экономике является более репрезентативной для студентов, зачисленных в школу управления бизнесом, чем высокая оценка по антропологии.) Положительное отношение к возвращению территории было выведено с большей уверенностью в первом случае. Точность воспоминания не влияла на эти вероятности, потому что она не затрагивала оцененную репрезентативность вспомненного поведения.

Принимая это во внимание, достоверность доклада сравнима с соображениями размера выборки и базовыми значениями. Размер выборки, на которой основывается свидетельство и базовое значение выведенного отноше-

ния не влияют на подобие между свидетельством и признаком. Понятие репрезентативности предполагает, и исследование показало, что люди не обращают внимания на эти соображения, делая выводы (Nisbett & Borgida, 1975, Tversky & Kahneman, 1974, 1). На основании этого можно ожидать, что достоверность будет приниматься во внимание в тех случаях, когда она влияет на репрезентативность доклада. Таким образом, неверное воспоминание реальных поведенческих особенностей (благодаря таким факторам, как промежуток времени, с которого особенность была зарегистрирована), которые детализированы больше, чем простые бинарные события, включённые в настоящее исследование, вероятно, будет влиять на ясность, полноту и яркость вспоминаемой информации. Такая неясная информация может быть менее репрезентативна рассматриваемому признаку, и соответственно, производит более регрессивные и более умеренные оценки вероятности ...

Часть VIII
Коррективные процедуры

28. Устойчивая привлекательность ошибочных линейных моделей при принятии решений*

Робин М. Дауэс

Книга Поля Мила (Paul Meehl) *Клиническое прогнозирование в сравнении со статистическим: теоретический анализ и обзор доказательств* (Clinical Versus Statistical Prediction: a Theoretical Analyses and a review of the Evidence) появилась 25 лет назад. В ней рассматриваются доказательства того, что числовые критериальные переменные психологической направленности (например, рейтинги способностей аспирантов, которые получили докторскую степень) от числовых предсказывающих переменных (например, оценки на выпускном экзамене, уровень средних оценок, рейтинг рекомендательных писем) лучше осуществляются посредством верных линейных моделей, чем клинической интуицией людей, имеющих опыт в таких предсказаниях. Основная идея этой статьи — это рассмотреть доказательства того, что даже ложные линейные модели могут превосходить клинические предсказания.

Верная линейная модель — это та, в которой веса, приписанные предсказывающим переменным, выбраны таким образом, чтобы оптимизировать отношение между предсказанием и критерием. Простой регрессионный анализ — это наиболее обычный пример верной линейной модели; предсказывающие переменные взвешиваются таким образом, чтобы максимизировать корреляцию между результатом по формуле и действительным значением критерия. Дискриминантный анализ — это другой пример верной линейной модели, веса приписываются предсказывающим переменным таким образом, чтобы результирующие линейные формулы максимизировали различия между двумя или более группами. Регрессионный анализ гребня (Darlington, 1978, Marquardt & Snee, 1975), другой пример, попытки приписать веса таким образом, что итоговые значения максимально коррелируют с реальным позателем в новом наборе данных.

Таким образом, существует множество типов верных линейных моделей, и они используются в различных контекстах. Один пример (Dawes, 1971) рассматривает прогнозирование факультетских рейтингов аспирантов. Все

* Этот раздел впервые появился в *American Psychologist*, 1978, 34, 571-582 Авторское право © 1979 принадлежит Американской психологической ассоциации. Перепечатано в соответствии с разрешением.

аспиранты отделения психологии Университета штата Орегон, которые были приняты в период с осени 1964 по осень 1967 – и те, кто выбыл из программы по неакадемическим причинам (например, психическое заболевание или вступление в брак) – были оценены весной 1969 года; сотрудники факультета оценили только тех студентов, в ком они были уверены. Использовалась следующая оценочная шкала: 5 – выдающийся, 4 – выше среднего, 3 – средний, 2 – ниже среднего, 1 – выбывший из программы по академическим трудностям. Такие оценивания являются психологически интересными критериями, потому что субъективные впечатления сотрудников факультета являются главными детерминантами работы (если он идет работать), которую студент получает после окончания аспирантуры. В качестве выборки выступили 111 студентов, количество сотрудников факультета, оценивающих каждого из этих студентов, варьировалось от 1 до 20, со средним – 5.67 и медианой – 5. Оценки были надежными. (Чтобы определить надежность, оценки были подвергнуты однофакторному дисперсионному анализу, при котором каждый оцениваемый студент принимался за единицу анализа. Результирующее отношение межгрупповых дисперсий (n^2) было 0.67, и находилось ниже уровня значимости 0.001). Эти факультетские рейтинги прогнозировались по проверенной линейной модели, основанной на оценках выпускных экзаменов, средних оценок на старших курсах и меры избирательности студентами образовательных курсов.¹ Подтвержденная перекрестными исследованиями множественная корреляция между факультетским рейтингом и предсказываемыми переменными была 0.38. В соответствии с результатами Мила, корреляция последних факультетских рейтингов со средними оценками сотрудников приемной комиссии, которая отбирала студентов составила 0,19², то есть она объясняла одну четвертую общей дисперсии. Этот пример типичен для психологических исследований в данной области: (а) корреляция с предсказаниями модели выше, чем корреляция с клиническим прогнозом, но (б) обе корреляции низкие. Подобные характеристики часто приводят психологов к следующей интерпретации вывода: в то время, как низкая корреляция модели показывает, что линейное моделирование является несостоятельным как метод, низкие корреляции экспертов означают лишь то, что использовались плохие эксперты.

¹ Этот индекс основывался на оценке избирательности Cass & Birnbaum (1968), в их книге «Сравнительное руководство к американским колледжам» (Comparative Guide to American Colleges). Этим устным категориям избирательности приписали численные значения, в соответствии со следующим правилом: наиболее избирательные – 6, сильно избирательные – 5, очень избирательные(+) – 4, очень избирательные – 3, избирательные – 2, не упомянутые – 1.

² К сожалению, только 23 из 111 студентов можно использовать в этом сравнении, потому что оценочная шкала, используемая комиссией, слегка изменялась от одного года к другому.

Ложная линейная модель — это та, в которой веса выбраны неким неоптимальным методом. Они могут быть выбраны с условием равенства, они могут быть выбраны, основываясь на интуиции человека, делающего прогноз, или они могут быть выбраны случайным образом. Тем не менее, ложные модели могут быть очень полезными. Когда, например, стандартизированные оценки выпускных экзаменов, средние оценки на старших курсах и индексы избирательности в предыдущем примере были равно оценены, итоговый результат коррелировал как 0.48 с последним факультетским рейтингом. Корреляция была не только выше, чем при клиническом оценивании комиссии (0.19), но она также была выше, чем полученная на основе значений, взятых из половины выборки.

Пример ложной модели, который может быть интересен — по крайней мере, широкой публике — был инициирован врачом, который входил в состав комиссии вместе со мной, когда я занимался системой прогнозирования. Позднее, находясь в баре со своей женой и со мной, он сказал, что моя статья могла быть интересна моим коллегам, но успешное обучение в аспирантуре в области психологии не представляет особого интереса: "Мог бы ты, например, использовать одну из своих ложных моделей, чтобы предсказать, насколько хорошо мы ладим с женой" — сказал он. Я понял, что мог бы или возможно смог бы. В то время психологическое отделение университета штата Орегон занималось гендерными исследованиями, большинство из которых были поведенчески ориентированы. Таким образом, участники этих исследований фиксировали, когда они занимались любовью, когда они ссорились и когда они встречались (например, с родственниками супруга/супруги) и т.д. Эти испытуемые также делали субъективные оценки относительно того, насколько счастливы они были в семейной жизни. Я сразу же подумал о ложных линейных моделях, чтобы спрогнозировать самооценки семейного счастья: количество раз занятия любовью минус количество ссор. Мой коллега Джон Ховард (John Howard) собрал такие данные о супружеских парах, когда он был аспирантом в университете штата Миссури, Канзас-Сити, где он работал с Александер (Alexander, 1971). После того, как была установлена достоверность оценивания занятий любовью и ссор, Александер просила одного супруга из 42 пар контролировать эти события. Она позволила нам анализировать данные, которые были получены: "В 30 парах, которые счастливы в браке (по мнению контролирующего супруга), только две ссорились чаще, чем занимались любовью. Все 12 пар, несчастливых в браке, ссорились гораздо чаще" (Howard & Dawes, 1976, с.478). Мы повторили это исследование в университете штата Орегон, где 27 контролирующих супругов оценили счастье по 7-бальной шкале, от "очень несчастны" до "очень счастливы" с нейтральным средним баллом. Корреляция частоты занятий любовью минус частота ссор с этими оценками супружеского счастья составила 0.40 ($p < 0.05$), ни одна из переменных не была значима сама по себе. Исследование было повторено в штате Миссури Эдвардсом и Эдвардсом (Edwards & Edwards, 1979) и в Техасе Торнтоном (Thornton,

1977а), которые обнаружили корреляцию 0.81 ($p < 0.01$) между различием секс-ссоры и самооценками супружеского счастья среди 28 новых пар. (Причина намного большей корреляции может быть в том, что Торнтон получил оценки супружеского счастья после, а не до того, как испытуемые фиксировали занятия любовью и ссоры; действительно, одна из испытуемых решила развестись, осознав, что она чаще ссорилась, чем занималась любовью; Thornton, 1977b). Вывод заключается в том, что если мы любим больше, чем ненавидим, мы счастливы; если мы ненавидим больше, чем мы любим, мы несчастны. Этот вывод не является глубоким, ни психологически, ни статистически. Однако эти грубые ложные линейные модели прогнозируют очень важную переменную: оценки семейного счастья.

Большой объем литературы (в действительности, весь) со времени публикации (1954) книги Мила поддерживает его обобщение о верных линейных моделях вместо интуитивных клинических оценок. Соьер (Sawyer, 1966) пересмотрел большое количество подобных исследований, а некоторые из этих исследований были весьма обширными (Goldberg, 1965). Через 10 лет после опубликования его книги, Мил (1965) смог сделать вывод, что существовал единственный пример, который показывал, что клинические оценки дают лучшие результаты, и это было сразу же оспорено Голдбергом (1968а) на основании того, что даже этот пример не показывает существенного превосходства. Холт (Holt, 1970) критиковал детали некоторых исследований, и он даже предположил, что прогноз, в отличие от понимания, не может являться важной частью клинического суждения. Однако обзор литературы не выявил других исследований, в которых клиническое оценивание превосходило статистическое прогнозирование, когда они были основаны на одинаковых исходных переменных. И хотя большинство противников позитивизма согласится, что понимание не синонимично прогнозированию, немногие оспорят то, что оно не подразумевает наличия способности прогнозировать.

Почему? Потому что люди – особенно эксперты в какой-либо области – намного лучше выполняют отбор и кодирование информации, чем ее интегрирование.

Но люди *действительно* важны. Статистическая модель может включать информацию оптимальным способом, но всегда индивиды (эксперты, клиницисты, испытуемые) выбирают переменные. Кроме того, именно эксперт знает направленные отношения между предсказывающими переменными и критерием и может закодировать переменные таким образом, что они имеют ясные направленные отношения. И именно в подобной ситуации, где предсказывающие переменные являются хорошими и где они имеют монотонные условные отношения с критерием, верная линейная модель хорошо функционирует³.

³ Отношения являются условными монотонными, когда переменные могут быть распределены на шкале таким образом, что более высокие значения на каждой прогнозируют более высокие значения по критерию. Это условие является комбинацией двух более фундамен-

Линейная модель не может заменить эксперта при решении вопроса "что искать", но именно знание, что искать, при достижении решения является именно той экспертизой, которую делают люди. Даже в такой сложной оценке как совершение хода в шахматах, именно способность закодировать доску правильным способом, чтобы "увидеть" верный ход, которая отделяет великого мастера от эксперта, а эксперта от новичка (deGroot, 1965; Simon & Chase, 1973). И не способностью интегрировать информацию различаются эти люди (Slovic, 1972b). Мастер рассматривает не больше ходов, чем эксперт; он просто знает, какие рассматривать. Разница между знанием того, что искать и способностью интегрировать информацию возможно лучше всего проиллюстрирована в исследовании Айнхорна (Einhorn, 1972). Врачи-эксперты закодировали биопсии пациентов, страдающих болезнью Ходжкина, а затем в целом оценили серьезность процесса. Эти оценки не спрогнозировали время жизни 193-х пациентов, все из которых умерли (корреляции оценок со временем жизни все были практически равны нулю, иногда даже в неверном направлении). Переменные, которые закодировали доктора, однако, предсказывали время жизни, когда они использовались во множественной регрессионной модели.

Таким образом, верные линейные модели работают по одной простой причине. Люди хорошо выбирают верные переменные и кодируют их, таким образом, что они имеют условное монотонное отношение с критерием. Однако люди плохо интегрируют информацию из разнообразных и несравнимых источников. Верные линейные модели хороши для таких интеграций, когда прогнозы имеют условное монотонное отношение к критерию.

Рассмотрим, например, проблему сравнения студента, желающего стать аспирантом, у которого средняя оценка выпускных экзаменов 750, а средняя оценка на старших курсах 3.3, с другим, у которого средняя оценка выпускных экзаменов 680, а средняя оценка на старших курсах 3.7. Большинство экспертов согласится, что эти показатели способностей и предыдущие достижения должны быть объединены неким компенсаторным способом, но вопрос, как компенсировать? Многие эксперты, которые пытались это разрешить, мало знали о характеристиках распределений средних оценок выпускных экзаменов и средних оценок на старших курсах, и большинство не знало об исследованиях, показывающих их надежность, как показателем успеха в аспирантуре. Кроме того, эти значения изначально не сравнимы без подобных знаний, оценки на выпускных экзаменах от 500 до 800

тальных условий измерения: (а) независимость (отношение между каждой переменной и критерием не зависит от значений оставшихся переменных) и (b) монотонность (порядковое отношение — это то, которое является монотонным). (см. Krantz, 1972, Krantz, et al., 1971). Истинные отношения не обязательно должны быть линейными, чтобы функционировали линейные модели; они должны быть просто приближены к линейным моделям. Верно то, что "для того, чтобы вычислить коэффициент корреляции между двумя переменными, отношение между ними должно быть линейным" (совет, обнаруженный в одном вводном тексте по статистике). Для начала всегда возможно что-либо вычислить.

для зачисленных претендентов на места в аспирантуре и средние оценки на старших курсах от 3.0 до 4.0. Интересно, статистическая схема оценивания функционирует в таких обстоятельствах лучше, чем эксперт?

Предположим, что в некоторых условиях невозможно построить верную линейную модель. Одна причина, по которой мы не сможем этого сделать, состоит в том, что размер нашей выборки является неадекватным. При множественной регрессии, например, b весовые коэффициенты являются как известно нестабильными; отношение наблюдений к предсказывающим переменным должно быть таким как 15 или 20 к 1 перед тем, как b веса, которые являются оптимальными, функционируют лучше на кросс-валидации, чем делают простые единичные весовые коэффициенты. Шмидт (Schmidt, 1971), Голдберг (1972) и Клауди (Claudy, 1972) эмпирически продемонстрировали эту необходимость посредством компьютерного моделирования, а Айнхорн и Хогарт (Hogarth, 1975) и Шринивизан (Srinivisan, 1977) рассмотрели проблему аналитически. Общее решение зависит от ряда параметров, таких как множественная корреляция в популяции и ковариация между предсказывающими переменными. Но практический вывод ясен. Стандартный регрессионный анализ не может использоваться в ситуациях, где нет "подходящей" пропорции наблюдений к предсказывающим переменным.

Другая ситуация, в которой верные линейные модели не могут использоваться, это та, в которой не существует измеряемых критериальных переменных. Мы могли бы знать, каковы были бы важные предсказывающие переменные фактора и направления, которые они задали бы критерию, *если бы мы могли измерить критерий*. Например, решая, какого студента принять в аспирантуру, нам бы хотелось предсказать долгосрочную переменную, которая может быть названа "профессиональная само-актуализация". Мы имеем представление о том, что понимать под этим понятием, но у нас пока нет точного хорошего определения. (И даже если бы оно было, невозможно провести исследование, используя записи данных студентов, потому что эта переменная может быть оценена только спустя, по крайней мере, 20 лет после написания студентами докторских диссертаций.) Однако мы знаем, что при всех вероятностях этот критерий положительно связан с интеллектом, с предыдущими достижениями и со способностью опередить своих коллег. В документах наших претендентов оценки выпускных экзаменов оценивали первую переменную, средние оценки на старших курсах — вторую, рекомендательные письма — третью. Разве бы нам не захотелось сформировать некий вид линейной комбинации этих переменных, чтобы оценить потенциалы претендентов? Учитывая, что мы не можем провести стандартный регрессионный анализ, разве ничего не остается, как вернуться к интуитивной интеграции без посторонней помощи этих переменных при оценивании наших претендентов?

Один возможный способ построения ложной линейной модели — это посредством использования *предыдущего значения для расчета последующего* (Dawes & Corrigan, 1974, L.R. Goldberg, 1970). Процесс — это построение вер-

ной линейной модели оценки эксперта о результирующем критерии и затем, использование этой модели вместо эксперта. На то, что подобная линейная модель может быть точной в предсказании оценок эксперта, указано в психологической литературе Хэммондом (Hammond, 1955) и Хоффманом (Hoffman, 1960). (Эту работу предвосхитил 32 годами ранее Генри Волейс (Henry Wallace), вице-президент во время правления Рузвельта, в 1923 году в сельскохозяйственной статье он предложил использование линейных моделей, чтобы проанализировать "что на уме у эксперта по кукурузе".) В своей влиятельной статье, Хоффман назвал использование линейных моделей *параморфической* репрезентацией эксперта, под которой он понимал, что психологические процессы эксперта не включают вычисление явного или существенного среднего значения входных переменных, но что оно может быть имитировано подобным оцениванием. Параморфические репрезентации были крайне успешными (для обзора, см. Dawes & Corrigan 1974, Slovic & Lichtenstein, 1971) в контекстах, в которых предсказывающие переменные имеют условные монотонные отношения к критериальным переменным.

Модели, вовлекающие предыдущее значение для расчета последующего, используют весовые значения, полученные от экспертов, потому что эти значения не получены из соотношения между предсказывающими переменными и критерием. Все же эти параморфические репрезентации последовательно лучше функционируют, чем эксперты, от которых они были получены (по крайней мере, когда оценивание качества осуществляется в терминах корреляции между предсказанными и действительными переменными).

Использование предыдущего значения для расчета последующего оказалось распространенным. Например, в исследовании, проведенном Виггинсом и Коухеном (Wiggins & Kohen, 1971), аспирантам в психологии университета штата Иллинойс предоставили 10 измерений способностей и индивидуальности, описывающих других (реальных) аспирантов по психологии штата Иллинойс, и их попросили предсказать средние оценки в аспирантуре на первом году обучения. Линейные модели экспертов в университете штата Иллинойс работали лучше, чем сами эксперты, предсказывая действительные оценки. Этот результат был также получен в исследовании, проведенном в сотрудничестве с Виггинсом, Грегори и Диллером (Wiggins, Gregory & Diller, цитированное в Dawes & Corrigan, 1974). Л.Р. Голдберг (L.R. Goldberg, 1970) продемонстрировали его для 26 из 29 экспертов в области клинической психологии, предсказывающих психиатрический диагноз невроза или психоза, исходя из данных Миннесотского Многофакторного Личностного Опросника (MMPI), а Даус обнаружил его в оценивании претендентов на поступление в аспирантуру штат Орегон. Единственным опубликованным исключением к успешному использованию предыдущего значения для расчета последующего, о котором я знаю, было исследование, проведенное Либби (Libby, 1976). Он просил 16 служащих кредитного бюро из относительно маленьких банков (расположенных в Шампейн-Урбана, Иллинойс, с активами от 3 до 56 миллионов долларов) и 27 служащих кре-

дитного бюро крупных банков (расположенных в Филадельфии, с активами от 6 миллиардов до 4.4 биллионов) оценить, какие 30 из 60 фирм обанкротятся в течение 3-х лет после их финансового утверждения. Служащие запросили 5 финансовых коэффициентов, на которых они основывали свои оценки (например, отношение настоящих активов к общей стоимости имущества). В среднем, служащие правильно оценили 44.4 предприятия (74%), как либо платежеспособные, либо будущих банкротов. Но в среднем, параморфические репрезентации служащих смогли верно охарактеризовать только 43.3 (72%). Эта разница оказывается статистически значимой, и Либби сделал вывод, что это был пример ситуации, где не работало использование предыдущего значения для расчета последующего – возможно, потому что его эксперты были хорошо подготовлены для определения высоко надежного критерия. Голдберг, однако, заметил, что многие отношения имели сильно ассиметричные распределения, и он пересмотрел данные Либби, нормализуя отношения перед тем, как построить модели служащих кредитного бюро. Либби обнаружил, что 77% служащих превосходили свои параморфические репрезентации, но Голдберг, используя измененные предсказывающие переменные, обнаружил противоположное: 72% моделей превосходили оценивающих, от которых они были получены.⁴

Почему работает использование предыдущего значения для расчета последующего? Баумен (Bowman, 1963), Голберг (1970) и Даус (1971) считали, что ее успех исходит из факта, что линейная модель извлекает основное содержание (в подразумеваемых значениях) от характеристик изменчивого поведения (например, оценки, подверженные влиянию эффектов контекста или внешних переменных).

Уверенность в эффективности использования предыдущего значения для расчета последующего была основана на построении валидности линейной модели эксперта с валидностью его оценок. Это только одно из двух возможных сравнений. Второе – это валидность линейной модели оценивающего против валидности линейных моделей в общем; то есть продемонстрировать, что использование предыдущего значения для расчета последующего функционирует, потому что линейная модель улавливает сущность валидной экспертизы оценивающего, исключая ненадежность; необходимо продемонстрировать, что значения, полученные от анализа поведения эксперта, превосходят значения, которые могли быть получены другими способами, например, случайно. Поскольку и модель эксперта, и модель, полученная случайно, являются совершенно надежными, сравнение случайной модели с моделью эксперта позволяет оценить основную линейную репрезентацию

⁴ Необходимо отметить, что верные линейные модели функционировали лучше, чем служащие либо их параморфические интерпретации. Используя ту же задачу, Бивер (Beaver, 1966) и Дикон (Deason, 1972) обнаружили, что линейные модели прогнозировали с точностью приблизительно 78% при кросс-валидации. Но я не могу не подчеркнуть, что самая простая ложная модель функционирует лучше всего. Отношение активов к пассивам верно охарактеризовали 48 (80%) случаев, изученных Либби.

оценивающего, или *содержание*. Если случайная модель функционирует одинаково хорошо, эксперт не будет "следовать валидным принципам, но следовать им недостаточно" (Dawes, 1971, с. 182), по крайней мере, не более валидным признакам, чем любые другие, которые оценивают переменные в правильном направлении.

Таблица 1 представляет 5 исследований, обобщенных Даусом и Корриганом (Dawes & Corrigan, 1974), в которых сравниваются валидности (то есть корреляции), полученные различными методами. В первом исследовании, 861 пациент заполнили ММРІ в различных психиатрических больницах, позднее они были разделены на невротических и психотических на основе более обширной информации. Данные ММРІ состоят из 11 множеств, каждое из которых представляет степень, в которой респондент отвечает на вопрос способом, похожим на тот, которым отвечают пациенты, страдающие явной формой патологии. С каждым пациентом связан набор из 11 множеств, и проблема – это предсказать, будет ли дальнейшим диагнозом психоз (обозначенный 1) или невроз (обозначенный 0). 29 клинических психологов "с различным опытом и образованием" (Goldberg, 1970, с. 425) попросили сделать этот прогноз по 11-этапному нормализованному распределению. Два других исследования рассматривают 90 аспирантов первого года обучения на психологическом отделении в университете штата Иллинойс, которые были выдвинуты на основании 10 переменных, способных предсказать академический успех. Эти переменные включали оценки теста на способности, средние оценки в колледже, отнесение к определенной категории (например, экстраверсия) и различные самооценки (например, добросовестность). Для всех аспирантов были посчитаны средние годовые оценки на первом году обучения. Задача состояла в том, чтобы спрогнозировать средние оценки, исходя из этих 10 переменных. Во втором исследовании этот прогноз был выполнен 80 (другими) аспирантами в университете штата Иллинойс (Wiggins & Kohen, 1971), а в третьем исследовании этот прогноз был выполнен 41 аспирантом в университете штата Орегон. Детали четвертого исследования уже были обсуждены, оно рассматривает прогноз поздних оценок способностей в Орегоне. В последнем исследовании (Yntema & Torgerson, 1961) экспериментаторы приписывали значения овалам, предложенным испытуемым, на основе размера фигур, эксцентricности и насыщенности серого цвета. Формула, которая использовалась, была $ij + kj + ik$, где i , j и k обозначают значения трех только что упомянутых измерений. Испытуемых в этом эксперименте попросили оценить значения каждого овала, и им предоставляли обратную связь исхода в конце каждого испытания. Задача состояла в том, чтобы предсказать истинное (т.е. приписанное экспериментатором) значение каждого овала на основе его размера, эксцентricности и насыщенности по цвету.

Таблица 1. Корреляции между значениями прогнозов и критерия

<i>Пример</i>	<i>Средняя валидность эксперта</i>	<i>Средняя валидность модели эксперта</i>	<i>Средняя валидность случайной модели</i>	<i>Валидность равнозве- шенкой мо- дели</i>	<i>Кросс- валидность регрессион- ного анализа</i>	<i>Валидность оптималь- ной и линей- ной модели</i>
Предсказания невротиков в сравнении с психотиками	0.28	0.31	0.30	0.34	0.46	0.46
Предсказания студентами Иллинойса средних годовых оценок	0.33	0.50	0.51	0.60	0.57	0.69
Предсказания студентами Орегона средних годовых оценок	0.37	0.43	0.51	0.60	0.57	0.69
Предсказание факультетского рейтинга в Орегоне	0.19	0.25	0.39	0.48	0.38	0.54
Эксперимент Ингема и Торгерсона (Yntema & Torgerson, 1961)	0.84	0.89	0.84	0.97	—	0.97

Первая колонка Таблице 1 представляет среднюю валидность экспертов в этих исследованиях, а вторая представляет среднюю валидность параморфической модели этих экспертов. Во всех случаях использовалось предыдущее значение для расчета последующего. А затем идут *случайные линейные модели*, созданные Корриганом и мною, модели, в которых значения выбирались наугад, за исключением знака, а затем применялись к стандартизированным переменным.⁵

Знак каждой переменной был определен на предшествующей основе так, чтобы он имел позитивное отношение к критерию. Затем из нормального распределения с единичной дисперсией было выбрано наугад нормальное отклонение, и абсолютное значение этого отклонения было использовано как вес переменной. Для каждого примера было построено 10 тысяч таких моделей. (Daws & Corrigan, 1974, с. 102)

В среднем, эти случайные линейные модели работают почти так же хорошо, как и параморфические модели экспертов. Эти средние значения предсказаны в третьей колонке таблицы. Равновзвешенные модели, представленные в четвертой колонке, функционируют даже лучше. (Существует чисто математическое обоснование того, почему равновзвешенные модели работают лучше.⁶) И, наконец, последние две колонки представляют кросс-валидизированную валидность стандартной регрессионной модели и валидность оптимальной линейной модели.

⁵ К сожалению, Daws и Corrigan не описали подробно, что эти переменные сначала должны быть стандартизированы, и что результат - зависимая стандартизированная переменная. Равное или случайное взвешивание несравнимых переменных - например, оценка ВЭ (Выпускного Экзамена) (GRE) и средняя оценка (GPA), - без предварительной стандартизации будет бессмысленным.

⁶ Рассмотрим набор стандартизированных переменных X_1, X_2, \dots, X_m , каждая из которых положительно коррелирует со стандартизированной переменной Y . Корреляция среднего значения $X - X_m$ с Y равна корреляции суммы значений X с Y . Ковариация этой суммы с Y равна:

$$\begin{aligned} & \left(\frac{1}{n}\right) \sum_i y_i (x_{i1} + x_{i2} + \dots + x_{im}) \\ &= \left(\frac{1}{n}\right) \sum_i y_i x_{i1} + \left(\frac{1}{n}\right) \sum_i y_i x_{i2} + \dots + \left(\frac{1}{n}\right) \sum_i y_i x_{im} \\ &= r_1 + r_2 + \dots + r_m \text{ (сумма корреляций)}. \end{aligned}$$

Дисперсия y равна 1, а дисперсия суммы значений X равно $M + M(M-1)\bar{r}$, где \bar{r} - это среднее значение внутренней корреляции между значениями X . Следовательно, корреляция средних значений X с Y равна $(\sum r_i) / (M + M(M-1)\bar{r})^{1/2}$; это больше, чем $(\sum r_i) / (M + M^2 \cdot M)^{1/2}$ = среднее r_i . Так как каждая из случайных моделей позитивно коррелирует с критерием. Корреляция среднего значения, которая является единично-взвешенной моделью, выше, чем среднее значение корреляций.

В сущности, те же самые результаты были получены в процессе отбора значения из прямоугольного распределения. Почему? Потому что линейные модели устойчивы к отклонениям из-за оптимального взвешивания. Другими словами, вывод использования предыдущего значения при расчете последующего, по крайней мере, в этих исследованиях, стал лишь подтверждением более раннего открытия, что правильные линейные модели превосходят суждения человека – веса, полученные от экспертов, настолько близки к оптимальным весам, что результаты моделей в высшей степени схожи. Одно из решений проблемы получения оптимального веса – в терминах фон Винтерфельдта и Эдвардса (von Winterfeldt и Edwards, 1973) – состоит в "категорическом максимуме". Веса, близкие к оптимальному уровню, приводят почти к такому же результату, что и оптимальное бета-значение. Так как эксперт знает, по крайней мере, хоть что-либо о направлении переменных, его суждения порождают веса, близкие к оптимальному (однако, важно, что во всех случаях равное значение превосходит модели, основанные на поведении оценивающего).

На тот факт, что различные линейные формулы в высокой степени коррелируют друг с другом, было впервые указано 40 лет назад Вилксом (Wilks, 1938). Он рассматривал только те ситуации, в которых наблюдалась позитивная корреляция между предсказываемыми переменными. Такой результат имеет силу обычно в тех случаях, когда эти внутренние корреляции не являются негативными, например, корреляция между $X+2Y$ и $2X+Y$ равна 0.80, когда X и Y – некоррелированные переменные. Относительная нечувствительность результатов к изменениям коэффициентов (при условии отсутствия изменений в знаке) совсем недавно исследовались Грином и Вайнером, Вайнером и Тиссеном, В. Эдвардсом, Гардинером и Эдвардсом (Green (1977), Wainer (1976), Wainer и Tissen (1976), W. Edwards (1978), Gardiner и Edwards (1975)).

Даус и Корриган (Dawes и Corrigan, 1974, с. 105) заключили, что "весь фокус в том, чтобы знать, какие переменные рассматривать, а затем уметь складывать". Этот принцип отлично иллюстрирует нижеследующее исследование, проведенное с момента опубликования статьи Дауса и Корригана. В нем Хеммонд и Адельман (Hammond и Adelman, 1976) как исследовали, так и влияли на решение о том, какого типа пули следует использовать в работе полицейским города Денвера, решение, имеющее гораздо более очевидное социальное значение, чем большинство обсуждаемых ранее. Цитата из Хеммонда и Адельмана (Hammond и Adelman, 1976):

В 1974 году отделение полиции города Денвера (ОПД), равно как и другие отделения полиции по всей стране, решили сменить амуницию. Основной причиной, предъявленной полицией, было то, что официально принятые пули с закругленным концом обеспечивали недостаточную "останавливающую эффективность" (то есть, возможность лишить способности двигаться и, таким образом, препятствовать раненому открыть ответный огонь по полицейскому и другим людям). По рекомендации начальника ОПД

(а также других начальников полиции) официальные пули были заменены пулями с полым наконечником. Такие пули, как утверждалось, выравнивают удар, таким образом, уменьшая степень проникновения, увеличивая останавливающую эффективность и уменьшая потенциал рикошета. На предполагаемые изменения отреагировали Американский Союз Гражданских Свобод, меньшинства и другие. Противники изменений заявляли, что новые пули были не чем иным, как запрещенными законом пулями "дум-дум", что они приводят к гораздо более серьезным телесным повреждениям, чем пули с круглым концом, и, следовательно, должны быть запрещены. Как обычно, мнения по этому поводу формировались в частном порядке, а затем публично отстаивались с должным энтузиазмом и упорством, а также проводились обычные публичные слушания. Обе стороны обращались к экспертам по баллистике за научной информацией и поддержкой (с. 392).

Дискутирующие стороны фокусировали внимание на оценке достоинств специфических пуль, смешивая их физическое воздействие и значение для социальной политики; то есть, стороны просто выступали в защиту той или иной пули, не отделяя вопрос о том, каким целям должна соответствовать пуля (вопрос социальной политики), от вопроса о ее баллистических характеристиках. Таким образом, как указывают Хеммонд и Адельман, адепты социальной политики, сами того не замечая, приняли роль (плохих) экспертов по баллистике, и наоборот. Заслуга Хеммонда и Адельмана в том, что они открыли важные политические измерения, основываясь на мнениях деятелей политики, а затем получили экспертную оценку пуль с учетом этих измерений. Оказалось, что этими измерениями были останавливающая эффективность (вероятность того, что человек, получивший ранение туловища, не сможет открыть ответный огонь), возможность серьезных ранений и возможность причинения вреда рядом стоящим. Когда эксперты по баллистике оценили пули с учетом этих измерений, то оказалось, что последние два были почти полностью перепутаны, но они не были перепутаны с первым. По одному из измерений, смешивающему эффективность со смертельностью, пули не различались. Возможность серьезного ранения или нанесения вреда рядом стоящим в высокой степени связана с проникающей способностью пули, тогда как возможность эффективно препятствовать открытию ответного огня в высокой степени связана с шириной входящего ранения. Так как политические деятели не могли договориться о значениях, приписанных этим трем измерениям, Хеммонд и Адельман предложили наделить их равным значением. Комбинируя равное значение с (независимыми) суждениями экспертов по баллистике, Хеммонд и Адельман обнаружили пулю, которая "имеет большую останавливающую эффективность и меньшую способность стать причиной повреждения (и угрожать рядом стоящим), чем стандартная пуля, используемая в то время ОПД" (Hammond и Adelman, 1976, с. 395). Эта пуля также могла причинить вред в меньшей степени, чем та, что была рекомендована ОПД ранее. Эта пуля была "принята Городским Советом и всеми другими заинтересованными партиями, и сейчас находится в использовании ОПД" (Hammond и Adelman, 1976,

с. 395).⁷ И снова, "весь фокус в том, чтобы решить, какие переменные рассматривать, а затем знать, как складывать" (Daws и Corrigan, 1974, с. 105).

Так почему же люди не делают этого чаще? Я знаю четыре университета (университеты штатов Иллинойс, Нью-Йорк, Орегон, Калифорния, Санта-Барбара – может быть и больше), которые используют линейную модель при отборе претендентов, но даже они используют ее, как механизм начального отсева и заменитель клинического суждения при окончательном отборе для тех, чьи оценки выше критических. Статистическая формула Голдберга (1965) для диагностики неврозов и психозов на основе данных ММРІ доказала свое превосходство над клиническими специалистами, пытающимися решить ту же задачу (никто, насколько известно мне или Голдбергу, не научил эксперта, который справлялся бы с этим лучше), и лишь однажды я прекратил использование формулы (в госпитале Управления ветеранов им. Энн Арбор (Ann Arbor Veteran Administration Hospital)), ввиду очевидных ошибок (интересная причина, обстоятельно обсуждавшаяся). В 1970 году я предложил, чтобы комитет нашего сообщества в Университете штата Орегон распределил сокращения сообществ Фонда национальной науки и Акта о защите национального образования по отделам на основе квазилинейной точечной системы, базирующейся на точно определенных индексах, достоинствах отделения и его нуждах; мне сказали: "Вы не можете систематизировать человеческие суждения". И только через шесть месяцев после того, как наш комитет осознал политическую и этическую невозможность сокращения сообществ на основе интуитивных суждений, такая система была принята. И так далее.

В течение последних трех лет я писал и говорил о пользе (и, по-моему, этическом превосходстве) использования линейных моделей при принятии социально важных решений. Периодически у различных отдельных читателей и аудиторий возникало множество подобных возражений. Я хотел бы закончить эту статью списком этих возражений и ответами на них.

Возражения против использования линейных моделей.

Эти возражения можно разделить на три обширных категории: технические, психологические и этические. Мы обсудим каждую категорию.

Технические

Наиболее распространенное техническое возражение направлено против

⁷ Следует отметить, что было лишь восемь пуль на *Границе Парето*; то есть, было только восемь пуль, не являющихся низшими по качеству, чем какая-либо другая, как в отношении останавливающей эффективности, так и в отношении возможности причинения вреда (или низшими по одной из переменных и равными по другой). Следовательно, любое правило распределения значения выбрало бы одну пулю из этих восьми.

использования коэффициента корреляции; например, Римас и Джиник (Remus и Jenicke, 1978, с. 221) писали:

"Ясно, что выбор Даусом и Корриганом коэффициента корреляции для установления полезности случайных и единичных правил является несоответствующим (несоответствующим чему?). Также необходима функция критерия в экспериментах, цитируемых Даусом и Корриганом. Несомненно, что существует и функция цены неправильной классификации невротиков и психозов или отказа в поступлении в аспирантуру образованным студентам, в то время как принимаются студенты с предельно низкими баллами."

Рассмотрим сначала проблему поступления в аспирантуру. Большинство учебных заведений имеют k мест и N претендентов. Проблема состоит в том, чтобы выбрать наилучших k (которых хотят принять в учебное заведение) из количества N . Какой же способ здесь будет лучшим, чем соответствующий разряд? Никакой. Римас и Джиник пишут так, как если бы это была проблема не сравнительного, а абсолютного выбора. Однако большинство социальных выборов включают в себя отбор лучших или наилучших альтернатив из набора имеющихся: студентов, которые будут успешнее; пуль, которые окажутся наилучшими, более выгодное место для аэропорта и т. д. Для оценки таких выборов коэффициент корреляции будет со всей очевидностью подходящим, так как он прекрасно отражает разряд.

Задача невротиков-психозов является более тонкой и поддерживает их доводы даже в меньшей степени. "Несомненно, — утверждают они, — здесь есть функция цены", но они не определяют подробно ни одного из кандидатов. Это и понятно: если они смогут это обнаружить, клинические суждения окажутся превосходящими линейные модели. Почему? При отсутствии подобного вывода с их стороны довод сводится на "нет". Но этот довод из вакуума может быть непреодолимым для людей (например, для проигрывающих генералов и проигрывающих футбольных тренеров, которые знают, что их планы несомненно сработают, в то время как планы в действительности обречены на провал, что бы ни произошло).

Другое связанное с этим техническое возражение касается сравнения средних коэффициентов корреляции экспертов и линейных моделей. Возможно, усреднение делает неясными суждения действительно выдающихся специалистов, но данные показывают иное. Например, в исследовании Голдберг (1970) только 5 из 29 обученных клинически психологов показали лучшие результаты, чем модель с равными весами, и ни один не превзошел правильную модель. В исследовании Виггинса и Коэна (1971) ни один из специалистов не превзошел модель с равными весами, и мы получили такой же результат в Орегоне. В исследовании Либби только 9 из 43 специалистов превзошли отношение активов и пассивов, прогнозируя банкротства (3 справились на том же уровне). Признавая то, что клиницисты должны уметь предсказать невротические и психотические диагнозы, что аспиранты должны быть способны предсказать успешность выпускных экзаменов, что банковские служащие отдела кредитов должны предвидеть банкротства,

то, возможно, эксперты, задействованные в исследованиях, были выбраны неправильно. Снова этот довод возникает из вакуума: если бы были задействованы другие эксперты, то и результаты были бы другими. Но нет таких экспертов, и опять-таки, соответствующим ответом будет вопрос, на каком основании эти гипотетически другие люди должны быть какими-то другими. (Как сказал мне один вице-президент университета, "Ваше исследование только доказывает то, что вы задействовали плохих специалистов; мы, несомненно, можем показать лучшие результаты, задействовав лучших специалистов" – очевидно, не с факультета психологии.)

И последнее техническое возражение касается природы критериальных переменных. Они, как признано, недолговечны и неглубоки (например, средние оценки, диагнозы); в противном случае большинство исследований было бы невозможным. Тогда поднимается вопрос, были бы другими результаты, если нужно было бы предсказывать точный важный длительный критерий. Ответ таков: результаты, конечно же, *могли бы* быть другими, но у нас нет оснований предполагать, что они *были бы* другими. Во-первых, отдаленное будущее обычно менее предсказуемо, чем непосредственное будущее, по той постоной причине, что на исходы деятельности индивида влияет большее количество непредвиденных внешних или внутренних факторов. (Заметим, что мы не обсуждаем совокупные исходы, такие, как необычайно холодная зима на Среднем Западе, продлившаяся три месяца.) Поскольку начинать с клинического предсказания хуже, чем с линейного, то гипотеза будет иметь силу только в том случае, если линейное предсказание со временем окажется намного хуже, чем клиническое. Нет никакого основания, полагать, что произойдет это дифференциальное ухудшение в предсказании, и ни об одном из них у меня никогда не возникало мысли. Определенно, не существует никаких свидетельств. Опять-таки, возражение состоит из довода, взятого из вакуума.

Особенно непреодолимым является факт, что люди, которые утверждают, что другие критерии, эксперты, переменные или временные рамки могут привести к другим результатам, имели в своем распоряжении 25 лет для того, чтобы представить примеры, но они не смогли этого сделать.

Психологические

Одно психологическое возражение против использования линейных моделей находится в нашей избирательной памяти о клиническом предсказании. Наша вера в такое предсказание подкрепляется доступностью (Tversky & Kahneman, 1974) примеров успешного клинического предсказания, – особенно тех, что являются исключением некоторой формулы: "Я однажды знал человека с..., который..." (например, "Я знал об одном человеке с IQ равным только 130, который получил высокую степень по психологии". Как показали Нисбетт, Боргида, Крендалл и Рид (Nisbett, Borgida, Crandall и Reed, 1976,7) такие единичные примеры часто имеют более сильное воздействие

на суждения, чем намного более достоверные статистические сведения, основанные на многочисленных примерах. Хорошей профилактикой для клинических психологов, основывающих возражение статистическому предсказанию на таких примерах, было бы сохранение точных протоколов своих собственных предсказаний о своих собственных пациентах, – будущие протоколы не попадают под прицел. Такие протоколы смогут сделать все примеры успешного и неуспешного предсказания в равной степени доступными для использования; в дополнение к этому, они могут быть полезны для другого клинического исследования, в противовес статистическому, в смысле использования как наилучшего эксперта – самого клинициста.

Более того, иллюзия хорошего суждения может быть усилена из-за отбора (Einhorn & Hogart, 1978) в тех ситуациях, когда предсказание положительного или отрицательного исхода производит эффект самоисполнения. Например, экзаменаторы, решившие, что кандидат хорошо подготовлен для выпускной программы, могут чувствовать, что их суждение оправдано, если этот кандидат успешен, даже хотя во многом успех кандидата является результатом положительного воздействия самой программы. (В противовес этому, линейная модель отбора оценивается по тому, насколько хорошо она прогнозирует деятельность выбранных претендентов *в пределах* этой группы.) Или официант, который полагает, что некоторые люди за столиком не платят хороших чаевых, может быть менее внимателен, чем обычно, и получить меньшие чаевые, тем самым, оправдав свое клиническое суждение.⁸

Второе психологическое возражение против использования линейных моделей основывается на их низкой "доказательной" валидности. Здесь существует имплицитный (как противоположный эксплицитному) аргумент из вакуума, так как не предполагаются ни изменения в оценочных процедурах, ни в экспертах, ни в критериях. Вернее, невысказанное предположение состоит в том, что эти критерии психологической выгоды, в сущности, в высокой степени прогнозируемы. Таким образом, следует, что, если один метод прогнозирования (линейная модель) не работает достаточно хорошо, другой может работать лучше (основательный вывод), что далее трансформируется в убеждение, что тот, другой, способ *будет* работать лучше (что не является основательным выводом), – если однажды он был обнаружен. Это возражение наилучшим образом выразил декан, рассматривавший принятие старшекурсников в аспирантуру, который написал: "Корреляция линейной комбинации с будущим факультетским рейтингом составляет только 0.4, тогда как корреляция суждений приемной комиссии – 0.2. Дважды ничего равняется ничему". В 1976 году я ответил следующее (Dawes 1976, с.6-7):

В ответ я могу только отметить, что 16% изменений лучше, чем 4%. Для меня, однако, очаровательной частью этого довода является подразумеваемое предположение, что другие 84% изменений прогнозируемы, и мы можем их каким-то образом предсказать.

⁸ Этот пример был приведен Einhorn (1979).

Так с чем же мы имеем дело? Мы имеем дело с личностью и интеллектуальными характеристиками (одинаково сообразительных) людей, которым около 20 лет.... Почему мы так убеждены, что этот прогноз может быть сделан вообще? Несомненно, не нужно читать *Экклезиаста* каждый вечер, чтобы понимать роль шанса.... Более того, есть очевидно положительные эффекты обратной связи в профессиональном развитии, которые преувеличивают пороговый феномен. Например, если однажды человека посчитали достаточно "выдающимся", то его уже приглашают в выдающиеся учреждения, он общается выдающимися коллегами, – и превосходство усиливается. С подобной проблемой сталкиваются и те, кто недостаточно богат для порогового уровня. Все эти факторы не только ослабляют успешное длительное прогнозирование, но и исследования успешности такого прогнозирования неизбежно ограничиваются до определенного круга принятых исследований, с возложенными проблемами ограничения предела и с негативной ковариантной структурой среди прогнозирующих факторов (Dawes, 1975).

Наконец, на профессиональный успех могут влиять разнообразные неинтеллектуальные факторы, которые не могут быть оценены перед принятием в аспирантуру. Например, успешность формирования удовлетворительных или вдохновляющих половых отношений, еще не очевидная генетическая предрасположенность к наркотической или алкогольной зависимости, несчастье присоединиться к исследовательской группе, которая лопнет и т. д., и т. п.

С интеллектуальной позиции, я нахожу в некоторой степени примечательным то, что мы в силах прогнозировать даже 16% изменений. Но я уверен, что мой собственный эмоциональный ответ является показательным на фоне ответов моих коллег, которые лишь просто предполагают, что будущее более прогнозируемо. *Я хочу, чтобы оно было прогнозируемо, особенно в тех случаях, когда аспект того, что я хочу прогнозировать, важен для меня.* Это желание, я полагаю, трансформирует само себя в подразумеваемое предположение о том, что будущее, в действительности, прогнозируемо в высшей степени, и за этим логически следует, что, если нечто не является хорошим прогностическим элементом, что-нибудь другое может оказаться в этом плане лучшим (хотя неправильно будет утверждать, что оно обязательно окажется).

Статистическое предсказание, так как оно включает спецификацию (обычно низкий коэффициент корреляции) того, как плохо мы можем предсказывать, резко поражает нас фактом, что жизнь не является полностью предсказуемой. Несистематическое клиническое предсказание (или "постдикция"), наоборот, предлагает нам утешительную иллюзию, что жизнь в действительности предсказуема, и мы ее можем предсказывать.

Этические

Когда я был прошлым летом на Ярмарке Ренессанса в Лос-Анджелесе, я случайно услышал жалобы одной женщины, что она была "ужасно несправедливо" отвергнута на психологическом факультете в Университете Калифорнии, Санта-Барбара, на основании одних лишь цифр, даже без интер-

вью. "Как они могут сказать, какая я?" Ответ состоит в том, что они не могут. Равно как и с интервью (Kelly, 1954). Тем не менее, многие люди придерживаются мнения, что делать решающий социальный выбор без интервью негуманно. Я думаю, что вопрос о том, обращаются ли с людьми справедливо, более соответствует вопросу о том, были или нет с ними гуманны, чем к вопросу о том, встречались ли с ними лично. (Некоторые из наилучших докторов проводят большое количество времени в разговорах со своими пациентами, не читают медицинских журналов, делают мало или не делают вовсе анализов и горюют на похоронах.) Средние оценки представляет поведение претендента в течение 3,5 лет. (Конечно же, не все профессора имеют предубеждение относительно особой формы креативности претендента.) Балл выпускных экзаменов является более тщательно продуманной проверкой. Неужели мы действительно считаем, что мы справимся с этой работой лучше или справедливее 10-минутной оценкой документов из личного дела или получасовым интервью, чем могут сделать эти две простые цифры? Такое когнитивное самомнение (Dawes, 1976, с.7) неэтично, особенно учитывая факт отсутствия очевидного доказательства того, что мы это сделаем лучше, чем линейное уравнение. (Но даже исключения необходимо делать с предельной осторожностью, если это должно быть этично; так как если мы примем кого-нибудь с предельно низкими линейными показателями на основании того, что у него или нее есть особый талант, мы автоматически откажем кому-то с более высокими оценками, кто с большой вероятностью может иметь в равной степени впечатляющий талант, если бы мы приняли его оценивать.)

Не имеет значения, насколько сильно измененным мы хотим видеть тот или иной аспект в том или ином исследовании, рассмотренном в этой статье. Не имеет значения, насколько психологически независимыми или неприятными мы находим их результаты. Не имеет значения, насколько этически неудобно мы себя чувствуем, "приравнивая людей к простым цифрам". Фактом остается то, что наши клиенты – это люди, заслуживающие возможного наилучшего отношения. Если это означает, – как это происходит в настоящее время, – что отбор, диагностика и прогнозирование должны основываться не более, чем на сумме чисел, представляющих оценки по важным характеристикам, то так оно и должно быть. Поступать иначе означает обманывать людей, которым мы служим.

29. Жизнеспособность мифических чисел*

Макс Зингер

Вообще предполагается, что за год в Нью-Йорке наркоманами, зависимыми от героина, разворовывается имущества на сумму от 2 до 5 млрд. долларов и совершается половина имущественных преступлений. Эти оценки преступности среди наркоманов используются такой организацией как RAND, таким политическим деятелем как Говард Самюэлс, и даже Генеральным Прокурором Соединенных Штатов. Оценка того, что половина имущественных преступлений совершается наркоманами, первоначально приписывалась полицейским чиновникам, но использовалась так часто, что стала частью общественного сознания.

Стоимость имущества, украденного наркоманами, обычно оценивается следующим образом:

Насчитывается 100.000 наркоманов, чей средний дневной расход на наркотики составляет \$30. Это означает, что наркоманам необходимо где-то \$1,1 млрд. в год на покупку героина ($100.000 \times 365 \times \30). Так как наркоман должен продать краденую собственность только за четверть ее реальной стоимости, или даже меньше, он должен украсть где-то от \$4 до \$5 млрд. в год на покупку героина.

Эти вычисления могут быть выполнены с большей или меньшей точностью. Можно допустить, что те наркоманы, которые ведут незаконный образ жизни, проводят больше четверти своего времени в тюрьме, что уменьшило бы сумму преступлений на четверть. (*The New York Times* недавно сообщал о смерти Вильяма Рейли (по прозвищу "Осел"). 74-летний бывший наркоман, бывший в зависимости в течение 54 лет, 30 из которых провел в тюрьме.) Определенную часть того, что ворует наркоман, составляют наличные деньги, которые не являются материальным имуществом. Значительную

* Эта глава первоначально появилась в *The Public Interest*, 1971, 23, 3-9. Авторское право © 1971 у National Affairs, Inc. Печатается по разрешению.

часть стоимости героина покрывает, скорее, участие в наркоторговле, чем обкрадывание общества, а другая большая часть покрывается за счет проституции, включая мужчин-наркоманов, живущих за счет проституток. Но независимо от того, насколько тщательно это рассматривать, если пытаться дать оценку собственности, украденной наркоманами, предположив, что их 100.000, и оценив минимальную сумму, которую им необходимо будет украсть, чтобы поддержать себя и свои привычки (после оценивания законного дохода), то приходишь к цифре, соседствующей с \$1 млрд. в год для Нью-Йорка.

Но что произойдет, если подойти к этому вопросу с другой стороны? Предположим, мы спрашиваем: “Какое количество собственности украдено – наркоманами или кем-либо еще?” Краж наркоманов должно быть меньше, чем суммарное количество краж. Какова ценность собственности, украденной в Нью-Йорке в год? Когда я спросил в первый раз, мне показалось несколько удивительным, что на этот вопрос, как оказалось, достаточно трудно ответить, даже приблизительно. Ни у кого нет оценок, в которых они имели бы хоть наименьшую уверенность, и вопрос, казалось, задавался нечасто. Сумма официально заявленных краж в Нью-Йорке составляет приблизительно 300 млн. в год, из которых около 100 млн. приходится на автомобильные кражи (преступления, редко совершаемые наркоманами). Но ясно, что о большом количестве краж не заявлено; например, обычно не заявляют в полицию о магазинных кражах. (Об огромных потерях собственности не сообщается также и в страховые компании, и страховая индустрия не имеет истинной оценки суммарных краж.)

Однако оказывается, что если кто-то просто задает вопрос типа: “Возможно ли, что наркоманы украли собственности на \$1 млрд. в прошлом году в Нью-Йорке?”, то относительно легко оценить, на какую сумму украдено собственности. Ясно, что два больших компонента краж наркоманов – это магазинные кражи и кражи со взломом. Какова *может быть* ценность собственности, украденной наркоманами в магазинах? Вся розничная торговля в Нью-Йорке составляет порядка \$15 млрд. в год. Сюда входят автомобили, ковры, бриллиантовые кольца и другие изделия, недоступные для магазинных воров. Разумная потеря имущества в учреждениях розничной торговли составляет 2%. Сюда входят растраты руководства, кражи клерков, в транспортно-погрузочном отделе, водителей и т.д. (Универмаги, в частности, сообщают о большом увеличении магазинных краж в последние годы, но они находятся в числе наиболее уязвимых учреждений розничной продажи и не так существенны, чтобы намного увеличить общую норму в 2%.) Общепринято, что существенно больше, чем половина собственности, исчезнувшей из учреждений розничной торговли, выносятся сотрудниками, а остаток – магазинными ворами со стороны. Но давайте припишем им 1% всей украденной собственности, продававшейся в розницу в Нью-Йорке, – это будет \$150 млн. в год.

А как же насчет кражи со взломом? В Нью-Йорке где-то 2,5 млн. домашних хозяйств. Предположим, в среднем 1 из 5 домов подвергается грабежу или краже со взломом каждый год. Здесь учитывается, что в некоторых районах кражи со взломом являются обычным явлением, и что некоторые дома подвергаются краже со взломом даже более 1 раза в год. Средняя ценность собственности, выносимой при краже со взломом, может быть порядка \$200. При некоторых кражах со взломом, конечно же, выносятся гораздо больше собственности, но такие ценные кражи совершаются намного реже и часто профессиональными ворами, не являющимися наркоманами. Если мы используем цифры $\$200 \times 500.000$ краж, мы получим, что за год в Нью-Йорке из домов людей совершаются кражи имущества на сумму \$100 млн.

Очевидно, ни одно из этих оцененных значений не является ни несомненным, ни достаточно обоснованным. Вы можете сделать собственную оценку. Оценки здесь носят такой характер, что было бы очень удивительно, если бы они ошибались в 10 раз, и не очень важно для заключения, если бы они ошибались в 2 раза. (Это хорошая позиция, в которой может находиться оценивающий.)

Очевидно, что не все кражи наркоманов – это собственность, взятая из магазинов и домов. Наиболее ужасающие типы преступлений, совершаемые наркоманами – это кражи собственности при разбойном, хулиганском нападении на жителей Нью-Йорка. И мы можем их оценить. Предположим, что в среднем каждый год один человек из десяти подвергается разбойному или хулиганскому нападению с кражей собственности. Это будет 800.000 таких нападений, и если в среднем человек отдает \$100 (что маловероятно), в такой форме краж ежегодно изымается \$8 млн.

Таким образом, мы видим, что, если приписать наркоманам *все* магазинные кражи, *все* кражи в домах, *все* кражи у отдельных людей, общая стоимость собственности, украденной наркоманами в Нью-Йорке, составляет где-то \$330 млн. за год. Можно перебросить сюда все “поправочные множители”, какие пожелаешь, прибавить все другие разнообразные преступления, совершаемые наркоманами, но что бы ни было сделано, трудно найти основания для той оценки, что наркоманы воруют более половины миллиарда в год, а четверть миллиарда выглядит более правдоподобной оценкой, хотя и, возможно, по максимуму. В конце концов, есть же воры, не являющиеся наркоманами.

Таким образом, я уверен, мы показали, что, несмотря на общепринятое предположение о том, что наркоманы в Нью-Йорке воруют от \$2 до \$5 млрд. в год, реальная цифра в десять раз меньше, и что это можно продемонстрировать пятью минутами размышлений.¹ Так что же? Даже на сумму в четверть миллиарда – это все же много собственности. Эта цифра превышает суммы,

¹ Мифические числа могут быть более мифическими и иметь большую жизнеспособность в области преступлений, чем в большинстве других. В начале 50х годов Kefauver Committee опубликовало оценку годового “изъятия” в азартных играх в США в \$20 млрд. Цифра в “действительности была взята с неба”. Один из сотрудников сказал: “Мы не имели ни малейшего представления о потраченных деньгах. Калифорнийская Комиссия по Преступле-

трапящиеся ежегодно на реабилитацию наркоманов и на другие программы, направленные на предотвращение и контроль наркомании. Кроме того, ценность собственности, украденной наркоманами, это малая часть той цены, которую платит общество за воровство наркоманов. Намного большая цена — это страх, измененная атмосфера соседства, вынужденная предосторожность и другие побочные реакции на воровство и его опасность.

Один из моментов в попытке оценить стоимость собственности, украденной наркоманами, состоит в том, чтобы пролить свет на отношение людей к цифрам. Люди чувствуют, что наркоманами совершается множество преступлений, и что \$2 млрд. — это большая цифра, таким образом, они склонны верить, что кражи наркоманов составляют \$2 млрд. в год. Но \$250 млн. — это тоже большая цифра, и если бы наше ощущение перспективы не было бы искажено ежедневным осознанием федеральных расходов, большинство вполне довольствовалось бы принятием того факта, что \$250 млн. в год — это большая сумма краж.

В продолжение нити рассуждений, эта попытка оценки является еще одним напоминанием, что даже ответственные должностные лица, ответственные газеты и ответственные исследовательские группы подбирают и передают как сплетни цифры, которые в действительности не имеют реального основания. Этот опыт напоминает нам, что только потому, что оценка широко использовалась различными людьми, которые должны знать, о чем они говорят, нельзя предполагать, что она является даже приблизительно верной.

Тот факт, что наркоманами не может совершаться так много краж, как в том уверены люди, подразумевает еще более важный момент, а именно: не может быть и такого количества наркоманов, как многие полагают. Большая часть денег в уплату за героин в розницу идет за счет воровства, и большинство наркоманов покупают героин в розницу. Таким образом, количество наркоманов, в основном, — хотя и не точно, — ограничено суммой краж. (По оценке, полученной в исследовании Гудзоновского института (Hudson Institute), около половины объема героина потребляется людьми, участвующими в системе его распространения, которые не покупают его в розницу и не платят за него за счет украденной собственности, а только за счет своих “услуг” в этой системе распространения.²) Но, несмотря на то, что люди, участвующие в этом бизнесе (на более низких уровнях), потребляют около половины героина, они составляют всего лишь 1/6 или 1/7 от всего количества наркоманов. Они могут позволить себе дорогие привычки.

нием сказала — \$12 млрд., Вирджил Петерсон из Чикаго сказал — \$30 млрд. Мы выбрали \$20 млрд. как баланс между этими двумя.

Необычный пример мифических чисел, имевший огромную силу, — утверждение, что полицией было застрелено 28 Черных Пантер, — подробно описывался Эдвардом Джейм Эпштейном в газете *The New Yorker* от 13 февраля, 1971 года. (Оказалось, что было убито 19 Пантер, 10 — полицией, из которых 8 — в ситуациях, когда Пантеры, вероятно, были активны).

² Параллельные данные были получены в более позднем исследовании в Больнице Св. Луки 81 наркомана — средний возраст 34 года. Больше половины героина, потребляемого этими

Наиболее популярная, неофициальная оценка количества наркоманов в Нью-Йорке – 100.000 с плюсом (обычно с акцентом на “плюс”). Федеральный реестр в Вашингтоне насчитывает 30.000 наркоманов в Нью-Йорке, а реестр имен наркоманов Нью-Йоркского департамента здоровья – 70.000. Несмотря на то, что не все люди в этих списках все еще являются активными наркоманами, – многие из них умерли или в тюрьме, – большинство полагают, что многие наркоманы вообще не попали ни в один из этих списков. Обычно оценку, что в Нью-Йорке 100.000 наркоманов, считают очень консервативной. В начале 1970 года широко цитировали оценку доктора Джудиан Денсен-Джербер (Dr. Judianne Densen-Gerber), что к концу лета будет более 100.000 подростков-наркоманов. Очевидно, однако, что есть еще и наркоманы старше 20 лет.³

Обсуждая в этой статье вопрос количества наркоманов, мы будем говорить о таком человеке, о котором думаешь, когда употребляешь термин “наркоман”.⁴ Лучшим термином, возможно, будет “уличный наркоман”. Это человек, обычно употребляющий героин каждый день. Это такой человек, который выглядит и действует согласно обычному образу наркомана. Мы исключаем здесь представителей медицинской профессии, которые часто используют героин и другие наркотики или пристрастились к ним, студентов, употребляющих героин от случая к случаю, состоятельных людей, хотя и зависимых, но которым не нужно воровать и часто посещать обычные притоны наркоманов, и т.д. Когда мы обращаемся к “проблеме наркомании”, гораздо менее важно включать эти случаи; несмотря на то, что они все же являются проблемами в той или иной степени, однако проблемами совершенно другого типа, чем та, что поставлена типичным уличным наркоманом.

наркоманами за год, было оплачено его продажей. Между прочим, в среднем каждый из этих наркоманов украл за предыдущий год собственности на \$9.000.

³ Среди последних исследований, пытающихся дать оценку, мы можем отметить марксиста Sol Yurick, давшего нам “500.000 отбросов” (Monthly Review, декабрь, 1970), и William R. Corson, который утверждал в Penthouse за декабрь 1970 года, что “сегодня по меньшей мере 2. 500.000 черных американцев посажено на героин”.

⁴ Существует интересная аномалия, касающаяся слова “наркоман”. Большинство людей, если бы их попросили дать определение слову “наркоман”, сказали бы, что это человек, который регулярно принимает героин (или подобный наркотик), и у которого, если не удается достать обычную дозу героина, будут наблюдаться неприятные и болезненные ответные симптомы. Но это определение не будет применимо к большей части тех, кого обычно признают как “совокупность наркоманов”. В действительности, оно не будет применимо к большинству настоящих наркоманов. Наркоман, которого очистили от токсинов, или который находился в тюрьме и был огражден от наркотиков в течение недели или около того, не соответствует обычному определению “наркомана”. У него уже нет физических симптомов в результате воздержания от героина. “Осел” Рейли (“Donkey” Reilly), конечно же, отвечает понятиям большинства людей о наркоманах, но в течение 30 лет из 54, когда он был “наркоманом”, он находился в тюрьме и, конечно же, не был активно зависим от героина большую часть времени, проведенного там, что составляет больше половины его карьеры “наркомана” (хотя определенное количество наркотиков доступно и в тюрьме).

По количеству собственности, украденной наркоманами, можно полагать, что в Нью-Йорке их скорее 70.000, чем 100.000, и почти наверняка не может быть 200.000, как иногда утверждают. Некоторые другие простые способы оценки количества уличных наркоманов приводят к подобным результатам.

Исследователи, имеющие дело с совокупностью наркоманов, пришли к оценке, что средний уличный наркоман проводит от 1/4 до 1/3 своей жизни в тюрьме. (Некоторые из тех, кто изучает эту тему, такие как Edward Preble и John J. Casey, Jr., полагают, что в среднем более 40%.) Это подразумевает, что в любое время от 1/4 до 1/3 всех наркоманов находятся в тюрьме, и что общее число наркоманов может быть оценено при умножении количества наркоманов в тюрьмах на 3 или 4. Конечно же, количество наркоманов в тюрьме не является известной величиной (и, в действительности, как мы указывали выше, даже не совсем точно отвечает понятию). Однако можно все же сделать приемлемую оценку количества наркоманов в тюрьме (и для этой цели мы можем задействовать наркоманов в различных учреждениях принудительного лечения). Это количество приблизительно равно 14.000 – 17.000, что вполне совместимо с оценкой количества уличных наркоманов в Нью-Йорке, составляющей 70.000.

Другим способом оценки общего количества уличных наркоманов в Нью-Йорке является использование имеющейся демографической информации о совокупности наркоманов. Например, мы можем небезосновательно быть уверенными в том, что где-то 25% уличных наркоманов в Нью-Йорке – это пуэрториканцы, а где-то 50% – афро-американцы. Мы знаем, что приблизительно 5 из 6 уличных наркоманов – это мужчины, и что 50% попадают в возрастную категорию от 16 до 25 лет. Это будет означать, что 20% общего количества уличных наркоманов – это чернокожие мужчины в возрасте от 16 до 25 лет. Если наркоманов 70.000, то это будет означать, что 14.000 афро-американцев в возрасте от 16 до 25 лет являются наркоманами. Но всего в городе около 140.000 афро-американцев в возрасте от 16 до 25 лет – возможно, половина из них живет в бедных районах. Это означает, что если в городе 70.000 наркоманов, то 1 из 10 молодых афро-американцев – наркоман, если 100.000 наркоманов – то 1 из 6, если 200.000 – то 1 из 3. Вы можете решить сами, какая из степеней охвата группы молодых мужчин афро-американцев более вероятна, но понятно, что 200.000 наркоманов – неправдоподобное количество. Подобным образом, количество уличных наркоманов 70.000 будет подразумевать, что 7.000 молодых пуэрториканцев-мужчин – наркоманы, а общее количество пуэрториканских парней в возрасте от 16 до 25 лет в Нью-Йорке – около 70.000.

Ни одно из вышеперечисленных вычислений ни в коем случае не пытается свести на “нет” значение проблемы героиновой зависимости. Героин – это ужасное проклятие. Когда думаешь об индивидуальной трагедии, то 70.000 – это ужасающе большое количество наркоманов. И если тебе необходимо зарабатывать на жизнь, то \$250 млн – это ужасающе огромные деньги, которые воруют у граждан города с тем, чтобы, пройдя через руки нар-

команов и торговлю краденым, они осели в карманах тех, кто ввозит и распространяет героин, и тех, кто берет взятки и оказывает другие услуги для героиновой индустрии.

Главным моментом этой статьи может быть иллюстрация того, как далеко можно зайти в рассмотрении проблемы, серьезно относясь к цифрам, к тому, что они подразумевают, проверяя различные выводы друг относительно друга и относительно общей информации (такой как количество людей или домов в городе). Незначительные усилия в этом направлении могут иметь большое значение для помощи обычным людям и должностным лицам в том, чтобы справиться с различного рода экспертами.

30. Интуитивное прогнозирование: предубеждения и корректирующие процедуры*

Даниель Канеман и Амос Тверски

Предисловие

Любая значительная деятельность по прогнозированию включает весомый компонент суждений, интуиции и развитой догадки. В самом деле, мнения экспертов являются источником многих технологических, политических и социальных прогнозов. Мнения и интуиция играют важную роль даже там, где прогнозы получают в результате использования математических моделей или имитации. Интуитивные суждения используются при выборе переменных, рассматриваемых в таких моделях, приписанных им факторов влияния, а также их предполагаемых начальных значений. Критическая роль интуиции в прогнозировании любого рода требует анализа факторов, которые ограничивают точность экспертных суждений, а также разработки процедур, направленных на улучшение качества этих суждений...

Единичные случаи и данные распределения

От экспертов часто требуют предоставления наилучших догадок, оценок или прогнозов, касающихся неопределенных величин, таких как индекс Доу-Джонса в определенный день, будущие продажи продукта или результаты выборов. Следует различать два типа информации, имеющейся в распоряжении прогнозиста: данные случая и данные распределения. Единичная информация или данные случая, состоит из очевидных свидетельств по частному рассматриваемому случаю. Данные распределения, или данные базового значения, состоят из знаний о распределении исходов в подобных си-

* Эта глава является сокращенным вариантом статьи, которая появилась в редакции S. Makridakis и S.C. Wheelwright, "Forecasting", TIMS, Studies in Management Science, 1979, 12, 313-327. Авторские права © 1979 у North-Holland Publishing Co. Печатается в соответствии с разрешением.

туациях. Например, при прогнозировании продаж нового романа то, что известно об авторе, стиле и сюжете, является единичной информацией, тогда как то, что известно о продажах романов, является данными распределения. Таким же образом, при прогнозировании долголетия пациента единичная информация включает его возраст, состояние здоровья и прошлые истории болезни, тогда как данные распределения состоят в соответствующей статистике населения. Единичная информация описывает характерные черты проблемы, которые отличают ее от других, тогда как информация распределения характеризует исходы, наблюдавшиеся в случаях подобного рода. Современное понятие информации распределения не совпадает с понятием распределения предшествующей вероятности Байеса. Первая определяется природой данных, тогда как последняя определена в терминах последовательности приобретения информации.

Многие проблемы прогнозирования уникальны в том смысле, что в наличии имеется очень мало, если есть вообще, релевантной информации. Примерами могут быть прогноз спроса ядерной энергии в 2000 году, или прогноз того, когда будут изобретены эффективные лекарства от лейкемии. В таких проблемах эксперт должен полагаться исключительно на единичную информацию. Однако действительность показывает, что люди недостаточно чувствительны к данным распределения, даже если она имеется в наличии. В самом деле, недавнее исследование показало, что люди полагаются преимущественно на единичную информацию, даже если она скудна и ненадежна, и приписывают недостаточный вес данным распределения (Kahneman & Tversky, 1973, 4; Tversky & Kahneman, 10).

В контексте планирования есть очень много примеров, в которых игнорировалось распределение исходов предыдущего опыта. Ученые и писатели, как известно, склонны недооценивать время, необходимое для завершения проекта, даже если у них есть значительный опыт прошлых неудачных попыток работать согласно планируемому графику. Подобная склонность была задокументирована в оценке инженерами времени, необходимого для завершения ремонта электростанции (Kidd, 1970). Хотя эта ошибка планирования относится к мотивационным факторам, таким как принятие желаемого за действительное, очень часто она допускается даже тогда, когда недооценка длительности или стоимости является наказуемой.

Ошибка планирования является следствием тенденции пренебрегать распределительной информацией и принимать то, что можно определить как внутренний подход к прогнозированию, при котором сосредотачиваются скорее на составляющих специфической проблемы, чем на распределении исходов в подобных случаях. Внутренний подход к оценке планов с большой вероятностью приведет к недооценке. Строительство здания может быть завершено в сроки, например, если не будет задержек в доставке материалов, забастовок, необычных погодных условий и т.п. Хотя все эти помехи маловероятны, возможность того, что хотя бы одна из них может произойти, существенна. Такое комбинаторное соображение, однако, неадекватно

представлено в интуиции людей (Bar-Hillel, 1973). Попытки побороть эту ошибку путем прибавления фактора погрешности редко являются адекватными, поскольку отрегулированное значение слишком близко к начальному значению, так что действует, как закрепитель (anchor) (Tversky & Kahneman, 1974, 1). Принятие внешнего подхода, который рассматривает специфическую проблему как одну из многих, может помочь преодолеть эту склонность. При таком подходе не надо пытаться предсказать особый образ действий, при котором план может потерпеть неудачу. Скорее, необходимо соотнести насущную проблему с распределением сроков выполнения для подобных проектов. Полагают, что вероятнее получить соответствующую оценку, задавая внешний вопрос: Как долго обычно длятся такие проекты? А не просто внутренний вопрос: Каковы специфические факторы и трудности в рамках специфической проблемы?

Тенденция пренебрегать данными распределения и полагаться главным образом на единичную информацию усиливается любым фактором, увеличивающим ощущаемую уникальность проблемы. Релевантность данных распределения может быть скрыта подробным ознакомлением со специфическим случаем или сильной вовлеченностью в него. На кажущуюся уникальность проблемы также влияет и формулировка вопроса, ответ на который должен дать эксперт. Например, вопрос о том, сколько будет стоить разработка нового продукта, можно решить, используя внутренний подход, при котором общая стоимость будет разбита на компоненты. Эквивалентный вопрос о процентном превышении стоимости над текущим бюджетом, вероятно, заставит принять во внимание распределение превышения стоимости для разработок подобного рода. Таким образом, смена единицы измерения, – например, стоимости на превышение, – может изменить способ рассмотрения проблемы.

Превалирующая тенденция недооценивать или игнорировать данные распределения, возможно, является основной ошибкой интуитивного прогнозирования. Принятие во внимание данных распределения, конечно же, не гарантирует точности прогнозов. Однако обеспечивает некоторую защиту от совершенно нереалистичного прогнозирования. Аналитика, таким образом, следует предпринимать все попытки, чтобы сформулировать прогностическую проблему с целью облегчения использования данных распределения, доступной эксперту.

Регрессия и интуитивное прогнозирование

В большинстве прогностических проблем эксперт имеет как единичную информацию о специфическом случае, так и данные распределения об исходах в подобных случаях. Примерами могут быть консультант, прогнозирующий вероятные достижения студента, банкир, оценивающий потенциальную прибыль малого бизнеса, издатель, оценивающий продажу учебника, или экономист, прогнозирующий некоторый индекс экономического роста.

Как люди прогнозируют в таких ситуациях? Психологическое исследование полагает, что интуитивное прогнозирование генерируется согласно простому правилу сопоставления (*matching*): прогнозируемое значение выбирается таким образом, чтобы положение случая в распределении исходов соответствовало его положению в распределении впечатлений (Kahneman & Tversky, 1973, 4; Ross, 1977). Следующий пример иллюстрирует это правило. Редактор просмотрел рукопись романа и был приятно удивлен. Он сказал: “Эта книга читается как бестселлер. Среди книг такого типа, издаваемых за последние годы, я бы сказал, только 1 из 20 оказала на меня большее впечатление”. Если бы редактора попросили оценить продажу романа, он бы, возможно, предсказал, что она будет на уровне первых пяти процентов распределения продаж.

Имеются значительные свидетельства того, что люди часто прогнозируют путем сопоставления прогноза и впечатления. Однако это правило прогнозирования необоснованно, так как оно не может принимать во внимание неопределенность. Редактор из нашего примера, несомненно, отметил бы, что продажа книг в высшей степени непредсказуема. В подобной ситуации с высокой неопределенностью наилучший прогноз продажи книг должен бы находиться где-то между значением, соответствующим чьему-то впечатлению, и средним значением продаж книг подобного типа.

Один из основных принципов статистического прогнозирования, который также является одним из наименее интуитивных, состоит в том, что предельность прогнозирования необходимо сдерживать рассмотрением предсказуемости. Представьте, например, что издатель знает из прошлого опыта, что продажа книг совсем не связана с его начальными впечатлениями. Рукописи, которые произвели на него благоприятное впечатление, и те, которые ему не понравились, в равной степени могли бы продаваться как хорошо, так и плохо. В таком случае нулевой прогностичности наилучшая догадка издателя – это одинаковая продажа всех книг, например, среднее значение соответствующей категории, независимо от его личного впечатления об одной из книг. Позволительно сопоставлять прогнозирование и впечатление только в случаях совершенной прогнозируемости. В промежуточных ситуациях, которые, конечно же, наиболее часты, прогнозирование должно быть регрессивным; то есть, оно должно быть между средним значением по классу и значением, наилучшим образом представляющим чье-либо впечатление в конкретном случае. Чем ниже прогнозируемость, тем ближе должен быть прогноз к среднему значению по классу. Интуитивное прогнозирование типично является не регрессивным: люди часто делают предельный прогноз на основе информации, чья надежность и прогностическая валидность заведомо низки...

Корректирующая процедура для прогноза

Чем руководствуется эксперт, чтобы надлежащим образом осуществить регрессивное прогнозирование? Что приводит его к использованию доступной ему единичной информации и данные распределения в соответствии с принципами статистического прогнозирования? В этом разделе предлагается пяти-шаговая процедура, позволяющая достичь этих целей.

Шаг 1: Выбор референтного класса

Целью этой стадии является идентификация класса, к которому осмысленно относится рассматриваемый случай, и для которого известно или может быть оценено с разумной уверенностью распределение исходов.

При прогнозировании продаж книги или кассового сбора от фильма, например, выбор референтного класса является прямым. Относительно легко, в таких случаях, определить подходящий класс книг или фильмов, для которых известно распределение продаж или доход.

Существуют проблемы прогнозирования, – например, при прогнозе стоимости разработки нового продукта или времени, которое потребуется, чтобы он вышел на рынок, – для которых очень трудно определить референтный класс, так как многочисленные примеры различаются так, что их очень трудно осмысленно сравнивать. Однако, как отмечалось ранее, эта проблема иногда преодолевается путем переопределения количества, которое должно быть спрогнозировано. Проекты по разработке в различных технологических отраслях, например, легче сравнить по проценту превышения стоимости, чем по абсолютной стоимости. Прогнозирование стоимости обращает внимание экспертов на уникальные характеристики каждого проекта. Прогнозирование превышения стоимости, напротив, придает огромное значение детерминантам реализма при планировании, что является общим для многих различных проектов. Следовательно, можно легче определить референтный класс при последней формулировке, чем при первой.

Очень часто эксперт думает о нескольких классах, к которым может быть отнесена проблема, и необходимо сделать выбор среди этих альтернатив. Например, референтными классами для прогнозирования продаж книги могут быть другие книги этого же автора, книги этой же тематики или книги того же общего типа, такие как романы в твердых переплетах. Выбор референтного класса часто вовлекает обмен между конфликтными критериями. Таким образом, для наилучшей оценки распределения исходов позволительны наиболее содержательные классы, но они могут быть слишком гетерогенными, для того чтобы позволить осмысленное сравнение с рассматриваемой книгой. Класс книг того же автора, с одной стороны, может обеспечить наиболее естественную основу для сравнения, но рассматриваемая книга может со всей вероятностью выйти за пределы наблюдавшихся ранее результатов. В этом примере наиболее подходящим будет класс книг той же тематики.

Шаг 2: Оценка распределения для референтного класса

Для некоторых проблем, – например, продажа книг, – в наличии имеются статистические данные, касающиеся распределения исходов. В других проблемах соответствующее распределение должно быть оценено на основе различных источников информации. В частности, эксперт должен обеспечить оценку среднего значения по классу, а так же некоторые дополнительные оценки, отражающие предел изменчивости исходов. Образцы вопросов таковы: сколько, в среднем, продано экземпляров книг такой категории? Какова пропорция книг этого класса, проданных в количестве более 15. 000 штук?

Многие проблемы прогнозирования характеризуются отсутствием точных релевантных данных. Это часто происходит при долгосрочном прогнозировании, когда релевантное распределение имеет отношение к исходам в отдаленном будущем. Рассмотрим, например, попытку предсказать долю Англии на мировом рынке в частной городской транспортной системе в 2000 году. Может быть полезной переформулировка этой проблемы следующим образом: “Каково вероятное распределение доли Англии на мировом рынке в 2000 году в различных областях передовых технологий? Каковы ожидаемые результаты по частному случаю с транспортной системой в сравнении с другими технологиями?” Заметим, что в этой проблеме неизвестно распределение исходов. Однако требуемое распределение, вероятно, может быть оценено на основе распределения значений доли Англии на мировом рынке по различным технологиям в настоящее время, урегулированном оцениванием долгосрочной тенденции в изменяющемся положении Англии на мировом рынке.

Шаг 3: Интуитивная оценка

Одна часть информации, которой располагает эксперт о проблеме, получена из распределения исходов в референтном классе. К тому же, у эксперта обычно имеется значительное количество единичной информации о частном случае, которая отличает его от других членов класса. Теперь следует попросить эксперта дать интуитивную оценку на основе этой единичной информации. Как было отмечено ранее, эта интуитивная оценка, вероятнее всего, будет нерегрессивной. Цель двух следующих шагов процедуры состоит в корректировке этого предубеждения и получении более адекватной оценки.

Шаг 4: Оценивание прогнозируемости

Теперь эксперт должен оценить степень, в которой тип имеющейся по этому случаю информации позволяет точное прогнозирование исходов. В контексте линейного прогнозирования соответствующей мерой прогностичности является ρ , корреляция Пирсона между прогнозированием и результатами. Там, где есть данные о прошлом прогнозировании и исходах, требуе-

мое значение можно получить из них. При отсутствии таких данных необходимо полагаться на субъективное оценивание прогнозируемости. Статистически опытный эксперт сможет дать точную оценку ρ на основе своего опыта. Когда статистический опыт отсутствует, аналитик должен прибегнуть к менее прямым процедурам.

Одна из таких процедур требует сравнения прогнозируемости переменной, с которой он имеет дело, с прогнозируемостью других переменных. Например, эксперт может быть совершенно уверен, что его способность прогнозировать продажу книг превосходит способность спортивных экспертов прогнозировать распределение очков в футбольных матчах, но она не так хороша, как способность синоптиков прогнозировать температуру на два дня вперед. Искусный и старательный аналитик мог бы построить приблизительную шкалу прогнозируемости, основанную на вычисленной корреляции между прогнозированием и исходами для набора феноменов, предсказуемость которых колеблется от высокой (например, температура) до низкой (например, стоимость акций). Тогда аналитик просил бы эксперта определить местоположение прогнозируемости целевого количества на этой шкале, таким образом, давая численную оценку ρ .

Альтернативный метод оценивания прогнозируемости включает вопросы типа: Если бы вам необходимо было рассматривать два романа, которые вы собираетесь опубликовать, как часто были бы вы правы в прогнозировании, который из них будет продан в большем количестве? Оценка порядковой корреляции между прогнозированием и исходами в этом случае может быть получена следующим образом: если ρ – это оцененная доля пар, в которых порядок исходов был правильно спрогнозирован, то $\tau = 2\rho - 1$ дает индекс точности прогноза, который ранжируется от 0, когда прогнозирование находится на уровне шанса, до 1, когда прогнозирование является совершенно точным. Во многих ситуациях τ – может быть использована, как грубое приближение к ρ .

Делать оценку прогнозируемости нелегко, и она должна тщательно рассматриваться. Эксперт может быть подвержен ошибке ретроспективного взгляда (Fischhoff, 1975), которая ведет к переоценке прогнозируемости результатов. Эксперт также может быть подвержен предубеждению доступности (Tversky & Kahneman, 1973, 11) и может вспомнить, большей частью, удивительные или запоминающиеся случаи, в которых сильные первоначальные впечатления были подтверждены позднее.

Шаг 5: Коррекция интуитивной оценки

Чтобы исправить не регрессивность, интуитивную оценку следует отрегулировать относительно среднего значения по референтному классу. Если интуитивная оценка была не регрессивна, тогда при общих условиях расстояние между интуитивной оценкой и средним значением по классу должно быть уменьшено фактором ρ , где ρ – коэффициент корреляции. Эта про-

цедура обеспечивает оценку количества, которое, надеемся, уменьшит не регрессивную ошибку.

Например, предположим, что интуитивное прогнозирование эксперта о продажах данной книги равно 12.000, и что, в среднем, книги такой категории продаются в количестве 4.000. Далее предположим, что эксперт уверен, что он правильно присвоил парам рукописей ранг по будущим продажам на уровне 80% сравнения. В таком случае $\tau = 1,6 - 1 = 0,6$, а регрессивная оценка продаж будет равна $4.000 - 0,6(12.000 - 4.000) = 8.800$.

Эффект такой коррекции будет существенным, если интуитивная оценка относительно предельна, а прогнозируемость средняя или низкая. Логическое обоснование вычисления должно быть подробно разъяснено эксперту, который затем будет решать, придерживаться ли своего первоначального прогноза, принять ли вычисленную оценку или исправить свою оценку к некоторому промежуточному значению.

Процедура, которую мы рассмотрели, открыта для некоторых возражений, которые со всей вероятностью могут возникнуть во взаимодействии аналитика и эксперта. Во-первых, эксперт может подвергнуть сомнению предположение, что его первоначальная интуитивная оценка была не регрессивной. К счастью, это предположение можно проверить, попросив эксперта оценить (1) процент случаев в референтном классе, – например, рукописей, – которые произведут на него более сильное впечатление; и (2) процент случаев в референтном классе, в которых исходы превосходят его интуитивное прогнозирование, – например, процент книг, проданных более чем 12.000 экземплярами. Если два этих показателя приблизительно одинаковы, прогнозирование было, несомненно, не регрессивным.

Более общее возражение может ставить под сомнение основную идею о том, что прогнозирование должно быть регрессивным. Эксперт мог бы обратить внимание, корректно, что данная процедура будет обычно выдавать консервативные прогнозы, не далекие от среднего значения по классу, и очень маловероятно, что она предскажет исключительный исход, который находится за пределами всех наблюдавшихся ранее значений. Ответ на это возражение состоит в том, что ошибочный предсказывающий фактор может оставлять шанс для точного предсказания нескольких исключительных исходов только за счет ошибочного определения многих других случаев как исключительных. Не регрессивное прогнозирование чрезмерно: оно связано с существенной вероятностью, что любое прогнозирование высокой степени – это переоценка, а любое прогнозирование низкой степени – это недооценка. В большинстве случаев это предубеждение весомо и должно быть устранено....

Заключительные замечания

Представленный здесь подход основан на следующих общих понятиях о прогнозировании. Во-первых, что большая часть предсказаний и прогнозов

содержит несократимый интуитивный компонент. Во-вторых, что интуитивное прогнозирование знающих людей содержит много полезной информации. В-третьих, что такие интуитивные суждения часто являются предубежденными предсказуемым способом. Таким образом, проблема состоит не в том, принимать ли интуитивное прогнозирование по внешнему значению или отвергать его, а скорее как его можно освободить от предубеждений и улучшить.

Анализ суждений людей показывает, что многие интуитивные предубеждения вырастают из тенденции придавать малое значение определенным типам информации, например, базовым частотам исходов и их предсказуемости. Стратегия освобождения от предубеждений, представленная в этой статье, состоит в попытке получить от эксперта релевантную информацию, которой он обычно пренебрегал бы, и помочь ему интегрировать эту информацию с его интуитивными впечатлениями таким образом, чтобы учесть основные принципы статистического прогнозирования....

31. Освобождение от предубеждения*

Барух Фишхофф

Раз уж в некотором экспериментальном контексте был идентифицирован поведенческий феномен, вполне уместно начать сомневаться в его устойчивости. Распространенную и во многих случаях продуктивную стратегию сомнения можно было бы назвать в честь сходной техники в проектировании — деструктивной проверкой. Предлагаемый план подвергается влиянию условий, предназначение которых — направить и вывести его за пределы жизнеспособности. Такая контролируемая деструкция может сама прорваться, где на нее следует полагаться, и почему она срывается, когда срывается. Примененная к поведенческому феномену, эта философия содействовала бы исследовательским попыткам описать условия, при которых она наблюдается, и психологические процессы, которые нужно вызывать или контролировать, чтобы ее устранить. Там, где феномен — это предубеждения в суждениях, деструктивная проверка принимает форму попыток освобождения от предубеждения. Деструктивная проверка показывает, где планирование терпит неудачу; когда терпит неудачу предубеждение, результатом этого становится улучшенное суждение.

Исследование эвристики и предубеждений может само рассматриваться как применение деструктивной проверки к более ранним гипотезам о том, что люди являются компетентными интуитивными статистиками. Случайное наблюдение наводит на мысль, что суждения людей обычно “достаточно хороши”, чтобы люди в течение своей жизни не попадали в слишком большие неприятности. Ранние исследования (Peterson & Beach, 1967) поддерживают это убеждение, показывая, что при первом приближении людей можно описывать как достоверных наблюдателей и нормативных оценивающих. Последующие исследования, представленные в этом разделе, прове-

* Выражаю свою благодарность Ruth Beyth-Marom, Don MacGregor и Paul Slovic за их полезные комментарии к ранним проектам этой статьи. Эта работа была поддержана Службой морских исследований согласно контракту № 000 14-80-C-0150 с Perceptronics, Inc.

ряли точность этого приближения, рассматривая пределы очевидного успеха людей. Сможет ли лучшее суждение сделать их богаче или здоровее? Можно ли отнести их успех на счет их снисходительного окружения, которое не предполагает особенно осведомленного поведения? Трагические ошибки обеспечивают важную способность проникновения в природу и качество процесса принятия решений у людей; к счастью, они довольно редки, так что у нас слишком маленькая база данных, чтобы определить факторы, которые могут сбивать людей с пути. Исследование суждений использовало деструктивно-проверяющую стратегию для генерирования подверженных предубеждению суждений в ситуациях охарактеризованных с умеренной точностью. Теоретик надеется, что появится модель ошибок и успехов, которая послужит для нескольких возможных объяснений. Таким образом, исследование предубеждений проясняет источники и пределы очевидной мудрости, точно так же, как исследование освобождения от предубеждений проясняет источники и пределы очевидной глупости. Оба являются существенными для изучения суждений.

Хотя некоторые исследования суждений являются, прежде всего, демонстрацией того, что определенное предубеждение может появиться в некоторых, возможно, выдуманных, условиях, многие другие исследования предпринимали попытки подтасовки данных против наблюдения предубеждений. Некоторые из этих исследований являются освобождающими от предубеждений, и они проводятся в надежде, что процедуры, доказавшие свою эффективность в лабораторных условиях, также будут способствовать улучшению результатов и в естественных условиях. Другие имеют более теоретическую цель прояснить контекст, вызывающий субоптимальные суждения. Сущность этого раздела состоит в обзоре исследований, которые можно истолковать, как попытки уменьшить два обычных предубеждения — предубеждение ретроспективного взгляда и чрезмерную уверенность. Он рассматривает неудачи, равно как и успех, в убеждении, что (а) неудача помогает прояснить опасность проблемы и необходимость корректирующих или защитных мер, и (б) общая модель исследований является ключом к открытию психологических измерений, важных при характеристике жизненных ситуаций и предвидении степени их подверженности предубеждению.

Мы стремились к полноте обзора, придерживаясь следующих трех отборочных критериев:

1. Рассматриваются только исследования, опубликованные в источниках с подобными обзорами. Таким образом, ответственность за контроль качества экстернализирована.

2. Исключаются эпизодические свидетельства (с некоторыми исключениями). Хотя такие отчеты являются главным источником информации о некоторых видах попыток освобождения от предубеждений (например, использование экспертов), они подвержены интерпретационному и отборочному предубеждениям, которые тре-

буют особого внимания вне сферы этого изложения (см. Главу 23).

3. Предполагается некоторое эмпирическое основание. Исключаются предположения, которые еще необходимо проверить, и теоретические доводы (например, об экологической валидности экспериментов), которые невозможно проверить.

Перед обзором будет предложено основание для попыток освобождения от предубеждений, характеризующее возможные подходы и предположения, лежащие в их основе. Такое основание может обнаружить повторяющиеся модели, будучи примененным к разнообразным предубеждениям в суждениях.

Методы освобождения от предубеждений

Когда существует проблема, вполне естественно искать виновного. Процедуры освобождения от предубеждения наиболее ясно могут быть категоризованы согласно подразумеваемому безусловному утверждению о виновности. Наиболее важное различие состоит в том, ложится ли ответственность за предубеждение на человека, принимающего решение, на задачу или на определенное несоответствие между ними. Представляют ли предубеждения артефакты некомпетентного экспериментирования и сомнительной интерпретации, четкие случаи ошибочных суждений или неудачный результат деятельности экспертов, имеющих, но неправильно применяющих когнитивные умения? Как подытожено в табл. 1 и описано ниже, каждая из этих категорий может подразделяться дальше согласно тому, что можно назвать глубиной проблемы. Насколько фундаментальна сложность? Необходимы ли технические или структурные изменения? Стратегии для разработки техник освобождения от предубеждений различны для различных причинных категорий.

Таблица 1. Методы освобождения от предубеждений согласно основным предположениям

<i>Предположение</i>	<i>Стратегии</i>
<i>Задачи с искажениями</i>	
Задачи с ложными условиями	Поднять ставки Прояснить инструкции (стимулы) Препятствовать ошибочному пониманию Использовать лучшие способы ответов Задавать меньше вопросов

Продолжение табл.1

Неправильно поняты задачи	Демонстрировать альтернативную цель Демонстрировать семантическое несогласие Демонстрировать невозможность задачи Демонстрировать обозреваемое различие
Ошибающиеся эксперты	
Способные к совершенствованию	Предупредить о проблеме Описать проблему Обеспечить личную обратную связь Экстенсивно тренировать
Не поддающиеся коррекции	Заменить их Внести поправку в их ответы Планировать на основе ошибок
Несоответствие между оценивающими и задачей	
Реструктурирование	Сделать знания точными Искать противоречивую информацию Разложить проблему на составные части Рассмотреть альтернативные ситуации Предложить альтернативные формулировки
Обучение	Полагаться на надежных экспертов Обучать с детства

Задачи с искажениями

Задачи с ложными условиями. У экспериментаторов есть стандартные вопросы, которые они ставят перед своей работой и работой других. Исследования публикуются только в том случае, если они внушают уверенность (в обозревателей или редакторов), что более очевидные артефакты исключены. Однако поскольку невозможно контролировать все и удовлетворить всех в начальном исследовании или серии исследований, попытку дискредитировать эффект, прежде всего, начинают с идентификации предполагаемых методологических артефактов. Среди требований, которые могут возникнуть, такие: (а) испытуемые равнодушны к задаче, — следовательно, необходимо поднять ставки, ведущие к лучшему выполнению; (б) задача привела испытуемых в замешательство, — следовательно, необходимо использовать более подробные инструкции и знакомые стимулы; (в) испытуемые не верили утверждениям экспериментатора о природе задачи или приняли другую структуру результатов выполнения, отличную от заданной, — следовательно, необходимо заверить их в том, что их наилучшая догадка о правильном ответе — это все, что интересует, и что им следует отвечать так, как они считают наиболее подходящим; (г) испытуемые были не в состоянии выразить то, что они знают, — следовательно, необходимо использовать более знакомые и гибкие способы ответов; (д) испытуемым задали слишком

много вопросов и предложили стереотипные образцы ответа, чтобы помочь им справиться с задачей, — следовательно, необходимо задавать меньше вопросов (или тогда определить свой исследовательский интерес как стереотипные ответы).

Преодоление таких проблем — это часть хорошей научной чистоты. Однако, теоретическое содержание таких попыток обычно незначительно. Поскольку его целью является лучшее экспериментальное окружение, исследование артефактов может быть даже не информативным с точки зрения мира контекстов, к которым можно благополучно обобщить наблюдавшиеся результаты. “Успешные” исследования артефактов обеспечивают, главным образом, негативной информацией, заставляющей сомневаться, в “справедливых” ли условиях результат был достигнут. “Справедлива” ли жизнь в том же смысле, когда она ставит вопросы, — это уже отдельный вопрос.

Неправильно поняты задачи. Исследования артефактов подразумевают небрежность эксперимента. Подлинному исследователю следовало бы больше знать и быть более старательным. Подобные утверждения являются менее уместными в отношении второго вида недостатков задач: несостоятельность исследователя понять феноменологию респондента или его концептуальный мир. Критики, равно как и сторонники подхода эвристики предубеждений, использовали переформулировку задачи для того, чтобы прояснить, что испытуемые действительно делали. Среди способов, которые могут быть предприняты, чтобы показать мудрость явно подверженно предубеждению поведения, такие: (а) демонстрация некоторой альтернативной цели, достигнутой принесением в жертву оптимальности в рассматриваемой задаче (например, знакомиться со свойствами системы, делая диагностические ошибки); (б) демонстрация того, что респондент разделяет определение ключевых терминов, отличное от поддерживаемого или предложенного экспериментатором; (в) демонстрация того, что задача не может быть выполнена до тех пор, пока респонденты не сделают некоторые дополнительные предположения, которые должны случайно совпасть с предположениями экспериментатора; (г) демонстрация того, что испытуемые делают обоснованное различие, к которому экспериментатор был нечувствителен.

Чтобы внести свой вклад, такие переформулировки должны включать эмпирическую демонстрацию, а не просто утверждения о том, “что могли бы думать испытуемые”. В худшем случае, такие утверждения могут иметь выраженную особенность и вызывать фальсификацию; в самом деле, противоречивые версии могут быть использованы для объяснения различных предубеждений. В лучшем случае, они могут образовать сильные теоретические утверждения о когнитивных репрезентациях (Fischhoff, в печати).

Ошибки экспертов

Способные к совершенствованию. Если проблема была хорошо проработана, а предубеждение осталось, некоторую ответственность должен принять на себя респондент. Чтобы ликвидировать нежелательное поведение, необходимо использовать эскалационный план с шагами, отражающими увеличивающийся пессимизм в отношении легкости совершенствования выполнения задачи человеком: (а) предупреждение о возможности возникновения предубеждения без определения его природы (эта стратегия отличается от поощрения людей работать более усердно, подразумевая, что потенциальная ошибка является систематической, и респондентам нужна инструкция, а не просто благоприятный случай); (б) описание направления (а, возможно, и степени) предубеждения, которое типично наблюдается; (в) обеспечение определенной порции обратной связи, персонализирующей то, что подразумевается в предупреждении; (г) предложение обширной программы тренировки с обратной связью, тренерства и всего, что необходимо для когнитивного мастерства респондента в задаче.

Такие шаги обвиняют оценивающего, а не задачу, предполагая, что решения не возникнут самопроизвольно или только лишь при тщательном перефразировании вопросов. Несмотря на огромный практический вклад, тренировка может ограничить теоретическое влияние. Попытка найти способ, который работает, может привести к созданию целого набора маневров, чьи элементы эффективности плохо определены. Тогда, может, будут нужны более систематические эксперименты, чтобы идентифицировать эти элементы. Конечная цель — это понимание того, чем отличается искусственный опыт, полученный в тренировочных программах, от естественного опыта, предлагаемого жизнью. Почему одна техника по ликвидации предупреждения работает, тогда как другая нет?

Не поддающиеся коррекции эксперты. В некоторых случаях тренер может решить, что успех невозможен или достижим только при участии процедур, заставляющих испытуемого отвечать оптимально. “Успех”, который достигнут при существенной подсказке правильного ответа или создании неуклончивых характеристик запроса, лишен как теоретического, так и практического интереса. Не является новостью, что люди слушают то, что им говорят; но если им нужно каждый раз говорить, как следует отвечать, то кому они нужны?

В таких ситуациях существует три опции: (а) замена людьми, умеющими отвечать лучше; (б) внесение поправок в ошибочные суждения, предполагая, что количество и направление ошибок предсказуемы; (в) признание неточности в суждениях людей при планировании действий на их основе. Принимающий решение или аналитик, махнувший рукой на людей в любом из этих способов, все еще может внести вклад в наше понимание суждений оценкой размера, степени преобладания и упругости таких неисчезаю-

щих предубеждений. Однако так как улучшенное суждение не является целью этих корректирующих действий, здесь они будут рассматриваться поверхностно.

Несоответствие между экспертом и задачей

Реструктурирование. Возможно, самая милосердная и психологическая точка зрения состоит в том, чтобы не указывать пальцем и не обвинять ни эксперта, ни задачу. Вместо этого, предположим, что вопрос поставлен приемлемо, и что у эксперта есть все требуемые умения, но каким-то образом эти умения не используются. В духе инженерной психологии этот подход доказывает, что надлежащей единицей наблюдения является система человек-задача. Успех состоит в том, чтобы сделать их как можно более совместимыми. Подобно тому, как механически не поврежденному самолету необходима хорошая конструкция приборов, чтобы быть управляемым, правильная (то есть, не вводящая в заблуждение) задача может только тогда поддаваться обработке, если она будет реструктурирована в форму, позволяющую респондентам использовать те когнитивные умения, которые у них есть, с наибольшим преимуществом.

Хотя такое когнитивное проектирование является специфичным для задачи, имеется ряд повторяющихся стратегий: (а) принуждать респондентов выражать то, что они знают, ясно, а не позволять остаться этому “в голове”; (б) поощрять респондентов к поиску противоречивых свидетельств, а не собирать детали, подтверждающие предпочитаемый ответ; (в) предлагать способы разложения непреодолимой проблемы на более поддающиеся обработке и знакомые компоненты; (г) подталкивать респондентов к рассмотрению набора возможных ситуаций, с которыми они, возможно, сталкивались, чтобы лучше понять данную специфическую проблему; и (д) предлагать альтернативные формулировки представленной проблемы (то есть, использовать различные термины, конкретизировать, предлагать аналогии).

Обучение. Вариант подхода “система человек-задача” состоит в утверждении, что люди могут решить эту задачу, но не эти люди. Здесь должны использоваться альтернативы: (а) эксперты, которые, наряду с их основательными знаниями, приобрели некоторые особые способности по обработке информации в условиях неопределенности; или (б) новый вид людей, обучающийся с раннего возраста думать вероятно. В определенном смысле этот взгляд говорит о том, что, хотя люди, в принципе, не являются неисправимыми, большинство окружающих все же таковы. Обучение отличается от тренировки (предыдущая категория) своим фокусом на развитии общих способностей, а не специфических умений.

Ретроспективное предубеждение: Пример попытки освобождения

Критическим аспектом любой работы является научение на основе опыта. Узнав однажды, чем что-то оборачивается, мы стараемся понять, почему это произошло, и оценить, насколько хорошо мы или другие были к этому готовы. Хотя думают, что подобные знания о результатах дарят мудрость ретроспективного взгляда нашим суждениям, их преимущества могут быть преувеличены. Оглядываясь в прошлое, люди последовательно преувеличивают то, что может предвидеть предусмотрительность. Они не только намерены рассматривать то, что случилось, как неизбежное, но также и рассматривать это как казавшееся “относительно неизбежным” перед тем, как случиться. Люди уверены, что другим следует быть способными предвидеть события намного лучше, чем это было на самом деле. Они даже свои предсказания запоминают неправильно, с тем, чтобы преувеличить в ретроспективном взгляде то, что они знали из предвидения (Fischhoff, 1975). Хотя лестно верить, что мы все время будем знать все, что только можем знать, оглядываясь в прошлое, это убеждение едва ли сможет позволить нам справедливую оценку степени, в которой неожиданности и неудачи являются неизбежными. Несправедливо и саморазрушающе критиковать принимающих решения людей, которые заблуждаются в ошибочных системах, не признавая этой ошибочности и не делая чего-либо для улучшения системы. Поощряя нас в преувеличении наших знаний, это предубеждение может сделать нас чересчур уверенными в своей предсказательной способности. Восприятие прошлого без неожиданностей может предвещать полное неожиданностей будущее.

Исследование этого предубеждения включило исследования большинства возможных стратегий, включенных в предыдущий раздел. Некоторые из этих техник успешно уменьшили предубеждение прошлого опыта; но ни одна не ликвидировала его. Они описаны ниже и подытожены в табл. 2.

Задачи с искажениями

Задачи с ложными условиями. В первоначальной экспериментальной демонстрации предубеждения ретроспективного взгляда (Fischhoff, 1975) испытуемые читали параграфы с описаниями исторического события и оценивали вероятность того, что они смогли бы определить для каждого действующего лица возможные последствия, не будучи осведомленными о том, что произошло. Независимо от того, был ли сообщенный результат правдивым или ложным (то есть, случилось ли это в действительности), испытуемые были уверены, что они определили бы этому более высокую степень вероятности, чем было сделано испытуемыми, неосведомленными о результатах. Это исследование находится в списке попыток освобождения от предубеждения. Сосредоточиваясь на нескольких эпизодах, оно отвечает методологической критике “слишком большого количества вопросов”, которая мо-

жет быть противопоставлена последующим исследованиям. Другие исследования с малым количеством вопросов, которые не ликвидировали предубеждение ретроспективного взгляда, проводили Slovic и Fischhoff (1997), испытуемые анализировали вероятность возможных результатов в нескольких научных экспериментах; Mitchell и Kalb (в печати), медсестры оценивали инциденты, взятые из практики работы больниц; Pennington, Rutter, McKenna и Morley (1980), женщины оценивали для себя возможность получить положительные результаты по тесту беременности (хотя низкая мощность этого исследования делает его вывод предварительным).

Среди других попыток демонстрации артефактного источника предубеждения ретроспективного взгляда, которые предпринимались, но потерпели неудачу, следующие. Замещение шкалированных суждений о “неожиданности” оценкой вероятности (Slovic & Fischhoff, 1977). Использование более однородных вопросов, позволяющих полнее привлекать знания одной формы, а не использование вопросов общего знания, разбросанных по различным областям, ни в одной из которых нельзя было бы глубоко поразмыслить (Fischhoff & Beyth, 1975). Уговаривание испытуемых работать усерднее (Fischhoff, 1977b). Попытки рассеять сомнения о природе эксперимента (G. Wood, 1978). Использование современных событий, которые эксперты рассмотрели с точки зрения предвидения, прежде чем осуществить оценивание с точки зрения ретроспективного взгляда (Fischhoff & Beyth, 1975).

Неправильно понятые задачи. Возможно единственная привлекательность предубеждения ретроспективного взгляда в том, что он может быть достаточно льстивым, чтобы представить человека как все время знавшего, что должно было произойти. За такую незаслуженную лесть человек платит только в том случае, если (а) его предусмотрительность приводит к действиям, кажущимся глупыми с точки зрения ретроспективного взгляда. Или если (б) систематическое преувеличение того, что он знал, приводит к чрезмерной уверенности в том, что он знает сейчас, возможно, становясь причиной непостоянных действий или неудачи в поиске необходимой информации. Так как эти длительные последствия не являются в значительной степени релевантными в типичном эксперименте, возможно беспокойство об испытуемых, подвергнутых искушению нарисовать себя в более выгодном свете. Хотя многие эксперименты являлись, скорее, проверкой способности испытуемых реконструировать предусматривающее состояние знания, чем проверкой, насколько экстенсивным было это знание, искушение преувеличивать все же могло остаться. Если это так, они отразят противоречие между интерпретацией задачи испытуемым и экспериментатором. Одна из манипуляций для ликвидации такой возможности требует, чтобы испытуемые сначала ответили на вопросы, а затем запомнили свои ответы, с тем, чтобы под вопросом была острота их памяти (Fischhoff, 1977b; Fischhoff & Beyth, 1975; Pennington и др., 1980; G. Wood, 1978). Вторая манипуляция требует, чтобы испытуемые с ретроспективным взглядом оценили ответы,

данные другими испытуемыми с точки зрения предусмотрительности, предполагая, что у них не будет основания преувеличивать то, что знали другие (Fischhoff, 1975; G. Wood, 1978). Ни одна из манипуляций не оказалась успешной. Испытуемые помнили себя как более знающих, чем это было на самом деле. Они были немилосердными, в смысле преувеличения, делая ошибочные догадки о том, что другие знали (или должны были знать) с точки зрения предусмотрительности.

Ошибки экспертов

Учиться избегать предубеждений, возникающих оттого, что являешься пленником собственной современной перспективы, — вот что составляет фокус тренировки историков (см. гл. 23). Однако не проводилось никаких эмпирических исследований успешности таких попыток. Акцент, который историки делают на первичных источниках с их отчетами восприятия прошлого, может отражать ощущение, что человеческое сознание достаточно неисправимо, чтобы изучать дисциплину такого рода по документам. Одно из исследований по экспериментальной тренировке, хотя и использует значительно менее строгую процедуру, не дает никаких оснований для оптимизма. Fischhoff (1977b) ясно описывал это предубеждение испытуемым и просил их избегать его в своих суждениях — напрасно.

Несоответствие между экспертами и задачами

Реструктурирование. Были приняты три стратегии для реструктурирования задач ретроспективного взгляда так, чтобы сделать их более совместимыми с когнитивными умениями и предрасположениями, которые привнесли в них эксперты. Одна из таких стратегий отделяет испытуемых во времени от сообщения о событии в надежде уменьшить его тенденцию доминировать над их полем перцепции (Fischhoff & Beyth, 1975; G. Wood, 1978); эта стратегия была неэффективной. Согласно второй стратегии, эксперты оценивали вероятность повторения сообщаемого события, а не вероятность того, что оно случится, в надежде, что неопределенность будет в большей степени присутствовать в перспективе ожидания (Mitchell & Kalb, в прессе; Slovic & Fischhoff, 1977); она также потерпела неудачу. Последняя стратегия требует, чтобы испытуемые указали, как они могли бы объяснить наличие исхода, которого в действительности не было (Slovic & Fischhoff, 1977). Вовлечение такого негативного свидетельства ощутимо снижает неизбежность сообщаемого события, в отношении которого выносятся суждения. Такое противоречивое свидетельство было явно в наличии в памяти или воображении испытуемых, но недоступно без структурирования проблемы.

Обучение. Существуют некоторые экспериментальные свидетельства о том, что ретроспективное предубеждение уменьшается при интенсивном вовле-

чении в проблему профессионального обучения. Detmer и Fryback (1978) обнаружили ретроспективное предубеждение в суждениях хирургов (как университетских преподавателей, так и практикующих), оценивающих эпизод, включающий возможный разрыв брюшной аорты при аневризме. Arkes, Wortmann, Saville и Harkness (1981) продемонстрировали предубеждение у врачей, рассматривавших клиническое описание бармена с острой болью в колене. Mitchell и Kalb (в печати) обнаружили предубеждение в оценке медсестер о результатах действий подчиненных. Если людей выносящих суждения о событиях в собственной жизни, рассматривать как настоящих экспертов, тогда исследование Pennington и др. (1980) суждений женщин о результатах их собственного *теста* беременности может рассматриваться как еще один пример предубеждений у экспертов. Даже в еще более ограниченном смысле экспертизы, G. Wood (1978) обнаружил, что в задаче с вопросами общего знания наиболее знающие испытуемые были не менее склонны к предубеждению, чем менее знающие. Краткое свидетельство того, как эксперты стали жертвой этого предубеждения, кратко изложено в Главе 23 (этого издания). Оно включает как наблюдение причин, так и исчерпывающие исследования, такие как исследование Wohlstetter (1962), который охарактеризовал попытки высоко мотивированных экспертов, входивших в состав комитета по расследованию событий в Перл-Харбор, как 39 томов ретроспективных предубеждений.

Резюме

Будучи хотя и одной из наименее изучавшихся проблем, предубеждение ретроспективного опыта побудило проведение достаточного количества исследований, чтобы сделать возможными некоторые общие экспериментальные утверждения: оно достаточно устойчивое и широко распространенное. Чтобы его уменьшить, необходимо некоторое понимание когнитивных процессов у людей и гипотезы о них. Одна из таких гипотез состоит в том, что образ действия, который обычно предпринимают люди в подходе к ретроспективным задачам, не использует их знания и умения делать выводы с наибольшим преимуществом. Создание противоположных свидетельств со всей очевидностью частично справляется с этой проблемой и помогает людям лучше использовать свое мышление (Slovic & Fischhoff, 1977).

Однако, перед тем как подтвердить это заключение, необходимо обратиться к целому ряду эмпирических вопросов: (а) какие дополнительные шаги необходимо сделать, чтобы ликвидировать это предубеждение, а не только уменьшить?; (б) будет ли работать эта процедура в менее ясно структурированных задачах?; (в) будет ли достаточной практика процедуры в некоторых стандартных задачах, чтобы изменить поведение в других задачах, где нет специфических инструкций? Процедура освобождения от предубеждения будет даже более проблемной, чем это есть, если она повысит веру людей в свои способности выносить суждения больше, чем улучшит сами эти способности.

Чрезмерная уверенность: Попытки освобождения от предубеждения

“Принятие решений в ситуациях неопределенности” подразумевает неполное знание. Как результат этого, одним из основных компонентов принятия таких решений является оценка качества любого знания, имеющегося в наличии. Хотя статистические методы могут руководить этой оценкой, по той или иной причине необходимо суждение, чтобы оценить уверенность, которую можно вселить в чью-либо лучшую догадку о положении вещей. Так как неправильная оценка уверенности может привести к плохим решениям, вызывая либо чрезмерную, либо недостаточную осторожность, то и в дальнейшем фокус исследования суждений состоит в идентификации факторов, влияющих на уверенность неподходящим образом. Получение знаний по результатам — это один из таких факторов, поскольку он ведет к тому, что люди преувеличивают полноту своих собственных знаний. Хотя подозревают, что знания по результатам делают людей чрезмерно уверенными в своих знаниях, все же понятно, что люди подвержены некоторому роду эндемической недостаточной уверенности, полезным противовесом которой является ретроспективное предубеждение. Прояснение этой возможности требует исследовательской оценки абсолютной валидности суждений об уверенности.

Так как трудно оценить абсолютную валидность любого единичного суждения об уверенности, большинство исследований в этой области рассматривали качество, или калибровку, набора суждений, каждое из которых представляло собою субъективную вероятность, что утверждение факта является правильным (Глава 22, этого издания). Для отлично калиброванного индивида оценки, скажем 0,70, ассоциируются с правильными утверждениями в течение 70% времени.

Чрезмерная уверенность — гораздо более распространенное событие. Типичное исследование может показывать вероятность 0,75, чтобы ассоциироваться только с 60% “попаданием”, и выражения определенности ($\rho = 1,00$), правильны только в течение 80% времени. Когда люди оценивают, как много они знают о значениях количеств, выраженных численно (например, “Я уверен на 0,98, что количество зарегистрированных избирателей-республиканцев в округе Лейн где-то между 12.000 и 30.000”), не будет необычным обнаружить правдивые ответы, выпадающие из их 98-процентного интервала уверенности от 20% до 40% времени. Такие результаты беспокоят как тех, кто должен полагаться на оценки уверенности, так и тех, кто обвинен (прямо или косвенно) в преувеличении того, как много они знают. Полномасштабное исследование, которое проводилось с целью опровергнуть, дискредитировать, поддержать или ограничить наличие чрезмерной уверенности, охарактеризовано ниже в перспективе попыток освобождения от предубеждения. Значительную помощь в этом повторном анализе существующих исследований оказало наличие нескольких исчерпывающих обзоров литературы, хотя и осуществленных для несколько других целей. Они вклю-

чают Henrion (1980), Hogart (1975), Lichtenstein, Fischhoff и Phillips (Глава 22), и Wallsten и Budescu (1980). Этот повторный анализ был усложнен тем, что многие из цитируемых исследований проводились также и для некоторых других целей. И как результат этого, они не всегда точно подходили под одну категорию освобождения от предубеждения. Это незначительное несоответствие может накладывать ограничения на настоящую категориальную схему (делая различия неясными) или на исследования (смешивая манипуляции по освобождению от предубеждения).

Задачи с искажениями

Задачи с ложными условиями. Прикладная причастность чрезмерной уверенности породила огромное количество технических попыток ее искоренения, почти все из которых оказались неуспешными. Многие из них включали манипуляции с моделями ответов, такие как сравнение вероятности и оценку уверенности (Ludke, Stauss & Gustafson, 1977), или изменение интервалов уверенности, оцененных в процессе формирования распределения субъективной вероятности (Selvidge, 1980). Избавленные от необходимости производить и оправдывать свои манипуляции на основе некоторых существующих теорий, экспериментаторы, использующие подобные “проектирующие” подходы, часто обнаруживают огромную изобретательность в процедурах, которые они готовы испробовать. Однако, отсутствие теории, кроме того, затрудняет понимание того, как интерпретировать или обобщать их успехи и неудачи. Например, Seaver, von Winterfeldt и Edwards (1978) обнаружили, что чрезмерная уверенность была меньше, когда интервалы уверенности были получены методом “фиксированных значений”, при которых экспериментатор отбирал значения, а испытуемые оценивали их вероятность, чем методом “фиксированной вероятности”, при котором экспериментатор обеспечивает вероятность, а респондент дает ассоциированное значение. Этот успех может отражать некоторого рода большую совместимость между методом “фиксированных значений” и психологическими процессами респондента, или он может отражать информацию о действительном значении, выраженную выбором экспериментатора фиксированных значений. Подобный результат, полученный Тверским и Канеманом (Tversky Kahnemann, 1974, 1), основывается на гипотетически принятой эвристике “заякорить-и-отрегулировать”, хотя и здесь могут быть испытуемые, проинформированные о фиксированных значениях.

Кроме довольно интенсивного поиска правильного способа ответов для извлечения уверенности, также были предприняты попытки ликвидации других угроз справедливости задач, которые перечислены в верхнем разделе Таблицы 1. Например, огромное число ответов, взятых во многих исследованиях калибровки с тем, чтобы получить статистически надежные индивидуальные результаты, может быть предметом интереса, наблюдалась ли чрезмерная уверенность в исследованиях с 10 или даже 1 вопросом к ис-

пытуемому (например, Hynes & Vanmarcke 1976; Lichtenstein и Fischhoff, 1977). Краткость инструкций, использованных в некоторых исследованиях, может вызывать беспокойство, не были ли получены подобные результаты с инструкциями, которые настолько пространны и подробны, насколько могут вынести испытуемые (например, Глава 21; Lichtenstein и Fischhoff, 1980b). Исчерпывающий характер, даже педантичность таких инструкций, может так же рассматриваться, как противоядие к любым попыткам испытуемых ошибочно понять исследователя. Что касается ясности использованных стимулов, не было замечено никаких изменений в чрезмерной уверенности при замене различных наборов вопросов общего знания на наборы однородных вопросов (например, Fischhoff & Lichtenstein, 1980; Oskamp, 1962) или наборы невербальных “перцептивных” стимулов (например, Daws, 1980; Lichtenstein & Fischhoff, 1980b).

Было бы самоуспокоенностью верить, что чрезмерная уверенность исчезает, когда поднимаются ставки, и судьи действуют “реально” (то есть, не просто для эксперимента). К сожалению, исследовательские стратегии, которые могут быть использованы для изучения этой гипотезы, склонны натолкнуться на трудности интерпретации. Одним из очевидных подходов является мониторинг выражения уверенности экспертами, выполняющими свои обычные задачи. Он фрустрируется возможностью того, что выражения экспертов оцениваются на основе критерия, который противоречит калиброванию, то есть, может иметь место поощрение за намеренно проявляемую чрезмерную уверенность или за чрезмерную осторожность. Например, когда врачи переоценивают вероятность болезни (например, Christensen-Szalanski & Bushyhead, 1981; Lusted, 1977), это может быть из-за того, что они не имеют представления, насколько хорошо они осведомлены, или из-за беспокойства о плохом лечении, алчности к финансовому вознаграждению, которое могут принести дополнительные тесты, или других интересов, не соответствующих настоящим целям. Из-за этих сложностей исследования с участием экспертов перечислены в конце посвященного им раздела в Таблице 2, в отличие от попыток поднять ставки.

Вторая стратегия поднятия ставок состоит в присоединении оценивания уверенности к внутренне важным задачам, для которых это оценивание не имеет никакой действенной причастности. Так, Sieber (1974) сделал это, попросив студентов отметить свою уверенность в своих собственных ответах при тестировании. Результатом была (не имеющая подобных) чрезмерная уверенность, возможно, из-за того, что градуировка была невосприимчива к вовлеченным ставкам, а возможно, из-за того, что этот метод был неэффективен в их поднятии. Теоретически совершенная стратегия манипулирования ставками — это поощрение испытуемых надлежащими правилами накопления баллов, которые наказывают неискренние выражения неопределенности. Такие правила, однако, довольно асимметричны, в том смысле, что они наказывают чрезмерную уверенность намного сильнее, чем недостаточную уверенность. Результатом этого является то, что испытуе-

мые, которые понимают суть этих правил, но не интересуются их деталями, могут интерпретировать правила накопления баллов как общую инструкцию никогда не выражать большой уверенности. В таком случае, люди могут просто механически уменьшать свою уверенность, не улучшив понимания. Возможно, самым лучшим из всех способов заставить испытуемого работать старательно является применение стандартных техник экспериментаторов для усиления внутренней мотивации задачи и вовлеченности в нее испытуемого.

Таблица 2. Опыт по освобождению от предубеждения

<i>Стратегии</i>	<i>Исследования, рассматривающие ретроспективное предубеждение</i>	<i>Исследования, рассматривающие чрезмерную уверенность</i>
<i>Задачи с искажениями</i>		
<i>Задачи с ложными условиями</i>		
Поднять ставки	4	1,30
Прояснить инструкции (стимулы)	6	3,10,13,14,21
Препятствовать ошибочному пониманию	11	13,21
Использовать лучшие способы ответов	9	13,14,20,22,23,32,34,35?,36,40
Задавать меньше вопросов	3,7,8,9	16
<i>Неправильно поняты задачи</i>		
Демонстрировать альтернативную цель	3,4,6,8,9	14
Демонстрировать семантическое несогласие	—	3,14,19,30
Демонстрировать невозможность задачи	—	13
Демонстрировать обозреваемое различие	—	15
<i>Ошибки экспертов</i>		
<i>Способные к совершенствованию</i>		
Предупредить о проблеме	—	13
Описать проблему	4	3
Обеспечить личную обратную связь	—	21
Экстенсивно тренировать	5?	1,2,4,17,21,26,27,31,34
<i>Не поддающиеся коррекции</i>		
Заменить их	—	—
Внести коррекцию в их ответы	—	2,5,24
Осуществлять проект на основе ошибок	—	—

Продолжение табл. 2.

Несоответствие между экспертами и задачей**Реструктурирование**

Сделать знания точными	—	18
Искать противоречивую информацию	9	18
Разложить проблему на составные части	6,11	—
Рассмотреть альтернативные ситуации	—	—
Предложить альтернативные формулировки	7,9	35?
Обучение		
Полагаться на надежных экспертов	1,2,7,8,10,11	11,16,20,24,29, 33,38,39/ 8,9, 23,28,31,32*
Обучать с детства	—	6,7

Примечания: *Ключ к исследованиям идет за примечаниями. Манипуляции, которые оказались, по крайней мере, частично успешными, отмечены жирным шрифтом. Те, которые все еще необходимо проверить эмпирически, или результаты которых неясны, помечены вопросительным знаком. Записи с индексом ^a перед разделом — это исследования с участием экспертов, не получивших тренировку по калибровке; записи после вертикальной черты — исследования, использующие различные уровни сложности.*

Ключ к исследованиям**Ретроспективные предубеждения**

1. Arkes, Wortmann, Saville, & Harkness (1981)
2. Detmer, Fryback, & Gassner (1978)
3. Fischhoff (1975)
4. Fischhoff (1977b)
5. Fischhoff (1980)
6. Fischhoff & Beyth (1975)
7. Mitchell & Kalb (in press)
8. Pennington, Rutter, McKenna, & Morley (1980)
9. Slovic & Fischhoff (1977)
10. Wohlstetter (1962)

11. G. Wood (1978)

Чрезмерная уверенность

1. Adams & Adams (1958)
2. Adams & Adams (1961)
3. Alpert & Raiffa (1969, 21)
4. Armelius (1979)
5. Becker & Greenberg (1978)
6. Beyth-Marom & Dekel (in press)
7. Cavanaugh & Borcovsky (1980)
8. Clarke (1960)
9. Coccozza & Steadman (1978)
10. Dawes (1980)
11. Dowie (1976)

12. Ferrel & McGoey (1980)
13. Fischhoff & Slovic (1980)
14. Fischhoff, Slovic, & Lichtenstein (1977)
15. Howell & Burnett (1978)
16. Hynes & Vanmarcke (1976)
17. King, Zechmeister, & Shaughnessy (in press)
18. Koriat, Lichtenstein & Fischhoff (1980)
19. Larson & Reenan (1979)
20. Lichtenstein & Fischhoff (1977)
21. Lichtenstein & Fischhoff (1980b)
22. Lichtenstein & Fischhoff, Phillips (Глава 22)
23. Ludke, Stauss, & Gustafson (1977)
24. Moore (1977)
25. Morris (1974)
26. Murphy & Winkler (1974)
27. Murphy & Winkler (1977a)
28. Nickerson & McGoldrik (1965)
29. Oskamp (1962)
30. Phillips & Wright (1977)
31. Pickhardt & Wallace (1974)
32. Pitz (1974)
33. Root (1962)
34. Schaefer & Borcharding (1973)
35. Seaver, von Winterfeldt, & Edwards (1978)
36. Selvidge (1980)
37. Sieber (1974)
38. Stael von Holstein (1971a)
39. Stael von Holstein (1972)
40. Tversky & Kahneman (1974)

Ошибочно понятые задачи. Как бы тщательно ни описывалась задача респондентам, все же останутся некоторые сомнения относительно того, действительно ли они поняли ее и приняли заданную структуру поощрения. Стандартный маневр для проверки того, “сработала” ли манипуляция – это посмотреть, останутся ли участники при своих ответах, которые они уже дали, если эти ответы используются в новой задаче со структурой поощрений, заданной для старой задачи.

Fischhoff, Slovic и Lichtenstein (1977) приняли эту стратегию, спрашивая людей, будут ли они готовы осуществить рискованное предприятие, основанное на оценке уверенности, которую они только что сделали. Это рискованное предприятие будет для них благоприятно, если оценка была искренней или имела тенденцию недооценить их уверенность, но оно покарает их, если они, из-за какой бы то ни было причины, преувеличили свои знания. Намеренное преувеличение может, например, служить альтернативной цели действовать более осмотрительно, чем это есть на самом деле. Такие испытуемые определенно стремились принять рискованное предприятие, несмотря на то, что были также чрезмерно уверены, как и испытуемые, наблюдавшиеся где-либо еще.

Еще одно основание для утверждения, что испытуемые поняли задачу иначе, нежели было задано экспериментатором, идет от наблюдения, что “степени определенности часто используются в повседневной речи (как упоминания о погоде), но они редко выражены численно, и часто нет возможности валидизировать их.... Неспособность людей оценить соответствующим образом 0,80 вероятности может быть не более удивительна, чем сложность,

с которой они могут столкнуться при оценивании яркости свечей или температуры в градусах по Фаренгейту” (Fischhoff и др., 1977, с. 553). Одним ответом относительно этой возможности является ограничение внимания к пределам шкалы вероятности в убеждении, что “100% уверенность, что утверждение правильно, охотно понимает большинство людей, и его уместность охотно оценивается” (Fischhoff и др., 1977, с. 553). Второй ответ — обеспечение вербальных ярлыков для количественных вероятностей, чтобы сделать их более легкими для понимания (например, Глава 21; Larson & Reenan, 1979). Ни одна из манипуляций не оказалась по настоящему эффективной. В утверждениях, что “неопределенность” сама по себе может иметь различные интерпретации, не все из которых являются значащими для всех индивидов, можно обнаружить более глубокое понятие семантического несогласия между экспериментатором и респондентом (Howell & Burnett, 1978; Phillips & Wright, 1977). Эмпирические попытки освобождения от предубеждения, основанные на этих концепциях, могут оказаться плодотворными.

Наиболее экстремальная чрезмерная уверенность наблюдалась в задачах, относительно которых респонденты вообще ничего не знали. Хотя экспериментаторы обычно пытаются не давать никаких намеков о том, насколько уверенными следует быть испытуемым, все же может иметь место подразумеваемое предположение, что “экспериментатор не дал бы мне задачи, которую невозможно выполнить”. Если испытуемые имеют подобные ожидания, тогда будет невозможным соответствующий уровень уверенности. Fischhoff и Slovic (1980) проверяли эту возможность в серии задач, чье содержание (например, диагностика язв, прогнозирование цен неизвестных акций) и инструкции были разработаны с целью сделать их кажущимися настолько невозможными, насколько это и было в действительности. Однако чрезмерная уверенность была уменьшена только (да и то частично) тогда, когда испытуемых предостерегли, что “может быть совершенно невозможно сделать подобного рода распознавание. Постарайтесь сделать все, что можете. Но если вы чувствуете, что совершенно не уверены (в своих ответах), без колебаний отвечайте 0,5 (показатель догадки) по каждому из них” (с. 752). Любые более сильные инструкции можно было бы подозревать в том, что они имеют характеристику требования.

Ошибки экспертов

Способные к совершенствованию. Если незначительно изменить интерпретационные положения, то последнее упоминавшееся исследование в предыдущем разделе может стать первым упоминающимся в настоящем. Уверение испытуемых в том, что они могли допустить, чтобы каждый ответ был просто догадкой, может рассматриваться как способ рассеивания какого бы то ни было остаточного ошибочного понимания задачи или как шаг в сторону исправления испытуемых, которые понимают задачи, но не себя. Оно несет подразумеваемое предостережение, что неспособность допустить до-

гадки может быть проблемой. Это предостережение было явным в инструкции (Глава 21) испытуемым “разворачивать цепочки” своих субъективных вероятностных распределений, чтобы избежать чрезмерной уверенности. Отражает ли частичный успех этих манипуляций улучшение понимания или чувствительность к распоряжениям, пока не ясно. Такая неясность может объяснить малое количество исследований, принявших эти подходы.

Это беспокойство о характеристиках требования исчезает в исследованиях намеренной тренировки, где “воздействия экспериментатора” составляют повестку дня. Как показано в табл. 2, различные попытки тренировки предпринимались с восхитительным успехом, хотя, может возникнуть беспокойство, что недостаточный энтузиазм журналов отражать исследования с негативными результатами мог уменьшить заметность неудач. Стремление тренеров делать все возможное для получения эффекта сделало попытки тренировки достаточно сложными манипуляциями, чьи эффективные элементы в некоторой степени неясны. Некоторыми из условий, изучение которых более необходимо являются: получение обратной связи по большой выборке ответов, получая информацию о конкретном выполнении задания (а не просто об общей проблеме), и наличие возможности обсуждения отношений между субъективным ощущением неопределенности и ответами о численной вероятности. К собственному удивлению, Лихтенштейн и Фишхофф (1980b) обнаружили, что один цикл тренировок с интенсивной, персонализированной обратной связью был так же эффективен, как долгая серия проб. Неясно, в какой степени эти различные успехи представляют тренировку, в узком смысле овладения конкретной задачей (например, изучение распределения ответов, требуемых экспериментатором), или приобретение более общих умений.

Не поддающиеся коррекции индивиды. Нетерпение в тренировочных исследованиях или скептицизм по поводу их всеобщности привели к тому, что целый ряд исследований приняли ошибочное оценивание уверенности как неизбежное и сконцентрировались на помощи принимающим решения справиться с ним. Одни предлагают заменить индивидов группами экспертов, чья оценка получена при непосредственном взаимодействии или схемой механического сбора (например, Becker & Greenberg, 1978; Morris, 1974); другие призывают к либеральному использованию анализов чувствительности всякий раз, когда в анализе решений поднимается вопрос об оценивании уверенности (например, Jennergren & Keeney, в печати); тогда как другие предлагают перекалибровать оценку, используя корректирующий фактор, который показывает должную степень уверенности оценивающего как функцию наличной степени уверенности (Lichtenstein & Fischhoff, 1977). Например, преобладание чрезмерной уверенности может наводить на мысль, что если объявлена определенность, то можно трактовать ее как 0,85 шанса быть правильным. К несчастью для этой стратегии, люди неправильно калибруют степень своей чрезмерной уверенности в зависимости от сложности конкретной задачи, с которой они столкнулись (Lichtenstein & Fischhoff, 1977). Как результат это-

го, необходимая калибровка может быть установлена, только если известна сложность наличной задачи, и есть возможность наблюдения (чрезмерной) уверенности респондентов в задаче подобной степени сложности, или, по крайней мере, возможность предположения отношений между наблюдаемой и предсказываемой чрезмерной уверенностью (Ferrel & McGoe, 1980).

Несоответствие между экспертами и задачей

Реструктурирование. Исследование калибровки, как и некоторые другие темы в отношении суждений, остались относительно изолированными от основного потока исследований сознания, извлекая из психологической литературы больше методологии, чем идей. Отражает ли этот недостаток контактов обособленность исследований суждений или неадекватное представление уверенности в распространенных моделях когнитивных процессов, он, вероятно, препятствует разработке методов снижения чрезмерной уверенности. Модели процессов должны как предлагать более сильные манипуляции, так и показывать, почему психофизиологические подходы работают или не работают, и насколько хорошо их результаты способствуют обобщению. Имеющиеся исследования показаний очевидцев, ощущения знания и мета-памяти могли бы при известных обстоятельствах обеспечить точки соприкосновения (например, Gruneberg, Morris & Sykes, 1978).

Одно из возможных направлений помощи людям в использовании их существующих когнитивных умений более совместимым образом с требованиями к оцениванию уверенности можно увидеть у Koriat, Lichtenstein & Fischhoff (1980), когда чрезмерная уверенность была уменьшена при перечислении респондентами причин, по которым выбранные ответы могут быть неправильными. Перечисление причин, почему респондент может быть прав, или называние одного основания “за” и одного — “против” выбранного респондентом ответа не дало никакого результата, показывая, что критический элемент — это не просто более усердная работа или большая точность, но и обращение к памяти иным образом, чем это обычно происходит в задачах оценивания уверенности. Без специфического побуждения “подумайте, почему вы можете быть не правы” люди оказываются недостаточно критичны или даже стремятся оправдать свой первоначальный ответ. Возможно, аналогичным образом Макмоен (1979) обнаружил, что 9-ти и 12-ти летние дети обнаруживали непоследовательность в текстовом материале только тогда, когда их просили ее отыскать.

Хотя и выдвинутая скорее на практической, чем на психологической основе, техника фиксированных значений Сивер и других (1978) может рассматриваться как еще один способ реструктурирования подхода респондента к задаче. Организация знаний человека вокруг набора значений, предположительно неправильных, может привести к более полной оценке того, что он знает, чем “традиционный” метод фиксированной вероятности, при котором внимание может фокусироваться на наилучшей догадке респондента о правильном ответе.

Обучение. Является ли исчезновение чрезмерной уверенности косвенным результатом реального обучения, которое получают эксперты в своей области? Как упоминалось ранее, очевидный способ исследования этого вопроса при рассмотрении выражений уверенности, сопровождающих выполнение реальной задачи, усложнен возможностью того, что реальное давление ограничит беспристрастие эксперта. Например, можно найти свидетельства чрезмерной уверенности в профессиях, где нужно выносить уверенные суждения, не демонстрируя валидность (например, предсказание изменений в биржевых ценах (Dreman, 1979; Slovic, 1972 c), психиатрическая диагностика опасности (Cocozza & Steadman, 1978)). Конечно же, если с такими экспертами консультируются (и платят им) в зависимости от уверенности, которую они внушают, они могут подвергаться искушению неправильно представлять свои реальные знания.

Несомненно, самые значительные попытки обеспечения беспристрастности предпринимались в случае с синоптиками, чья тренировка часто ясно поощряла их хорошей калибровкой. Выполнение ими задачи великолепно (например, Murphy & Winkler, 1974, 1977a). Относится ли этот успех за счет тренировки калибрования, или является продуктом их общего профессионального обучения, не ясно. Обзор других исследований, использующих экспертов, не получивших тренировок калибрования, наводит на мысль, что подобные тренировки, а не только реальное обучение, являются эффективным методом. Эксперименты, использовавшие проблемы, извлеченные из соответствующих областей экспертизы, но изолированные от давлений реального мира, обнаружили чрезмерную уверенность у выпускников психологического факультета (Lichtenstein & Fischhoff, 1977), банкиров (Stael von Holstein, 1972), клинических психологов (Oskamp, 1962), исполнительных лиц (Moore, 1977), гражданских инженеров (Humes & Vanmarcke, 1976) и нетренированных профессиональных синоптиков (Root, 1962; Stael von Holstein, 1971a).

Dowie (1976) обнаружил хорошее калибрование в прогнозировании у журналистов, ведущих колонку лошадиных бегах в газете. Хотя эти эксперты не получают ни поощрений, ни формальной обратной связи, можно предположить, что они получают удовлетворение, делая мониторинг своей работы изо дня в день, и вознаграждают себя хорошей калибровкой. Идею о том, что нас необходимо тренировать с детства для такого рода само-мониторинга, можно найти в недавних предложениях по введению в школьную программу вынесения суждений как предмета (например, Beyth-Marom & Dekel в печати; Cavanaugh & Borkovski, 1980). Перспективность таких предположений еще необходимо проверять.

И, наконец, существует довольно узкая форма экспертизы, оказавшаяся наиболее сильным (и наименее интересным) методом уменьшения чрезмерной уверенности. Один из примеров нечувствительности людей к объему своих знаний вырывается в том факте, что их средняя уверенность изменяется относительно медленно в ответ на изменения сложности задач, с ко-

торыми они сталкиваются (Lichtenstein & Fischhoff, 1977). Типичными парами пропорций правильных ответов и средней уверенности являются: 0,51, 0,65; 0,62, 0,74; 0,80, 0,78; 0,92, 0,86. Тогда как точность ранжируется в пределах 0,41, уверенность изменяется в пределах 0,23. Кривые калибрования, соответствующие этим итоговым статистическим данным, являются, в некотором смысле, одинаково плохими (или ровными); однако, их степень чрезмерной уверенности отличается значительно. Тогда как первые две пары представляют чрезмерную уверенность, третья показывает приемлемую общую уверенность, а четвертая — недостаточную уверенность. Эти примеры взяты у Lichtenstein и Fischhoff (1977), но подобная модель была обнаружена Clarke (1960), Nickerson и McGoldrick (1965), Pickhardt и Wallase (1974) и Pitz (1974), как и у многих других. В самом деле, любое сравнение чрезмерной уверенности в различных условиях должно принимать во внимание сложность используемых задач. В этом свете, превосходство чрезмерной уверенности в литературе отражает, в какой-то мере (возможно, естественную) тенденцию не ставить перед людьми слишком легких вопросов.

Резюме

Оценка уверенности, полученная у различных людей различными способами, почти всегда демонстрировала значительную нечувствительность к степени их знаний. Хотя все же не следует закрывать дверь перед методологическими манипуляциями, они оказались относительно неэффективными и их результаты трудно обобщать. То, что они сделали, это показали, что чрезмерная уверенность относительно сопротивляется многим формам исправления (отличным от изменений в уровне сложности). Ключом к разработке более сильных и предсказывающих манипуляций могла бы быть значительно большая опора на психологическую теорию. Эффективность тренировки калибровки наводит на мысль, что тщательный анализ того, какой уникальный опыт обеспечивается такими тренировками, а не профессиональным обучением, мог бы как руководить освобождением от влияния, так и обогатить психологическую теорию.

Обсуждение

Предположив, что рассмотренные здесь исследования были охарактеризованы точно, и что они исчерпывают (или, по крайней мере, достаточно представляют) совокупность соответствующих исследований, тогда их общее сообщение окажется довольно успокаивающим для когнитивного психолога. Оба предубеждения оказались умеренно устойчивыми, сопротивляясь попыткам интерпретировать их как артефакты и ликвидировать их “механическими” манипуляциями, такими как принуждение испытуемых работать усерднее. Эффективное освобождение от предубеждения обычно вовлекало изменение психологической природы задачи (и подхода испытуемых

к ней). В таких случаях, по крайней мере, некоторой чести заслуживает психологическая теория. Например, гипотеза о том, как люди восстанавливают информацию из памяти, прежде чем оценить уверенность, руководила манипуляцией Koriat и других (1980) этого восстановительного процесса. Даже тренировочная программа “забрасывания испытуемого всем” была основана на хорошо проверенных и общеприменимых принципах обучения.

Это изложение следует сопровождать несколькими концептуальными протестами (в дополнение к методологическим, которыми оно открылось). Одно состоит в том, что различие между артефактной и методологической манипуляциями может быть менее ясным, чем было сказано здесь. Например, увещание людей работать старательнее было бы артефактной манипуляцией, если бы оно коренилось в утверждении, что более непреднамеренные инструкции не вызывают “реального поведения”. Однако если бы исследователь мог выдвинуть основательную гипотезу о том, как различные инструкции воздействуют на процессы суждений, артефакт стал бы главным следствием с отдельными предсказаниями реального поведения в ситуациях с и без определенных увещаний.

Второй концептуальный протест состоит в том, что, когда мы подвергаем сомнению реальность предубеждений, это может наложить ограничения и сделать непродуктивным психологическое исследование. В продолжение примера из предыдущего абзаца, в жизни есть как несерьезные ситуации, так и ситуации, где следует работать усердно; внутренне ни одна из них не реальнее, чем другая. Кроме того, относительная валидность несерьезных и “усердных” ситуаций в лабораторных экспериментах зависит от ситуаций реального мира, на которые должны экстраполироваться их результаты. У каждой есть свое место. Понимание соответствия лаборатория-мир требует хорошего суждения для характеристики обоих контекстов. Например, ситуации с усердной работой не обязательно синонимичны важным ситуациям. Люди могут не работать усердно по важным проблемам до тех пор, пока не осознают как центральность суждений по результатам проблемы, так и потенциальную ошибочность этого суждения.

Использование исследований по освобождению от предубеждения для того, чтобы открыть пограничные условия наблюдения предубеждений, ведет к третьему концептуальному протесту. В этом обзоре итоговые таблицы косвенным образом приписали равный вес различным исследованиям, точно определенным, возможно, некоторым понятием окончательности каждого исследования (компетентностью, экстенсивностью и т.д.). Такое совпадение статистически значимых и статистически незначимых результатов является сомнительной процедурой, будучи построенной только лишь на методологической основе (например, Hedges & Olkin, 1980). Она становится концептуально сомнительной, когда начинаешь сомневаться относительно того, адекватно ли сделана выборка возможных исследований. В таких случаях собранные данные составляют концептуально зависимое наблюдение, и им не нужно присваивать равный вес. Любое изложение того,

как ведут себя люди, нуждается в тщательном определении подмира поведенческих ситуаций, по которым будет производиться выборка исследований. Например, некоторые критики обвиняли ранние исследования эвристики суждений в том, что они “искали неприятностей” в смысле поиска (выхватывания) ситуаций, в которых люди вели бы себя ошибочным образом. Если это утверждение правильно, тогда не нужно интерпретировать каждую демонстрацию поведения, подверженного предубеждению, как бойкотирование общей способности людей к суждениям; его уместность ограничена к такому роду ситуаций, какие изучались (или чрезмерно изучались) в тех экспериментах. Фокусируясь на пограничных условиях для оценивания предубеждений, большинство современных исследований подвержено собственному предубеждению выборки, которое необходимо рассмотреть в ходе обобщения их результатов.

Таблица 3. *Мир рассуждений о предубеждениях и попытках освобождения от предубеждений.*

-
1. *Лежащие в основе процессы, о которых требуется сделать выводы, являются вероятностными.* То есть, суждения выносятся в условиях неопределенности, с предубеждениями, вытекающими из конфронтации между детерминистским умом и вероятностным окружением.
 2. *Проблемы возникают в интеграции, а не в обнаружении свидетельств.* Хотя стимулы являются полными и недвусмысленными, насколько это возможно, они не много говорят о том, как может быть структурирована задача. Задача испытуемого состоит в интерпретации и использовании той информации, которая обеспечивается.
 3. *Предубеждения являются не связанными с существом проблемы.* Действие когнитивного процесса с данной информационной структурой должно быть одинаковым в любой содержательной области. Это ликвидирует “ошибки” вследствие неправильной информации и “неправильные концепции” вследствие намеренного обмана.
 4. *В наличии некоторая нормативная теория, характеризующая соответствующее суждение.* Этот критерий выводит проблему из области предпочтения (например, противоречивые отношения), где ни один ответ не может определяться как оптимальный.
 5. *Не предлагается и не позволяется никакая помощь в вычислениях (кроме карандаша и бумаги).* Этот фокус на интуитивных суждениях исключает такую помощь, как ручные калькуляторы, консультанты по статистике и интерактивные компьютеры.
 6. *Нет никаких очевидных побуждений к субоптимальному поведению.* То есть предубеждения когнитивны, а не мотивационны по природе. Суть исследования предубеждения состоит в том, конечно же, что там, где у людей нет весомой причины действовать субопти-

мально, ошибки наводят на мысль, что они просто не знают, как действовать лучше.

Дальнейшие вопросы

Вопрос, возникнут ли подобные модели по другим видам предубеждений, требует обзора аналогичной литературы. Таблица 3 предлагает характеристику сферы предубеждений, внутри которой могут разыскиваться повторяющиеся модели, отличая содержание этого издания от других предубеждений, причинявших беспокойство психологам.

Затягивающийся мета-вопрос, обращенный к таким обзорам, это: Насколько хороши люди? Являются ли они когнитивными калеками или знаками? Чтобы дать один ответ, необходимо ответить на невесомые вопросы о природе жизни и общем сходстве жизненного опыта людей с лабораторными условиями. Уклончивый момент настоящего обзора состоит в том, что резервуар умений по вынесению суждений у людей как наполовину пуст, так и наполовину полон. Люди достаточно умелы, чтобы прожить, достаточно неумелы, чтобы совершать предсказуемые и последовательные ошибки; они достаточно умны, чтобы изобрести широко и легко применимую эвристику, которая часто хорошо им служит, достаточно неискушенны, чтобы не осознать пределов такой эвристики. Более специфическая оценка способности людей может быть дана только в контексте конкретной задачи по вынесению суждения. Такие общие утверждения (или отговорки) о “людях” отражают общую черту исследований суждений — отсутствие интереса к индивидуальным различиям. Хотя это предпочтение групповых результатов может быть просто делом вкуса, оно может быть оправдано теоретически утверждением, что главные результаты в исследованиях суждений так обширны и используются так неадекватно, что индивидуальные различия могут подождать. Некоторую эмпирическую поддержку этого утверждения обеспечивает довольно скудное понимание, являющееся результатом исследования групп с известными характеристиками. Особенно поразительным было отсутствие различий в экспериментальных исследованиях наиболее логичных из известных групп — эксперты, выносящие суждения в своих областях экспертиз. Случайные и полученные в ходе изучения проблемы свидетельства, собранные Dawes (1976), Eddy (Глава 18, этого издания), Fisher (1970) и другими, также показывают, что экстенсивная тренировка и высокие ставки — не гарантия успехов в суждениях. Тем не менее, необходимо дальнейшее исследование, как из-за твердости, с которой многие уверены, что эксперты лучше, так и из-за прикладного значения использования экспертного суждения с большим преимуществом.

Для непосредственной практической цели наилучшего использования экспертов так, чтобы избежать предубеждения, достаточно знать, лучше ли они, чем непрофессионалы, или, по крайней мере, лучше ли осознают свои ограничения в суждениях. Для окончательной практической цели освобож-

дения всех судей от предубеждений очень важно знать, как эксперты оказались на том этапе, где оказались, или почему они не пошли дальше. Далее следует перечень условий, которые обычно сопровождают изучение. Для каждого показаны способы, при которых эксперты могут быть в особенно выгодном или невыгодном положении, в зависимости от обстоятельств:

1. Обильная практика с набором умеренно однородных задач. У экспертов должен быть такой опыт. Они могут использовать его, чтобы **оттачивать свои умения выносить суждения, или могут разрабатывать ситуативно-специфические привычные решения, освобождая себя от необходимости анализировать (и думать).**

2. События критерия точности. Хотя от эксперта часто требуют достаточно точных суждений, объекты таких суждений являются часто компонентами таких сложных (природных, социальных или биологических) систем, что трудно оценить уровень понимания судьи. Суждения, не попавшие в цель, могут появляться из-за непредвиденных случайностей, тогда как суждения, попавшие в цель, могут быть правильными согласно ошибочным доводам.

3. Специфичное для задачи подкрепление. Экспертам, в принципе, оплачивается их работа. Однако, даже когда видна мудрость их суждений, они могут получить поощрение на другой основе (например, принесли ли они хорошие известия? разрушили ли они планы? изменится ли все к лучшему?).

4. Точное принятие необходимости обучения. Становится учеником какой-либо программы, которая присуждает право экспертизы, – это, несомненно, признак скромности. Тем не менее, на каждой стадии этого процесса и профессиональной жизни, следующей за ним, определенные преимущества у тех, кто демонстрирует хорошие манеры и проявляет компетентность.

Существуют вполне оперантные принципы обучения, манипулирующие поведением без предположения каких бы то ни было знаний о лежащих в его основе когнитивных процессах. Прояснение и эксплуатация этих процессов является, очевидно, главной теоретической и практической задачей для исследований по освобождению от предубеждений, особенно когда подобные манипуляции рассматриваются как имеющие в какой-то мере лучший послужной список, чем более механические попытки. Хотя исследование предубеждений и освобождения от них преодолело достаточно большую часть пути от основного исследования до применения в полевых условиях, ему все же еще необходимо адекватно приблизиться к основам с обеих сторон. Уже сейчас очевидно, что достижение одного конца будет требовать достижения другого. Хорошая практика будет требовать лучшей теории о том, как работает мышление. Хорошая теория будет требовать лучшей практики, проясняющей и пытающейся выяснить условия, в которых мышление действительно работает.

32. Улучшение индуктивного вывода*

*Ричард Нисбетт, Дэвид Кранц, Кристофер
Дженсон и Джеффри Т. Фонг*

В этой главе мы обсудим возможность улучшения выводов, которые люди делают в повседневной жизни. Нисбетт и Росс (Nisbett и Ross, 1980) предположили, что улучшений можно достичь благодаря тому, что инструментарий ученого по формированию заключений становится доступным непрофессионалу в форме аксиом, таких как “Это эмпирический вопрос” или “Всегда можно объяснить исключения”. Мы обратимся к “статистическим эвристикам” за одним из наиболее важных неформальных руководств по рассуждению, такому как “Подумай об одном свидетельстве, как если бы это была выборка, и поразмышляй о размере этой выборки”. Мы уверены, что такие грубые руководства по выводу заключений, помещенные в матрицу понимания статистики и теории вероятности и поддерживаемые опытом применения эвристики в реальных проблемах, могут оказаться очень ценными в отношении уменьшения ошибок вывода в повседневной жизни.

Можно быть систематичным в исследовании того, что может произойти при использовании такой программы, и где вероятнее всего можно натолкнуться на сложности. Мы уверены, что сейчас видны три такие сложности, и мы уверены, что попытки решения каждой из этих проблем принесут дивиденды в смысле улучшения понимания как того, как люди делают выводы, так и того, как им следует это делать.

1. Может быть очень сложно установить, что данный сомнительный вывод на самом деле ошибочен. В большинстве работ Канемана и Тверски, вероятностные модели, являющиеся основой предписанных выводов, являются стандартными статистическими моделями, и их применение к рассматриваемым событиям неоспоримо. В то время как социально-психологическая ра-

* Написание этой статьи и некоторые из исследований, о которых в ней идет речь, были частично поддержаны грантом BNS 79-14094 Национального фонда науки. Мы благодарим Ли Росса, Сауля Штернберга и Поля Тагада за комментарии к более раннему варианту статьи.

бота в этом русле продолжается, становится все более ясно, что может быть очень сложно точно знать, какая модель для событий является правильной и, таким образом, знать, для чего требуются процедуры вывода.

2. Даже если можно быть достаточно уверенным, что произошла ошибка, очень не просто узнать, *как* она произошла. Трудно узнать, относится ли ошибка насчет ошибочного рассуждения, то есть, скудных процедур вывода, или насчет неправильных моделей, то есть, неверных предшествующих убеждений о природе рассматриваемых событий.

3. Даже если мы достаточно уверены в соответствующих моделях, у нас может не быть очевидных полезных руководств по выводу. В настоящее время мы не имеем представления о том, как перевести некоторые из наиболее фундаментальных статистических размышлений в руководство по рассуждениям в повседневной жизни. Это выступает особенно явно в примере размышлений о предубеждении выборки. Люди склонны уделять слишком мало внимания возможности того, что свидетельство подверглось предубеждению, но еще далеко до понимания того, какая “статистическая эвристика” будет соответствующей для использования в большинстве проблем реального мира.

Модели и эвристика в индуктивном рассуждении

Давайте начнем наше обсуждение ошибочного индуктивного рассуждения с анализа примера, в котором отсутствуют сложности, на которые мы только что ссылались. Многие из примеров в этой книге подошли бы, но проблема родильного отделения в работе Канемана и Тверски (1972, 3) особенно точна и поможет нам установить некоторые определения. Испытуемых попросили составить мнение о том, в большой или в маленькой больнице будет большее число дней в году, когда более 60% новорожденных будут мальчиками. Большинство испытуемых отметило “примерно одинаковое”, из оставшихся около половины отметили “в большой” и около половины – “в маленькой”. Другими словами, как группа, испытуемые были уверены, что подобные отклоняющиеся дни, когда рождение мальчиков превысит 60%, будут в равной степени вероятны как в большой, так и в маленькой больнице.

Канеман и Тверски предположили на основе этого и многих других экспериментов, что преобладающий способ рассуждения, ведущий к этому результату, — это использование *эвристики репрезентативности*. Испытуемые, использующие эту эвристику, будут фокусироваться на различии отклоняющегося результата (60%) от предполагаемого типичного исхода (около 50% для этого случая). Поскольку на степень сходства, или “репрезентативность”, не влияет размер больницы, испытуемые будут считать вероятность отклоняющегося результата одинаковой для обеих больниц.

Однако, при правильном подходе к этой проблеме, мы обращаемся к действительным подгруппам “мужских” и “женских” исходов в любой из дней в одной больнице, как к случайной выборке из совокупности 50 – 50. Затем

следует, из биномиальной формулы или закона больших чисел, что отклоняющийся процент выборки менее вероятен при большем размере выборки. заключаем, исходя из применения вероятностной модели, что дни, когда рождается 60 или более процентов мальчиков менее вероятны в большой больнице.

В этом примере рассуждение, основывающееся на *интуитивной эвристике* (эвристике репрезентативности), противопоставлено рассуждениям, осуществляющимся внутри математической модели. Мы выносим *нормативное* суждение, критикуя интуитивное суждение как ошибочное, поскольку было получено отличное от него заключение по модели, которая, как мы уверены, является подходящей репрезентацией для пола новорожденных. Наша процедура по вынесению нормативного суждения на основе модели иллюстрирует основной принцип: *Индуктивное размышление должно быть оправдано в терминах уместности основных моделей рассматриваемых событий.*

Математическая модель — лишь один вид модели, притом редко используемый в размышлениях людей: она требует математической тренировки и огромного количества времени для размышлений. Но мы будем использовать термин *модель* в отношении любой частичной *репрезентации* некоторых аспектов действительности. Мы включаем сюда физические модели (например, модель самолета, сделанная в масштабе), математические модели (например, уравнения, описывающие воздушную струю над крыльями самолета) и, более обобщенно, интуитивные концептуальные модели. Это есть умственные репрезентации людей о самолетах, или воздушной струе, или процессах, определяющих пол новорожденных, или о любых других аспектах действительности. Предельно важное различие среди моделей лежит между теми, которые являются чисто *детерминистскими* (то есть, они не содержат репрезентации непредсказуемости), и теми, которые являются *вероятностными* (то есть, они включают предположение, что события недостаточно предсказуемы, учитывая стандартные условия информации). Вероятностные модели могут быть физическими (выбрасывание игральные костей или вытягивание камушков из урны) или математическими (случайные переменные), но часто они менее точны. Например, человек, наблюдающий прыжки в длину может быть убежден, что следующий участник соревнования прыгнет где-то на 8 метров, но прыжок короче или длиннее не будет удивительным, и прыжок даже на 7,5 или 8,5 метров будет также возможным. Эта умственная репрезентация о прыжках в длину является интуитивной вероятностной моделью, включающей умственное “типичное расстояние” и, своего рода, “ошибочное распределение”, которое дает отклонения от типичного.

Природа модели событий имеет критическое значение для отбора используемых инструментов вывода, включающих разнообразную эвристику. *Эвристика* — это любой руководящий принцип трансформирования информации с целью решения проблемы или формирования суждения. Мы уже упоминали эвристику репрезентативности, но на самом деле она включает

два различных вида эвристики: один для построения моделей (согласно эвристике репрезентативности, основная модель должна близко соответствовать структурным особенностям наблюдаемых данных), а другой для оценивания вероятности исходов (исход более вероятен, если его структура более сходна со структурой предполагаемой основной модели). *Статистическая эвристика* поощряет думать об информации в терминах таких свойств данных, как надежность и валидность, и манипулировать информацией, используя вероятностные концепции, такие как базовое значение. Если имеется основная вероятностная модель для событий особого типа, то, вероятно, при рассуждении об этих событиях будет использоваться статистическая эвристика.

Давайте определимся, как применяются эти концепции моделей и эвристики в проблеме родильного отделения. Для полностью удовлетворительного решения этой проблемы использовалась бы математическая модель — модель случайной выборки из совокупности 50 – 50. Большинство испытуемых, возможно, используют слишком простую модель пола новорожденных — модель, которая определяет разделение 50–50, но не включает никаких отношений между случайностью процессов осуществления выборки и возможной ошибкой этого осуществления. Несомненно, большинство испытуемых знают что-то об отношениях между размером выборки и ошибочностью осуществления выборки, но это знание не включено в интуитивную модель, которую они умственно установили, чтобы решить проблему. Вместо этого они дополнили простую модель 50 – 50 репрезентативной эвристикой, которая привела их к заключению, что 60% рождений мальчиков в какой-то степени маловероятно, но в равной степени для обеих больниц. Заметим, что испытуемым не нужно быть статистиками, чтобы решить эту проблему, по крайней мере, качественным способом. Если их модель наблюдаемого отношения полов включала интуитивное понятие осуществления выборки, то они могли бы использовать статистическую эвристику. В этом случае, необходимой эвристикой будет понятие, что большие выборки с большей вероятностью будут иметь репрезентативную или типичную структуру, чем маленькие выборки.

Вкратце, размышление основано на моделях. В зависимости от своей модели человек может использовать различные алгоритмы эвристики. Математики могут выводить количественные следствия на основе математических моделей; люди с интуитивной моделью осуществления выборки могут успешно использовать статистическую эвристику; а люди со слишком простыми моделями могут использовать репрезентативную эвристику и, по крайней мере, в этой проблеме, будут вводиться ею в заблуждение.

Применение вероятностных моделей в повседневных выводах

Неудача использования вероятностной и статистической эвристики становится причиной ошибок не только в ответе на загадки типа проблемы ро-

дильного отделения; она также влияет на выводы людей в повседневных ситуациях, особенно в социальной сфере. Росс (Ross, 1977) предположил, что люди совершают *фундаментальную ошибку атрибуции*: они склонны приписывать поведение других людей личным наклонностям, не замечая ситуативных причин или временных внешних влияний на поведение. Мы дадим два примера этой ошибки и покажем, как рассмотрение вероятностных моделей помогает нам определять такой тип ошибок.

Нисбетт и Боргида (Nisbett и Borgida, 1975) показали, что испытуемые часто неспособны использовать информацию “консенсуса”: они подразумевают идиосинкразические, личные наклонности для поведения конкретного индивида, даже если их проинформировали, что большинство других людей в подобной ситуации вели себя сходным образом. Например, в одном исследовании испытуемые оценивали “Грега Р.” как равнодушного и жестокого человека, потому что он не пришел на помощь “жертве”, у которой, как он был уверен, был приступ. Тенденция давать “Грегу Р.” такие негативные оценки была высокой как у испытуемых, которым дали информацию консенсуса, а именно, что большинство людей в идентичной ситуации не оказали помощи жертве, так и у испытуемых контрольной группы, которым не давалась подобная информация консенсуса, и которые, таким образом, были уверены, что большинство людей помогли бы жертве.

Как люди могут выгодно использовать информацию консенсуса, чтобы смягчить свои выводы о личных чертах или склонностях? Нам кажется, что некоторые социальные психологи, которые все же делают акцент на ситуационных детерминантах поведения, оперируют при помощи интуитивной (а порой даже формальной) вероятностной модели, в которой некоторые ситуации ведут к высокой вероятности определенного поведения почти для всех индивидов, тогда как некоторые индивиды могут обладать высокой вероятностью определенного поведения почти во всех ситуациях. По структуре, предлагаемой подобной моделью, информация консенсуса легко интерпретируется как свидетельство того, что конкретная *ситуация* была высоко вероятной, и, следовательно, возникновение подобного поведения не является хорошим свидетельством того, что данный *индивид* обладает высокой вероятностью такого поведения в других ситуациях.

Мы считаем, что вероятностные модели такого типа являются хорошей репрезентацией социального поведения, и они обеспечивают основание для критики выводов, которые не использовали выгодно информацию консенсуса. Испытуемые в исследовании Нисбетта и Боргиды, возможно, полагались исключительно на репрезентативную эвристику, конструируя модель целевого человека, который будет соответствовать данным: Грег Р. жестокий и равнодушный, потому что его действия кажутся жестокими и равнодушными.

В этом примере, как и в случае с проблемой родильного отделения нет слишком большого соблазна защищать выводы испытуемых. Чистая модель черт социального поведения, *без какого бы то ни было* ситуационного влияния на вероятность, не является очень разумной, и сами испытуемые,

наиболее вероятно, допустили бы то же самое. Наш второй пример того, чего мы могли бы придерживаться, в качестве случая фундаментальной ошибки атрибуции, однако, ведет даже к большим спорам.

Предположим, что кандидату на значительную административную должность предложили встречу на основании очень хорошей рекомендации предыдущих работодателей. Встреча началась с группового ланча, во время которого кандидат демонстрировал нервное поведение. Позднее некоторые из присутствовавших на ланче людей сказали, что кандидату недостает необходимых для работы межличностных умений.

Социальный психолог мог бы сказать, что вывод работодателей о “межличностных умениях” сомнителен. Ситуативное объяснение нарушающего правила поведения (включая требования обстановки ланча-встречи) или вероятностная интерпретация (он вел себя не так, как обычно) могли бы быть правдоподобными, и, в самом деле, в свете рекомендаций предыдущих работодателей кажутся более правдоподобными, чем объяснение в терминах умений кандидата. Но работодатель мог бы ответить, что каждый, кто не в состоянии демонстрировать равновесие в ситуации ланча-встречи, тот вероятно, потерпит неудачу в других ситуациях, связанных с работой, которые в равной степени требуют равновесия.

Очень сложно опровергнуть этот довод. Работодатели оправдали свой вывод в терминах модели, которая постулирует высокую статистическую ассоциацию между неудачей в одной ситуации и неудачей в других ситуациях, сходных с первой по некоторым переменным, хотя по другим – нет. Трудно узнать, вовлекал ли действительный когнитивный механизм, лежащий в основе первоначального вывода, использование подобной модели или вместо этого просто полагался на эвристику репрезентативности (рассматривая сходство между поведением кандидата во время ланча и прототипным поведением во время ланча стереотипного идеального кандидата). И также трудно знать наверняка, что предложенная работодателями модель неверна. Социальный психолог знает по опыту, что подобная корреляция обычно очень слаба при эмпирической проверке; но эта конкретная корреляция не проверялась непосредственно, и чтобы проверить ее, применив существующие психологические техники, потребовалось бы невероятно подробное и недопустимо дорогое исследование.

Эти примеры иллюстрируют два различных способа, когда индуктивное размышление может быть ошибочным, и они указывают на два различных вида нормативных советов или обучения, которые могут потребоваться. (а) Иногда процедуры вывода ошибочны. Иногда люди могут соглашаться (по крайней мере, после размышления) с тем, какая модель приемлема, но на практике могут использовать другие, более простые модели и, соответственно, легкую эвристику. Нормативный совет делает акцент на избегании чрезмерного упрощения и слишком легкого использования эвристики репрезентативности. Цели обучения включают обучение моделям обширной полезности, таким как биномиальная модель осуществления выборки, обучение

конкретному применению этих моделей, и подчеркивание статистической эвристики, которая может иногда заменять строгое дедуктивное рассуждение в моделях. (б) Иногда рассуждение по данной модели правильно, но сама модель сомнительна или даже демонстративно неправильна. Нормативный совет подчеркивает, что известно (например, о слабости определенных типов поведенческих корреляций), что может делать конкретную модель предпочтительнее.

Этот анализ также указывает на два очень серьезных пробела в доступности нормативного совета. В некоторых случаях у нас могут отсутствовать знания, необходимые для формулирования адекватных моделей. В других случаях, у нас может не быть достаточно легкой и подходящей статистической эвристики. Образец информации консенсуса иллюстрирует пробелы в современных проблемах социальной науки. Мы можем критиковать испытуемых, которые не использовали информацию консенсуса вообще, но предполагаем, что испытуемые ответят: “Да, вы правы, ситуативные факторы являются релевантными; а теперь скажите нам, *насколько* мы должны исправить наше убеждение о равнодушии Грега Р. с точки зрения того факта, что большинство людей вели себя подобным образом?” Чтобы ответить на это, необходима была бы теория взаимодействия черт / ситуаций, которую специалисты по психологии личности и социальные психологи до сих пор не дали.

Демонстрация того факта, что у нас нет определенной существенной статистической эвристики, требует своего раздела.

Регулирование вывода для предубеждения выборки: Необходимость статистической эвристики

Правильные процедуры для учета *размера* выборки при выведении заключений хорошо разработаны, поскольку большая часть статистической теории посвящена проблемам такого рода. Частично, как следствие, мы не предвидим особых сложностей в обучении людей быть более чувствительными к рассмотрению размера выборки в задачах вывода, представленных в повседневной жизни. Однако когда мы рассматриваем вопросы *предубеждения* выборки, то дела обстоят несколько иначе. К сожалению, как доказали Нисбетт и Росс (Nisbett и Ross, 1980), ошибки вывода оказываются с гораздо большей вероятностью результатом предубеждения, чем небольшого размера выборки.

В обсуждении проблем, с которыми сталкиваешься, пытаешься разработать эвристику для преодоления предубеждения выборки, будет очень полезно различать два основных вида информации о выборке наблюдений: информация о *процедуре осуществления выборки* и информация о *типичности ковариантных черт*. В первом виде информации, а именно, о процедуре, мы будем довольно грубо различать случайные, статистические и иллюстративные выборки. Эти термины достаточно быстро выводят нас из сферы теоретической статистики, которая рассматривает только одну из катего-

рий, и толкают нас в сферу случайных наблюдений и взаимодействия наблюдаемых фактов.

Случайные выборки – это наблюдения, отобранные из всех возможных любым наиболее подходящим способом. Большинство выводов, не только в повседневной жизни, но даже в тщательной научной работе, основаны на случайных выборках. Например, испытуемые, использованные в исследованиях Канемана и Тверски, наблюдались потому, что были доступны. *Статистические* выборки используют известный вероятностный механизм для отбора наблюдений из некоторого множества возможных. Они являются единственным видом выборок, для которых возможны строгие статистические вычисления. Этот вид осуществления выборки преобладает в исследованиях путем опроса или в проверке контроля качества. Однако, умный статистик, составляющий модели, может перевести случайную выборку в статистическую путем формулирования разумной совокупности и вероятностной селективной модели, которая может быть удовлетворена действительной процедурой отбора. И, наконец, *иллюстративная* выборка является механизмом взаимодействия, используемым для увеличения конкретности или яркости отчета, основанного на значительно большем объеме данных. Телевизионные репортажи или журнальные статьи иногда иллюстрируют то, что рассматривается как характеристики совокупности, используя яркие единичные случаи. Но даже в науке большая презентация данных иллюстративна. В электрофизиологии, например, заключения, основанные на фрагментарном наблюдении множества нервных клеток, часто иллюстрируются отчетами по выборке клеток, которые хорошо демонстрируют типичные черты.

Другой тип информации о выборке относится к типичности ковариантных черт. Такая информация включает знания переменных, не являющихся главным интересом, которые могут быть коррелятами последних, и которые могут быть сравнены с известными значениями совокупности для оценки типичности выборки. Рассмотрим пример из повседневной жизни, где главная переменная – это доход, а возраст и пол каждого индивида также отмечаются. Предположим, мы хотим знать, какая группа зарабатывает больше – люди, пишущие стихи в качестве хобби, или люди, которые играют на музыкальных инструментах. У нас есть друг Джек, 50 лет, у которого много знакомых (преимущественно мужчины среднего возраста), пишущих стихи, а также у нас есть подруга Джейн, 25 лет, у которой много знакомых (преимущественно молодые женщины, играющие на музыкальных инструментах). Распределение доходов у поэтов намного выше, чем у музыкантов. Каждый согласится, что это ни о чем не говорит в отношении двух рассматриваемых совокупностей: выборки атипичны совокупностям по переменным, которые высоко коррелируют с доходом (в настоящее время). Это имеет большое значение в отношении того, *насколько атипичны* черты выборки, а также *насколько полезны* эти черты для предсказания значений центральной переменной. Во многих экспериментах Канемана и

Тверски были использованы в качестве испытуемых учащиеся израильской средней школы. И все же мы и они извлекаем намного более пространственные выводы из результатов, поскольку, хотя большинство взрослых людей не являются ни израильцами, ни учащимися средней школы, эти черты считаются слабо коррелирующими, если вообще коррелирующими, с центральными переменными в их исследованиях. Даже в сконструированной статистической выборке, следует отметить, должна рассматриваться атипичность: если в обзоре доходов зарегистрированных демократов была взята небольшая случайная выборка с преобладанием белых мужчин среднего возраста, то мы не примем стандартные типы выводов, основанные на случайном осуществлении выборки, как валидные.

Чтобы урегулировать атипичность, статистик, конструирующий модели, может попытаться более точно сформулировать отношения между целевой переменной и ковариантами и исправить вывод, используя эти отношения. Например, можно предположить или оценить отношения между возрастом, полом и доходами и сравнить урегулированные доходы поэтов и музыкантов. Такая адаптация требует опыта, размышления и часто количественных данных и умеренно больших выборок (для проверки модели и оценки параметров).

Мы отмечали, что случайное осуществление выборки или атипичные выборки иногда могут управляться сложными моделями, но неразумно ожидать от не экспертов управления предубеждением выборки этим способом. Какие интуитивные процессы используют обыватели, чтобы исправить предубеждение выборки? И какие рекомендации, исходящие из практических способов статистика, мы можем дать, чтобы улучшить их выводы?

К вопросу о том, как люди адаптируют предубеждение выборки, обращались в двух исследованиях Хемилл, Вилсон и Нисбетт (Hamill, Wilson и Nisbett, 1980). Общее заключение состояло в том, что люди адаптируют очень слабо: они не показывают никаких систематических различий между выводами, сделанными на основе случайных выборок, в сравнении с выводами, сделанными на основе типичных выборок, и даже извлекают подобного рода выводы на основе выборок, отмеченных как специфически атипичные (контр-иллюстративные, как это было).

Некоторые подробности этих исследований могут быть значимыми, если читателю необходимо серьезно подумать о том, какую статистическую эвристику следует использовать людям. В одном исследовании испытуемые просматривали видеозапись интервью с подразумеваемым тюремным надзирателем, который показался очень гуманным одной половине испытуемых и очень жестоким – другой. Некоторых испытуемых привели к убеждению, что представленный единичный случай был иллюстративным для большей выборки тюремных надзирателей. Другим испытуемым было сказано, что этот единичный случай был контр-иллюстративным: им ясно сказали, что интервьюируемый надзиратель был отобран как один из наиболее крайних (один из наиболее гуманных для тех, кто видел гуманного надзи-

рателя, и один из наименее гуманных для тех, кто видел жестокого надзирателя). Третьей группе не дали никакой информации об иллюстративности, и в результате была представлена случайная выборка единичного размера. В другом исследовании испытуемые читали журнальную статью с ярким негативным портретом получателя благотворительного пособия. Некоторых из испытуемых привели к убеждению, что случай, о котором они прочитали, был в высокой степени типичным в отношении продолжительности получения пособия. Им сказали (неправильно), что, как большинство получателей пособия среднего возраста, эта женщина получала пособие большую часть своей взрослой жизни. Другим испытуемым сказали правду об атипичности женщины в отношении этой черты: их проинформировали, что большинство получателей пособия получают его только в течение нескольких лет.

Отношения к тюремным надзирателям вообще (в первом исследовании) и к получателям пособия вообще (во втором исследовании) были оценены разнообразными мерами. Отношения к тюремным надзирателям были намного более благосклонными в группе, направленной на гуманного надзирателя, чем в группе, направленной на жестокого надзирателя. Это показывает, что вывод о надзирателях вообще был извлечен из единичного случая. Отношения к получателю пособия у испытуемых, которые читали журнальную статью, были значительно более негативными, чем отношения контрольной группы, снова показывая, что выводы о реципиентах пособия, в общем, были извлечены из статьи об одном получателе.

В исследовании надзирателей испытуемые, которым сказали, что надзиратель был типичным, из видеозаписи извлекли выводы, систематично не отличавшиеся от выводов, извлеченных испытуемыми, у которых не было информации о процедуре осуществления выборки. Даже испытуемые, убежденные, что надзиратель был атипичен, сделали выводы в том же направлении, что и другие группы, и в действительности различия между ними и другими группами не были статистически надежными. Подобным образом, в исследовании получателей пособия “типичная” и “атипичная” группы извлекли сходные выводы, и фактически наблюдавшееся различие между средними значениями их отношений было незначительным.

Очень важно для нашего последующего нормативного анализа поразмышлять о том, как можно объяснить эти результаты в терминах когнитивных процессов. Результаты довольно неожиданны, если предполагается, что у испытуемых имеются определенные предшествующие отношения, сформировавшиеся прошлым опытом и информацией, к получателям пособия и тюремным надзирателям. Сильное воздействие на такие предшествующие отношения о совокупности может быть, конечно же, приемлемо для испытуемых, которым был представлен иллюстративный или типичный случай, поскольку они достаточно убеждены, что он типичен для значительно большей группы; но это оправдание отсутствует для испытуемых “случайной” группы. И почему испытуемые, которым был представлен контр-иллюст-

ративный или атипичный случай, вообще должны делать какой-либо вывод по направлению к выборке, а не придерживаться своих первоначальных отношений или, возможно, даже продвинуться в какой-то мере в направлении, противоположном рассматриваемому случаю?

Современная теория о формировании отношения и изменении отношения предлагает несколько экспериментальных объяснений (Bem, 1967; Nisbett & Ross, 1980; Nisbett & Wilson, 1977). Мы можем рассуждать, что испытуемые, отвечая на вопросы об их отношениях, вообще не *восстанавливают* их из памяти. Скорее, они создают их из подручных материалов, включающих их текущую аффективную реакцию на объект, их семантические ассоциации с объектом и наблюдение их недавнего поведения по отношению к объекту. Большая часть процесса построения быстротечна и недоступна сознанию. Яркий единичный случай, вероятно, может вызвать аффективные реакции по отношению к целому классу объектов, который он представляет, несмотря на компенсирующее, но бледное уверение о типичности. (Аффективные реакции с большой вероятностью могут “разрушить когнитивную тюрьму”, как это и было. Никакое количество уверений об атипичности парижского таксиста, оскорбившего нас, не могут вернуть нас к нашему предыдущему ненастороженному отношению к классу парижских таксистов.) Яркий единичный случай может также служить побуждением к извлечению сходной информации из памяти. Хотя гуманные тюремные надзиратели могут напомнить испытуемым, совершенно произвольно, о добрых нацистских надзирателях в “Героях Хогана” или о милом южном перифе, изображенном Энди Гриффитом, в летних повторных показах фильмов их детства. Когда испытуемые приступают к созданию своих отношений о тюремных надзирателях вообще, они могут напомнить себе об игнорировании случайных или контриллюстративных свидетельств, которые они только что видели, и у них это даже может успешно получиться, но этого не происходит. Аффективные реакции и содержание памяти, подвергнутое теперешнему предубеждению, будут достаточными, чтобы выработать отношение, довольно отличающееся от отношения контрольных испытуемых.

Удерживая в уме все упомянутые понятия о том, как влияет на испытуемых наблюдаемый случай, давайте вернемся к нормативному вопросу: как *следует* людям регулировать свои выводы в свете знаний о процедуре осуществления выборки и типичности выборки с учетом всех ковариантных черт? Формальная теория и прикладная статистическая практика предлагают нам меньше, чем видит глаз, и значительно меньше, чем нам нужно.

Одно правило с неприятным и решительно научным вкусом утверждает, что все заключения о совокупности следует выводить лишь исключительно на основе должным образом случайно осуществленной выборки, иначе предположения о статистических процедурах нарушаются, и невозможно узнать, какую уверенность следует вкладывать в заключение. Согласно этому суровому правилу, испытуемые, не получившие никакой информации о процедуре осуществления выборки или типичности надзирателя в видеоза-

писи, должны были воздерживаться от всех выводов, и, таким образом, их отношение к тюремным надзирателям должно было быть одинаковым, независимо от того, видели ли они гуманного или жестокого надзирателя.

Это правило может быть отклонено на трех различных основаниях: гносеологическом, прагматическом и логическом.

1. Первый довод исходит из наблюдения Голдмана, что "...гносеологический совет или правила должны быть такими, чтобы им можно было следовать" (Goldman, 1978, с. 513). Наше обсуждение возможных механизмов создания отношений наводит на мысль, что "воздержание от выводов" может быть возможным правилом для повседневного сознания людей. Возможно, необходимо значительно более подробное правило, показывающее, как освободить от предубеждения содержание ассоциативной памяти и как компенсировать аффективные реакции.

2. Даже если мы можем применить в высшей степени запрещающее правило, это не окажет влияния на препятствование изучению большинства из того, что мы все-таки изучаем, как в повседневной жизни, так и в науке. Узнавая о деканах колледжа или об условиях движения на автомагистралях, человек стойко придерживается своей случайной выборки, на которую он натолкнулся лично или косвенно. И научные исследования выводов человека, цитируемые в этой статье, использовали испытуемых, материалы или обстановку, выборка которых осуществлялась не статистически, но была в наличии или легко создаваемой.

3. Наконец, случайность – это не непосредственно наблюдаемое или самоочевидное качество процесса осуществления выборки; это качество математических моделей, которые могут или не могут быть хорошими описаниями процессов реального мира. Случайная выборка может быть эффективно "случайной", потому что мы уверены, что определенные модели являются хорошим описанием мира и, как мы указывали, умный статистик, создающий модели, может облечь подобную интуицию в адекватную математическую формулировку.

Что мы могли бы посоветовать для родственной проблемы типичности ковариантных черт? Поверхностно, это могло бы казаться призывом формулировать критику против обобщения на основе выборки, которая существенно отличается от центральной популяции по чертам, являющимся возможно релевантными. Согласно этому правилу, испытуемые, которые знали, что выборка в случае, рассматривавшем пособия, была атипичной по одной релевантной черте (продолжительность получения пособия), должны были воздержаться от выводов обо всех других характеристиках этого случая вообще. Но это правило подвергается такой же критике, что и правило о случайности. Гносеологическая критика точно такая же: верное правило должно считаться с действительными когнитивными механизмами, используемыми в извлечении выводов в повседневной жизни, и должно быть таким, чтобы ему можно было следовать. Прагматическая критика подобна критике предыдущего правила. В большинстве научных исследований

можно найти некоторые черты выборки, которые атипичны, и которые возможно релевантны. Насколько можно быть убежденным в модели, которая утверждает, что такие черты являются в высокой степени релевантными? Наконец, даже если мы знаем или сильно подозреваем, что атипичная ковариантная черта важна, мы все же можем представить ее значение в модель и обеспечить соответствующую коррекцию. В примере о доходах музыкантов и поэтов-любителей мы могли бы постараться получить сырую оценку воздействия возраста и пола и сравнить адаптированные средние значения по двум группам. Конечно же, в этом примере было бы почти немыслимо сделать выводы с какой бы то ни было уверенностью, но это наверняка оттого, что у нас нет хорошей модели для воздействия возраста и пола на доход.

Возможно, существуют более мягкие или менее запрещающие принципы, которые могут быть использованы? Мы могли бы испробовать, например, принцип: “Знание является экспериментальным”. Он правилен, но кажется слишком неопределенным, чтобы широко использоваться. Нам необходима более подробная статистическая эвристика. *Насколько большую уверенность мы должны вкладывать в обобщения, извлеченные из случайных выборок? Какие экспериментальные выводы можно сделать на основе атипичных выборок?* Частично эти вопросы обращены к статистикам и философам. Существует ли общий подход к построению модели, которая может быть источником статистической эвристики для индуктивного вывода? Но наше обсуждение механизмов формирования отношений и бессознательных процессов, влияющих на вывод, должно прояснить, что, как Голдман (Goldman, 1978) доказывал, гносеологические принципы должны проходить проверку способности быть используемыми, равно как и валидности. И здесь мы сталкиваемся с еще одним вопросом. Часть процесса вывода, который мы желаем модифицировать, является автоматической и бессознательной (Bem & McConnell, 1970; Goethals & Reckman, 1973; Nisbett & Wilson, 1977). Какой ценой можно исключить автоматичность? Понимание инструкций, рассказов, мотивов и т. д., требование постоянного потока выводов. Разумно подозревать, что автоматичность может быть обнаружена в этих функциях в приблизительно такой же степени, как и в восприятии. Хорошая статистическая эвристика должна быть постижимой в отношении того, где ее можно легко использовать, даже автоматически.

Могут ли вероятностные модели использоваться в повседневной жизни?

Мы рассмотрим программу, в которой вероятностные модели и статистическая эвристика будут окончательно введены в повседневное размышление большинства людей. Мы обсуждали две основные сложности такой программы: игнорирование того, какие модели правильны, и отсутствие удобной статистической эвристики в некоторых важных областях. Третий, в такой же мере сложный вопрос, который, конечно же, логически является

предшествующим, касается способности человека вводить вероятностные модели и эвристику в повседневное мышление. Несмотря на трудности, есть основания для оптимизма, и некоторые вопросы могут поддаваться экспериментальному исследованию вида, который мы будем описывать позже.

Одним из поводов для оптимизма является то, что размышление человека изменяется с возникновением новых культурных изобретений. Подлинные продвижения вывода произошли в пределах последнего исторического времени. Современному понятию вероятности едва ли больше 300 лет (Hacking, 1975). И еще до 1660 года понятия, которые были в *любом* смысле вероятностными, применялись почти исключительно к пониманию результатов, генерированных механизмами случая, такими как игральные кости и карты. Однако фактически каждый образованный человек сегодня часто использует статистическое рассуждение в некоторых сферах, например, спорт и погода, и имеет статистическое понимание механизмов случая, таких как кости и карты, которое очень отличается от ранних концепций (Хеккинг заходит так далеко, что говорит, что играющий в кости в древние времена на основе современных концепций вероятности завладел бы всей Галлией в быстром порядке!)

Кроме чисто статистических понятий, довольно много общих индуктивных принципов и руководств появилось относительно недавно. Понятие о том, что корреляция не является достаточной для установления причинности, не появляется до Хьюма (Hume) и не получает точного общего утверждения до *Учебника политических ошибок* Бентама (Bentham, 1984/1952). Общий обвинительный акт “критерия сходства” (Nisbett & Ross, 1980), как основа для выведения заключения о причинно-следственных отношениях, не появляется до *Системы логики* Милла (Mill, 1843/1974). В самом деле, до конца 18 столетия совершенно противоположному правилу следовали врачи, которых учили рассуждать в соответствии с “доктриной подписей”. Согласно этой доктрине, от каждого естественного лечебного агента можно ожидать, что он по хорошо заметному внешнему свойству определит болезнь, против которой он эффективен. Поэтому тяжелые, подобные гравию объекты, были полезны в лечении желчекаменной болезни, желтые объекты были полезны в борьбе с желтухой, и т.д. К счастью, врачи оказались способными отказаться от этой индуктивной эвристики в пользу других продуктивных эвристик.

Второй повод для оптимизма состоит в том, что хорошее размышление может иногда быть полезнее и легче, чем ошибочное. Мы сказали уже, что ошибки идут от чрезмерного упрощения, и что точные модели могут быть сложными. Но обратное в той же мере может быть правдой. Ошибки могут проистекать от чрезмерно сложного причинного рассуждения, и отсутствие хорошего простого подхода к проблеме может быть причиной того, что люди генерируют несколько сложных плохих подходов. В одном из наших текущих экспериментов мы проводим телефонные интервью о спорте, в которые мы внедриli некоторые вопросы, которые поддаются вероятностным

ответам. Часто мы получаем краткие, неожиданные (и на наш взгляд, правильные) вероятностные ответы. Другие интервьюируемые, которые не думают о вероятностном подходе, нерешительно предлагают несколько длинных детерминистских объяснений, ни одно из которых, кажется, не убеждает даже самих интервьюируемых, что они на правильном пути.

Как исторические, так и эпизодические свидетельства отражают один и тот же основной смысл. Даже быстрые и автоматические процессы вывода применяются в хорошо изученных концепциях и моделях. Рассуждения человека и предпочитаемые формы доводов, следовательно, изменяются не только как функция созревания индивида, но также и как функция изменений в языке, культуре и образовании.

Недавно мы начали программу исследования, чтобы рассмотреть, как люди рассуждают, и как они учатся рассуждать, о проблемах, к которым следует, по нашему мнению, подходить с вероятностным уклоном. Мы использовали проблемы с большим количеством различных структур и с содержанием, которое имеет тенденцию, в большей или меньшей степени, вызывать вероятностное мышление. Это не место для попыток сообщить даже предварительные результаты, но может быть полезным проиллюстрировать материалы, которые мы используем, чтобы указать вопросы, которые, мы считаем, являются достойными для того, чтобы их задали. Чтобы проиллюстрировать материалы, мы предлагаем следующую проблему.

Выбор колледжа

Дэвид Л. был старшеклассником в школе на Восточном побережье, который планировал поступить в колледж. Он имел отличные оценки в школе, и у него была возможность выбора из двух наилучших альтернатив: небольшой колледж свободных искусств и Ivy League University. Оба заведения имели почти равный престиж и равную оплату за обучение. Оба находились в привлекательных городах Восточного побережья на приблизительно равном расстоянии от его родного города. У Дэвида было несколько старших друзей, посещающих колледж свободных искусств, и несколько друзей, посещающих Ivy League University. Все они были отличными студентами, как и он, и имели сходные с ним интересы. Все его друзья из колледжа свободных искусств сообщили, что им нравится это место, и что они находят его очень стимулирующим. Друзья из Ivy League University сообщили, что у них много жалоб как на личной и социальной основе, так и на образовательной основе. Дэвид первоначально думал, что он пойдет в колледж свободных искусств. Однако он решил сделать однодневные визиты в оба заведения лично. Ему не понравилось то, что он увидел в частном колледже свободных искусств: несколько людей, которых он встретил, показались ему холодными и неприятными; профессор, с которым он имел краткую встречу, казался резким и незаинтересованным в нем; и ему не понравилось "ощущение" университетского городка. Ему понравилось то, что он увидел в Ivy League University: несколько людей, которых он встретил, показались ему живыми, энтузиастами, приятными в общении; он встретился с двумя различными профессорами, которые проявили к нему личный интерес; и он ушел с приятным ощущением университетского городка.

Какое заведение выберет Дэвид Л. и почему? Постарайтесь проанализировать доводы с обеих сторон и объяснить, какая сторона сильнее.

Это особенно трудная проблема, потому что она содержит два различных вероятностных компонента: (а) довод основного значения, относящийся к маловероятному результату, что Дэвид Л. будет действовать на основе хвоста распределения (на основе текущих реакций своих друзей он может предположить, что модальный ответ по колледжу свободных искусств очень благоприятен, и что вариативность низка), и (б) довод размера/предубеждения выборки, касающийся адекватности его однодневного разоблачения. Мы предполагаем, на основе опытного исследования, что каждый из этих вероятностных компонентов блокирован обычным набором убеждений. А именно, что профиль предпочтений и антипатий индивида уникален и предсказуем только на основе опыта и личности этого индивида. Ощущения индивида, насколько таинственными они ни были бы в своем происхождении, являются, по крайней мере, надежными, с небольшим количеством информации о том, является ли объект достаточным для создания списка плюсов и минусов, который обеспечивает уверенное руководство для будущих реакций. Большинство незаконченных ответов по этой проблеме отражали эти убеждения (например, “он должен решать сам, а не его друзья”). Но не все ответы таковы. Вот один ответ выпускника, у которого не было курсовых работ по вероятности или статистике:

Я бы сказал, что он пойдет в колледж свободных искусств. Его негативный опыт был лишь кратким, очень поверхностным контактом с заведением. Его друзья, все истинные клоны его самого, были там (предположительно) некоторое время и знают место близко, и оно им нравится, тогда как для Университета правдивы обратные утверждения. Однако он будет оправдан, если будет иметь собственные впечатления об этих местах. Часто эта интуиция является высшим восприятием, которое мы не можем анализировать, и он будет прав, следуя ему. Хотя, я думаю, что первый упомянутый мной выбор более надежен, так как его опыт знакомства с двумя заведениями *очень* ограничен.

Мы считаем этот ответ полностью удовлетворительным статистическим ответом, хотя, конечно же, он не высказан на языке вероятностных моделей.

Мы начали эти исследования с подозрением, что в большинстве случаев, где формальная вероятностная модель может быть с пользой применена статистиком, существуют аналоги в повседневном мире, в которых подобное интуитивное использование вероятностного мышления часто происходит у сообразительных непрофессионалов. Например, мы отмечали раньше, что люди не могут формулировать модель, в которой поведение индивида рассматривается как функция и ситуативных переменных, и диспозиционных переменных, и в которой некоторые ситуации вызывают высокую вероятность поведения у большинства индивидов. Все же мы думаем, что люди все-таки оперируют чем-то похожим на этот вид моделей, когда поведение – это хорошее выполнение теста. Концепция “легкого теста”, когда большинство людей имеет высокую вероятность хорошего выполнения, широко известна и используется в нашей культуре. Мы подозревали, что другие виды вероятностных моделей, для других видов проблем, будут также применяться в повседневных выводах. Наши исследования пытались продемонстри-

ровать этот вопрос, показывая, что люди могут хорошо справляться с определенными проблемами, в то время как терпят неудачу в других с подобной формальной структурой, потому что в их моделях, лежащих в основе последних проблем, отсутствует вероятностный компонент.

Если подозрение подтверждается, это приводит нас к ряду вопросов, на которые мы можем постараться ответить экспериментально. В какой степени хорошее вероятностное размышление обязано применению абстрактной или формальной модели определенного типа? Почему формальная модель должна быть легко применимой в конкретных задачах определенного типа и редко в других? Если успешное размышление не является применением бессодержательной формальной модели, тогда как оно должно быть описано в теоретических терминах? Другие вопросы касаются индивидуальных различий в моделировании и использовании правил. В какой степени использование вероятностного рассуждения является устойчивой личной диспозицией? Коррелирует с интеллектом? Связана с полученным образованием?

Наконец, имеется набор важных вопросов, касающихся проблем возможности обучения созданию вероятностных моделей. Основной вопрос уже получил ответ в нашем исследовании: Людей можно обучить как курсами традиционной статистики, так и более быстрыми методами, чтобы отвечать, по крайней мере, по многим “словесным проблемам”, типа тех, что были упомянуты нами, более вероятностным образом. В какой степени поддерживаются действительные суждения повседневной жизни, обучающие техники которых наиболее эффективны, и каковы наилучшие индуктивные принципы и методы обучения – вот важные вопросы, которые мы только начали задавать. Мы надеемся, что получим помощь.

Часть IX
Восприятие риска

33. Факты против страха: понимание воспринимаемого риска *

*Пауль Словик, Барух Фишхофф
и Сара Лихтенштейн*

Люди реагируют на риск, который они воспринимают. Если их восприятие ошибочно, попытки личной, общественной и средовой защиты, вероятно, будут неправильно направлены. Для некоторых видов риска, например, дорожно-транспортных происшествий, имеются в наличии экстенсивные статистические данные, для других подобных деятельности, таких как потребление алкоголя и табака, оценка риска требует сложных эпидемиологических и экспериментальных исследований. Однако, даже если статистические данные в избытке, только “неопровержимые” факты могут вести так далеко, чтобы разрабатывать способ действий. В каком-то смысле, суждение человека необходимо для интерпретации результатов и определения их релевантности.

Однако, другие виды риска, такие как исследование модифицированного ДНК или атомная энергия, так новы, что оценка риска, скорее, должна быть основана на сложном теоретическом анализе, таком как дерево ошибок (см. рис.1), чем на непосредственном опыте. Несмотря на появление объективности, этот анализ также включает большой компонент суждения. Полагаясь на тренированную интуицию, необходимо определить структуру проблемы, последствия, с которыми надо считаться и важность различных ветвей дерева ошибок. Как только анализ будет представлен, с ним должны взаимодействовать те, кто в действительности имеет дело с риском, включая промышленников, специалистов по окружающей среде, регуляторов, законодателей и избирателей. Если эти люди не понимают или не верят данным, которые им демонстрируют, то тогда возможны недоверие, конфликт и неэффективное управление риском.

* Это пересмотренная версия статьи, которая первоначально появилась в книге под редакцией R. Schwing и W. A. Albers Jr. *Оценка риска общества: Насколько безопасность достаточно безопасна?* New York: Plenum Press, 1980. Авторские права © 1980 имеет Plenum Press. Печатается в соответствии с разрешением. Поддержка этой работы была обеспечена Программой оценки технологий и анализа риска фонда Национальной науки по Гранту PRA79-11934 Университету Кларка согласно субдоговору с Perceptronics, Inc.

ло ряд общих правил вывода, которые люди, по-видимому, используют в подобных ситуациях. Эти правила суждения, известные как эвристики, используются для упрощения сложных умственных задач до простых. Являясь валидными при одних обстоятельствах, при других они приводят к сильным и устойчивым предубеждениям которые серьезно вовлечены в принятие решений в таких различных областях, как финансовый анализ (Slovic, 1972) и управление природным риском (Slovic, Kunreuther & White, 1974).

Доступность

Одна из эвристик, которая имеет особую релевантность для восприятия риска, называется доступность (Tversky & Kahneman, 1973, 11). Люди, использующие эту эвристику, судят о событии как о вероятном или частом, если его примеры можно легко представить или вспомнить. Так как часто происходящие события обычно легче представить или вспомнить, чем редко происходящие события, доступность часто является подходящим признаком. Однако на доступность оказывают влияние многочисленные факторы, не связанные с частотой происхождения события. Например, недавнее бедствие или яркий фильм, такой как “Челюсти” или “Китайский синдром”, могли бы серьезно исказить суждения о риске.

Предубеждение доступности помогает объяснить неправильное восприятие людей и ошибочные решения в отношении определенных естественных видов риска. Например, обсуждая жителей затопляемых территорий, Кейтс (Kates, 1962) писал:

Главное ограничение способности людей использовать улучшенную информацию о риске наводнения – это то, что в основном они полагаются на опыт. Люди в затопляемых территориях оказываются в большой степени пленниками своего опыта... Недавние наводнения, кажется, устанавливают стремящиеся вверх границы размеров потерь, о которых, по убеждениям менеджеров, люди могли бы побеспокоиться заранее. (с. 140)

Кейтс приписывает сложность улучшения контроля наводнения, в основном, “неспособности индивидов концептуализировать наводнения, которые никогда не происходили” (Kates, 1962, с. 92). Он наблюдал, что индивиды, прогнозирующие потенциал наводнения, “находятся под сильным влиянием своего недавнего прошлого и ограничивают свою экстраполяцию к упрощенным конструкциям, рассматривая будущее как отражение своего прошлого” (с. 88). Подобным образом, заключение страховых договоров от землетрясения резко увеличивается после землетрясения, а затем постоянно снижается с угасанием воспоминаний (Steinbrugge, McClure & Snow, 1969).

Особенно важный результат эвристики доступности состоит в том, что обсуждение низковероятного риска может увеличить его запоминаемость и способность его представить и, таким образом, его кажущуюся рискованность, независимо от того, что показывают свидетельства. Например, лидеры в области исследований модифицированного ДНК быстро раскаялись в

представлении общественному вниманию слабого риска “загрязнения” новыми организмами. Розенберг (Rosenberg, 1978) подытожил реакцию, которая последовала обнаружению предположительного риска:

Первоначально ответом была похвала... социальной ответственности, проявленной вовлеченными в это дело учеными.... Однако, постепенно и предсказуемо дебаты стали накаленными. Размышления были в изобилии, и чем страшнее был сценарий, тем шире была огласка, которую он получал. Многие из дискуссий по этому вопросу полностью потеряли из вида тот факт, что опасность была, прежде всего, гипотетической, и предположили, что лаборатории, где проводились исследования по модификации ДНК, были полны бешеных животных. В конечном счете, сами ученые, чья сдержанность привела в движение весь процесс, были очернены. (с. 29)

Мнение о частоте смертельных событий. Предубеждение доступности проиллюстрировано несколькими исследованиями, в которых студенты колледжа и члены Лиги женщин-избирателей судили о частоте 41 причины смерти (Lichtenstein и другие, 1978). В одном исследовании этим людям сначала сообщили о годовой смертности по одной причине (автомобильные аварии) в США (50. 000), а затем попросили оценить частоту остальных 40 причин. В другом исследовании участников попросили вынести суждение, какая из двух причин смерти в паре была более частой. В обоих исследованиях суждения были умеренно точны в глобальном смысле: люди обычно знали, какие события были наиболее и наименее частыми. Однако внутри этой глобальной картины люди совершали серьезные ошибки в суждениях, многие из которых оказывались отражением предубеждения доступности.

Рисунок 2 сравнивает количество смертей в год согласно суждениям испытуемых с количеством смертей по данным статистики общественного здоровья. Если суждения о частоте были точными, они будут равны статистическим нормам, всеми точками находясь на линии идентификации. Хотя более вероятные виды риска обычно вызывают более высокие оценки, точки оказываются разбросанными по кривой линии, которая пролегает иногда выше, а иногда ниже линии точных суждений. В целом, редкие причины смерти были переоценены, а распространенные причины смерти были недооценены.

Кроме этого общего предубеждения, было очевидным специфическое предубеждение по размерам. Например, было вынесено суждение, что несчастные случаи являются причиной такого же количества смертей, что и болезни, тогда как болезни в действительности уносят в 16 раз больше жизней. Убийства неправильно рассматривались как более частые, чем смерти от диабета и рака желудка. Они также рассматривались такими же частыми, как и смерти от удара, хотя последние в действительности уносят в 11 раз больше жизней. Частота смертей от ботулизма, торнадо и беременности (включая роды и аборты) была также значительно переоценена. В таблице 1 перечислены смертельные события, суждения о частоте которых в различ-

ных наших исследованиях были наиболее неправильными. В продолжение рассмотрения доступности, переоцененные причины смерти были драматическими и сенсационными, тогда как недооцененные причины скорее были незрелищными событиями, которые вызывают одну жертву и являются привычными в своей несмертельной форме.

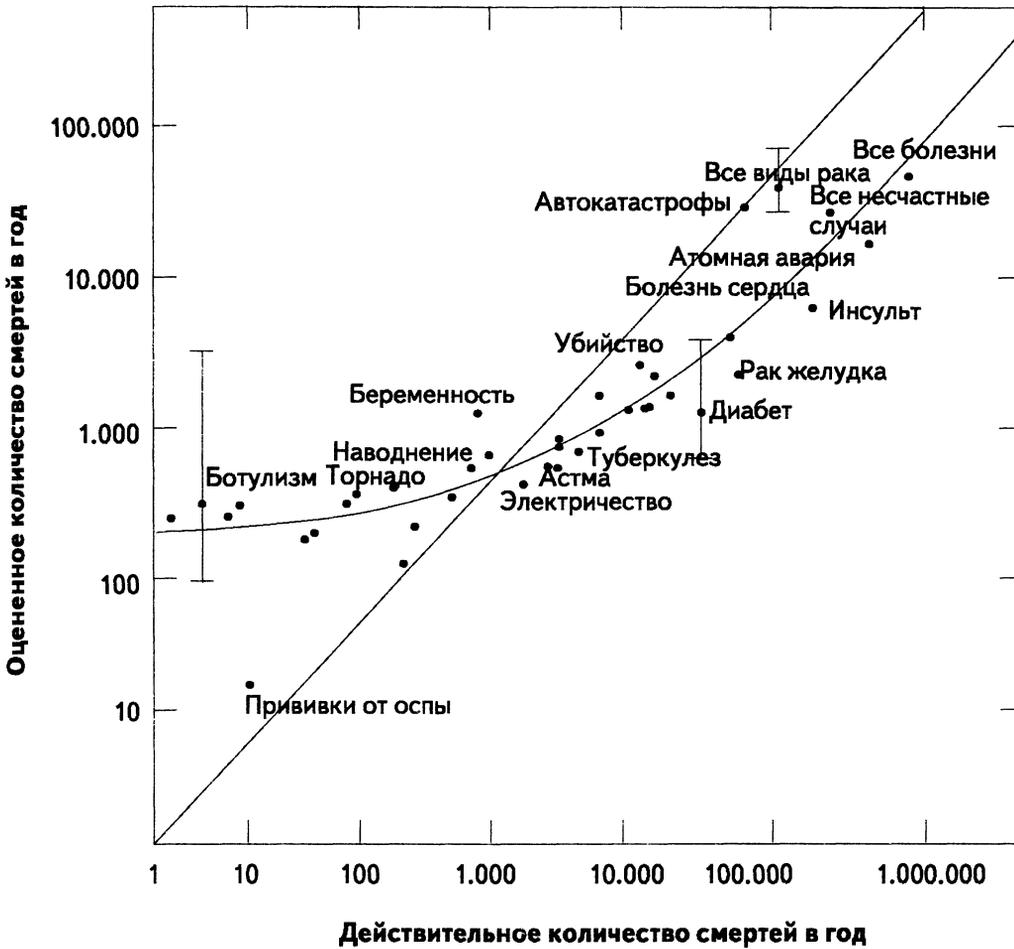


Рис. 2. Отношения между оцененной частотой и действительным количеством смертей в год по 41 причине. Если оцененные частоты и действительные частоты были равны, данные попали бы на прямую линию. Точки и кривая, им соответствующая, представляют средние ответы огромного количества непрофессионалов. Как индекс изменчивости среди различных индивидов, прочерчены вертикальные полосы, изображающие 25-й и 75-й процентиля суждений о ботулизме, диабете и всех несчастных случаях. Область распределения ответов по другим 37 причинам была сходна

Таблица 1. *Предубеждение о частоте смертей в суждениях*

<i>Наиболее переоцененные</i>	<i>Наиболее недооцененные</i>
Все несчастные случаи	Прививка от оспы
Автомобильные аварии	Диабет
Беременность, роды, аборт	Рак желудка
Торнадо	Молния
Наводнение	Инсульт
Ботулизм	Туберкулез
Все виды рака	Астма
Пожары и возгорания	Эмфизема
Ядовитый укус животного или насекомого	
Убийства	

Источник: Slovic, Fischhoff & Lichtenstein, 1976.

Подверженные предубеждению газетные репортажи и подверженные предубеждению суждения. Эвристика доступности выдвигает на первый план жизненно важную роль опыта как детерминанты воспринимаемого риска. Если опыт человека подвержен предубеждению, то его восприятие, вероятно, будет неточным. К сожалению, большинство информации, под воздействием которой находятся люди, обеспечивает искаженную картину мира рисков. Рассмотрим наблюдение автором Ричардом Бахом страха, высказанного супружеской парой, впервые летящей на самолете:

При всем этом ветре и гудении двигателя, и земле, кренящейся и становящейся все меньше под нами, я наблюдал за моими парнем и его девушкой из Висконсина и видел, как они меняются. Несмотря на их смех, они боялись самолета. Все, что они знали о полетах, было почерпнуто из газетных заголовков, сообщения о столкновениях, авариях и смертельных случаях. Они не читали ни единого сообщения о том, как маленький самолет взлетел, пролетел по воздуху и благополучно приземлился. Они могли только верить, что это возможно, несмотря на газетные статьи, и на эту веру они поставили три доллара и свои жизни (Vach, 1973, с. 37).

Как продолжение исследований, о которых сообщалось выше, Кобс и Словик (Combs и Slovic, 1979) рассмотрели сообщения о причинах смерти в двух газетах на противоположных побережьях США. Записывались различные индексы газетных репортажей по чередующимся месяцам в течение одного года. Результаты выявили, что обе газеты имеют сходное предубеждение в своих репортажах о событиях, угрожающих жизни. Например, рассмотрение Таблице 2 обнаруживает, что о многих из статистически частых причин смерти (например, диабет, эмфизема, различные формы рака) редко сообщалось в обеих газетах в течение периода исследования. Кроме того, о жестоких, часто катастрофических событиях, таких как торнадо, пожа-

ры, случаи утопания, убийства, автомобильные аварии и все несчастные случаи, сообщалось чаще, чем о менее драматических причинах смерти, имеющих подобную (или даже большую) статистическую частоту. Например, болезни уносят в 16 раз больше жизней, чем несчастные случаи, но в газетах было в 3 раза больше статей о несчастных случаях, отмечавших в 7 раз большее количество смертей. Среди наиболее частых событий, убийства были категорией, о которой сообщалось больше всего в пропорции к действительной частоте. Хотя болезни уносят в 100 раз больше жизней, чем убийства, в газетах было в 3 раза больше статей об убийствах, чем о смертях от болезней. Более того, статьи об убийствах имели тенденцию быть в 2 раза длиннее статей, сообщающих о смертях от болезней и в результате несчастных случаев.

Более того, предубеждения в газетных репортажах и суждениях людей были достаточно сходными. Корреляция между оцененной частотой смерти и количеством смертей, сообщаемым в газетах, была около 0.70. Эта высокая корреляция была не за счет общей ассоциации как тех, так и других смертей со статистической частотой. Когда последняя удерживалась постоянной, частичные корреляции между суждениями людей и количеством сообщаемых смертей были 0.89 и 0.85 для этих двух газет. Хотя на основе этих двух корреляций заманчиво сделать заключение, что репортажи в средствах массовой информации оказывают влияние на восприятие риска, дело может быть также в том, что мнения людей относительно того, что важно, влияют на СМИ. Журналистская литература насыщена примерами, в которых влияние происходит в каждом из направлений (Brucker, 1973).

Таблица 2. *Статистическая частота и газетные репортажи в "Register Guard" (Eugene, Oregon) и "Standard Times" (New Bedford, Massachusetts) по 41 причине смерти.*

Причина смерти	Норма по $2,05 \times 10^8$ респондентов США	Оценки испытуемых	Сообщаемые смерти		Случаи		Статьи	
			R-G	S-T	R-G	S-T	R-G	S-T
Оспа	0	57	0	0	0	0	0	0
Отравление витаминами	1	192	0	0	0	0	0	0
Ботулизм	2	183	0	0	0	0	0	0
Корь	5	168	0	0	0	0	0	0
Фейерверк	6	160	0	0	0	0	0	0
Прививка от оспы	8	23	0	0	0	0	0	0
Коклюш	15	93	0	0	0	0	0	0
Полиомиелит	17	97	0	0	0	0	0	0
Ядовитый укус животного или насекомого	48	350	0	0	0	0	0	0

Продолжение табл. 2

Торнадо	90	564	36	25	10	6	14	7
Молния	107	91	1	0	1	0	1	0
Неядовитое животное	129	174	4	2	4	2	4	2
Наводнение	205	736	4	10	2	2	2	2
Чрезмерное переохлаждение	334	314	0	0	0	0	0	0
Сифилис	410	492	0	0	0	0	0	0
Беременность, роды, аборт	451	1,344	0	0	0	0	0	0
Инфекционный гепатит	677	545	0	0	0	0	0	0
Аппендицит	902	605	0	0	0	0	0	0
Удар электрическим током	1.025	766	5	0	5	0	6	0
Столкновение поездов/автомобилей	1.517	689	0	1	0	1	0	1
Астма	1.886	506	1	0	1	0	1	0
Несчастный случай с огнестрельным оружием	2.255	1.345	8	1	8	1	9	1
Отравление твердым/жидким веществом	2.563	1.013	3	3	1	1	1	1
Туберкулез	3.690	658	0	0	0	0	0	0
Пожары и возгорания	7.380	3.336	94	46	33	9	38	10
Случаи гибели на воде	7.380	1.648	47	60	44	24	45	37
Лейкемия	14.555	2.496	1	0	1	0	1	0
Случайное падение	17.425	2.675	15	7	15	6	16	9
Убийство	18.860	5.582	278	208	167	122	329	199
Эмфизема	21.730	2.848	1	0	1	0	1	0
Суицид	24.600	4.679	29	19	28	18	36	20
Рак груди	31.160	2.964	0	0	0	0	0	0
Диабет	38.950	1.476	0	1	0	1	0	1
Автомобильная авария	55.350	41.161	298	83	245	69	180	73
Рак легких	75.850	9.764	3	2	3	2	4	2
Рак желудка	95.120	3.283	0	1	0	1	0	1
Все несчастные случаи	112.750	88.879	715	596	421	152	374	177
Инсульт	209.100	7.109	12	4	12	4	13	4
Все формы рака	328.000	45.609	25	12	25	12	26	15
Болезнь сердца	738.000	23.599	49	30	45	25	46	25

Продолжение табл. 2

Все болезни	1.740.450	88.838	111	87	100	76	104	78
Общее количество репортажей (причины 10, 11, 13, 29, 31, 37 и 41)			1.174	945	729	376	860	483
Корреляции (R-G против S-T)			$r = 0.97$	$r = 0.94$	$r = 0.98$			

Примечание: R-G = "Register Guard"; S-T = "Standard Times"

Источник: Combs & Slovic (1979).

Со мной это не случится. Суждения людей о причинах смерти, возможно, так хороши, как того можно ожидать, учитывая, что они не являются ни специалистами в рассматриваемых видах риска, ни обеспечены репрезентативными выборками информации. Точное восприятие вводящих в заблуждение выборок информации может также рассматриваться, как лежащее в основе другого очевидного предубеждения суждений – склонности людей рассматривать самих себя как обладающих личным иммунитетом от риска. Подавляющее большинство людей уверены, что они лучше, чем средние водители (Naatanen & Summala, 1975; Svenson, 1980), более вероятно, чем это есть в среднем, проживут более 80 лет (Weinstein, 1980), менее вероятно, чем это есть в среднем, понесут ущерб от потребляемых продуктов (Rethans, 1979) и так далее. Хотя такое восприятие является, очевидно, нереалистичным, в перспективе опыта отдельного индивида риск выглядит очень маленьким. Рассмотрим вождение автомобиля: несмотря на быстрое вождение, подрезание и т.д., плохие водители совершают поездку за поездкой без неудач. Этот личный опыт демонстрирует им их исключительное мастерство и безопасность. Более того, их косвенный опыт через средства массовой информации показывает им, что если происходят несчастные случаи, они происходят с другими. Получая подобный вводящий в заблуждение опыт, люди чувствуют себя достаточно оправданными, отказываясь принимать защитные меры, такие как использования ремня безопасности в автомобиле (Slovic, Fischhoff & Lichtenstein, 1978).

Вне поля зрения – вне размышлений. В некоторых ситуациях неспособность оценить пределы "имеющихся в наличии" данных, может успокаивать самодовольство людей. В исследовании Фишхоффа, Словики и Лихтенштейн (Fischhoff, Slovic и Lichtenstein, 1978) три группы студентов колледжа попросили оценить полноту дерева ошибок, показывающего риск, связанный с проблемой завести машину (см. рис. 3). Одна группа видела полное дерево. Каждая из двух других групп получила другое, сокращенное дерево. В одном варианте отсутствовали ветви пуска, зажигания и злонамерен-

ных действий; другому варианту недоставало ветвей, подробно описывающих аккумулятор, топливо и другие проблемы двигателя.

Инструкция к задаче гласила следующее (цифры в скобках были предложены людям, которые видели сокращенное дерево):

Каждый день по всем Соединенным Штатам миллионы водителей выполняют следующие действия: садятся в машину, вставляют ключ в замок зажигания, пытаются завести двигатель. Иногда двигатель не заводится, и поездка откладывается. Мы бы хотели, чтобы вы подумали о разнообразных проблемах, которые могут быть достаточно серьезными, чтобы быть причиной того, что машина не может завестись и поездка водителя откладывается, как минимум, на 1 минуту.

Диаграмма на следующей странице предназначена для того, чтобы помочь вам подумать об этих проблемах. Она показывает шесть (три) главных недостатков, являющихся причиной того, что двигатель машины не заводится. Эти главные категории, возможно, не охватывают всех возможностей, поэтому мы включили седьмую (четвертую) категорию: все другие проблемы.

Пожалуйста, тщательно рассмотрите эту диаграмму и ответьте на следующие вопросы: Оцените, в среднем, какое количество из 100 задержек, когда поездка откладывается из-за “неудачи запуска”, вызвано каждым из семи (четыре) факторов. Поставьте свои оценки в строках бланка согласно факторам, названным ниже.

Если люди, которые видели сокращенные деревья, были должным образом чувствительны к тому, что упущено, доля проблем, которые они отнесли к “другим проблемам”, уравнивает сумму долей проблем, отнесенных к сокращенным ветвям и к “другим проблемам” теми, кто видел полное дерево. Результаты в Таблице 3 показывают: что было вне поля зрения, было вне рассуждений. Например, в сокращенном дереве группы 1 “другие проблемы” должны были увеличиться в шесть раз (от 0.078 до 0.468), чтобы отразить долю нарушений из-за проблем запуска, зажигания и злонамеренных действий, которые были упущены в диаграмме. Вместо этого, “другие проблемы” были только удвоены, тогда как значение трех упомянутых систем было существенно увеличено. Второе исследование не только повторило эти результаты, но и показало, что люди, рассматривавшие сокращенные деревья, судили о неудаче пуска (по всем причинам), как менее вероятной, чем это делали те, кто наблюдал несокращенное дерево.

Машина не заводится

Недостаточный заряд аккумулятора

1. Отсутствие минусового контакта
2. Нет электрического контакта клем или их коррозия
3. Разряженный аккумулятор
 1. Краска
 2. Коррозия
 3. Грязь
 4. Плохое соединение контактов
1. Фары остались включенными при выключенном моторе
2. Возраст
3. Холодная погода
4. Поломанный генератор
5. Растянутый или поврежденный ремень вентилятора
6. Электродит низкой плотности
7. Кабельные провода разорваны
8. Неисправный генератор переменного тока
9. Неисправный гальванический регулятор (прибор)
10. Короткий внутренний цикл
11. Работает слишком много электрических аксессуаров
12. Утечка электричества
13. Постоянная небольшая утечка (коробка аккумулятора на переднем сиденье моделей 1974 года)
14. Слишком маленькая емкость аккумулятора

Неисправная пусковая система

1. Неисправные выключатели
2. Трансмиссия не нейтральная или рычаг коробки передач во включенном состоянии
3. Проблема ремня безопасности (машины 1974 года)
4. Неисправный электродвигатель стартера
5. Неисправный привод стартера
 1. Выключатель зажигания
 2. Реле стартера
 3. Нейтральное положение замка зажигания
 4. Соленоид
1. Не пристегнуты пояса
2. Неисправный выключатель ремня водителя
3. Тяжелый объект на переднем сиденье с не пристегнутым ремнем
4. Ремень застегнут до того, как сел водитель

Неисправная топливная система

1. Недостаточно топлива
2. Чрезмерное количество топлива (перелив)
3. Неисправная дроссельная заслонка
4. Неисправный воздушный фильтр
 1. В машине нет бензина
 2. Засоренный бензопровод
 3. Утечки бензопровода
 4. Грязь в топливном баке
 5. Замерзший бензопровод
6. Неправильно установленные дроссельные форсунки
7. Неисправный топливный насос
 1. Поврежденный карбюратор
 2. Многочисленные утечки прокладок
 3. Слишком высокое давление топливного насоса
1. Утечка впускного клапана
2. Не отрегулирован поплавок
3. Чрезмерный отсос акселератора
4. Чрезмерное давление топлива в жаркий день
5. Электрический топливный насос заливает карбюратор (в иностранных машинах)

1. Открыт клапан дросселя
2. Соединение клапана заклинило
3. Заслонка не работает
4. Неисправная функция электрического дросселя (Volkswagen)

Неисправная система зажигания

1. Неисправная катушка
2. Неисправный распределитель
3. Неисправные свечи зажигания
4. Неисправная проводка между компонентами

1. Треснула крышка распределителя
2. Окислившиеся электроды
3. Неправильный зазор между кулачками
4. Высокое сопротивление
5. Неисправный конденсатор
6. Несмазанная ось распределителя
7. Момент зажигания установлен неверно
8. Неправильно вставлен ключ

1. Не отрегулированные свечи
2. Замыкание в предохранителях
3. Неисправная проводка
4. Подача напряжения на предохранители происходит в неправильном порядке

Другие проблемы двигателя

1. Слишком густое масло
2. Заклинившие поршни
3. Плохая компрессия

1. Не та марка масла
2. Слишком холодная погода

1. Лопнувшее кольцо
2. Чрезмерная жара
3. Выемка кольца повреждена, испорчена или неточна

1. Протекающая прокладка головки
2. Треснута головка цилиндра
3. Клапан прогорел, неправильно отрегулирован или заклинивает
4. Поршень, кольца поршней или цилиндр изношены или повреждены
5. Бензин стекает на цилиндр

Злонамеренные акты вандализма

1. Кража или поломка жизненно важных частей (например, аккумулятора)
2. Отсос бензина
3. Разрыв проводки

Все другие проблемы

Рис. 3. Дерево ошибок, показывающее причины, по которым машина может не заводиться. Оно было использовано авторами для изучения того, являются ли люди чувствительными к полноте репрезентации этого типа. Улучшение обширных секций диаграммы, как было обнаружено, имеет незначительное влияние на степень полноты в суждениях. В результате, то, что было вне поля зрения, было вне размышлений. Профессиональные автомеханики не были оптимом лучше по выполнению этого задания, чем непрофессионалы. (*Источник: Fischhoff, Slovic & Lichtenstein, 1978*)

Таблица 3. Приписывание нарушений запуска для сокращенного и несокращенного деревьев

Группа	n	Средняя доля нарушений запуска по типу						
		Аккумулятор	Пусковая система	Топливная система	Система зажигания	Двигатель	Злонамеренные действия	Другие
Несокращенное дерево	93	0.264	0.195	0.193	0.144	0.076	0.051	0.078
Сокращенное дерево 1	29	0.432	-	0.309	-	0.116	-	0.140 ^a
Сокращенное дерево 2	26	-	0.357	-	0.343	-	0.073	0.227 ^b

Примечание: Тире означает, что данная ветвь была убрана

^a Должно быть 0.468

^b Должно быть 0.611

Источник: Slovic, Fischhoff & Lichtenstein (1978)

Чрезмерная уверенность

Знание с уверенностью. Особенно вредный аспект эвристики состоит в том, что люди типично имеют большую уверенность в суждениях, основанных на ней. В другом исследовании, развивающем идею изучения причин смерти, людей попросили указать шансы того, что они были правы, выбирая более частое из пары смертельных событий (Fischhoff, Slovic & Lichtenstein, 1977). Таблица 4 показывает процентное отношение правильных ответов по каждой из наиболее часто используемых категорий шанса. В эксперименте 1 испытуемые были разумно калиброваны, давая шансы 1:1, 1.5:1, 2:1 и 3:1. То есть, процентное соотношение их правильных ответов по этим шансам было близким к соответствующему правильному процентному соотношению. Однако тогда как шансы возрастали с 3:1 до 100:1, точность увеличивалась незначительно или вовсе не увеличивалась. Только 73% ответов с шансами 100:1 были правильными (вместо 99,1%). Точность подскочила до 81% при шансе 1.000:1 и до 87% при 10.000:1. Для ответов с шансами 1.000.000:1 или большими, точность составляла 90%; соответствующая степень уверенности была у шансов 9:1. 12% ответов, не перечисленных в табл. 3, так как они попали в промежуток между наиболее употребляемыми категориями шанса, показали сходную модель чрезмерной уверенности. Резюмируя, можно отметить, что испытуемые были часто не правы, даже на самых высоких уровнях шансов. Более того, они дали много предельных ответов по шансам. Больше половины их суждений имели шанс, больший, чем 50:1. Почти четверть – больший, чем 100:1.

Второй эксперимент пытался улучшить результаты, предлагая испытуемым больше инструкций. Экспериментальная сессия началась с 20-минутной лекции, которая подробно объясняла концепцию вероятности и шанса. Обсуждались тонкости выражения чувства неуверенности человека суждениями о численных шансах, с особым ударением на том, как использовать маленькие шансы (между 1:1 и 2:1), если довольно неуверен в правильном ответе. Была дана диаграмма, показывающая отношения между разнообразными шансами и соответствующей вероятностью. И, наконец, испытуемых обучили концепции калибровки (Глава 22) и настоятельно попросили вынести шансовые суждения таким образом, чтобы они были хорошо калиброваны. Хотя результаты немного улучшились, испытуемые снова демонстрировали неоправданную уверенность (см. табл.4). Они приписали шансы, большие или равные 50:1, приблизительно 1/3 вопросов. Только 83% ответов, связанных с этими шансами, были правильными.

Таблица 4. Процентное отношение правильных ответов для основных категорий шансов

Шансы	Соответствующий правильный % ^a	Летальные события						Вопросы на общие темы		
		Эксперимент 1 ^b			Эксперимент 2 ^b			Эксперимент 3 ^b		
		N	% N	Правильный %	N	% N	Правильный %	N	% N	Правильный %
1:1	50	644	9	53	339	8	54	861	19	53
1,5:1	60	68	1	57	108	2,5	59	210	5	56
2:1	67	575	8	64	434	10	65	455	1	63
3:1	75	189	2	71	252	6	65	157	3,5	76
5:1	83	250	4	70	322	8	71	194	4	76
10:1	91	1 167	17	66	390	9	76	376	8	74
20:1	95	126	2	72	163	4	81	66	1,5	85
50:1	98	258	4	68	227	5	74	69	1,5	83
100:1	99	1 180	17	73	319	8	87	376	8	80
1.000:1	99,9	862	13	81	219	5	84	334	7	88
10.000:1	100	459	7	87	138	3	92	263	6	89
100.000:1	100	163	2	85	23	0,5	96	134	3	92
1.000.000:1	100	157	2	90	47	1	96	360	8	94
Итого		6.098	88		2.981	70		3.855	75	
Общий правильный %				71.0			72.5			73.1

Примечание: % N относится к процентному отношению суждений о шансах, которые попали в каждую основную категорию. В Эксперименте 1 было 66 испытуемых, в Эксперименте 2 – 40 и в Эксперименте 3 – 42.

^a Для хорошо калиброванных испытуемых

^b Эксперименты 1, 2 и 3 были обозначены Эксперименты 2, 3 и 4 в первоначальном отчете.

Источник: Fischhoff, Slovic & Lichtenstein (1977)

В третьем эксперименте люди оказались такими же самоуверенными, отвечая на вопросы по общим темам (например, какой журнал имел наибольший тираж в 1970 году? (а) *Playboy* или (б) *Time*), как и в ответах на вопросы о частоте летальных событий (см. табл. 4). Дополнительное исследование Фишхоффа и других проверило веру людей в свои шансовые оценки, спрашивая, поставили бы они на них деньги, держа пари, описанное ниже:

Инструкция к игре по решению “пустяковых задачек”

Эксперимент закончился. Вы только что заработали \$2.50, которые Вы вскоре можете взять. Но перед тем как взять деньги и уйти, я бы хотел, чтобы Вы подумали, не захотели бы сыграть в определенную игру, чтобы возможно увеличить Ваш заработок. Правила игры таковы:

1. Посмотрите на Ваш ответный лист. Найдите вопросы, по которым Вы оценили Ваши шансы правильными как 50:1 или больше, чем 50:1. Сколько было таких вопросов? – – – (напишите количество).
2. Я дам правильные ответы на эти вопросы “50:1 или больше”. Мы посчитаем, сколько раз ваши ответы на эти вопросы были неправильны. Поскольку неправильный ответ при такой высокой уверенности будет удивительным, мы назовем эти неправильные ответы “вашими удивлениями”.
3. Передо мною мешочек с фишками для покера. В нем 100 белых фишек и 2 красных. Если я наугад выберу фишку, шанс, что я выберу белую фишку, равен 100:2 или 50:1, как и шанс, что ваши “50:1 или больше” ответы правильны.
4. За каждый ответ “50:1 или больше”, данный вами, я буду вытягивать из мешочка фишку. (Если хотите, вы можете сами вытягивать для меня фишки.) Я буду класть фишки в мешочек, перед тем как снова вытягивать, таким образом, шансы не изменятся. Вероятность того, что я вытяну красную фишку, равна 1/51. Поскольку вытягивание красных фишек маловероятно, каждая красная фишка, вытянутая мною, будет рассматриваться как “мое удивление”.
5. Каждый раз когда Вы удивлены неправильным ответом на вопрос “50:1 или больше”, Вы платите мне \$1 (увеличивая до 2.50 в некоторых условиях). Каждый раз, когда я удивлен, вытянув красную фишку, я плачу вам \$1.

6. Если вы хорошо калиброваны, эта игра для Вас выгодна. Это потому, что я ожидаю потерять \$1 один раз из 51 попытки вытягивания фишки, в среднем. Но поскольку ваши шансы иногда выше, чем 50:1, вы ожидаете иметь потери даже реже, чем я.

7. Будете ли вы играть в эту игру?

Это пари выгодно для отлично калиброванных и недостаточно уверенных участников и невыгодно для самоуверенных. Большинство участников нашего исследования с нетерпением хотело сыграть в эту игру. Поскольку их уверенность была неоправданна, они терпели ощутимые денежные потери (которые мы вернули им после окончания исследования).

Хотя психологическая основа неоправданной уверенности сложна, ключевым элементом оказывается недостаточное осознание людьми того, что их знания основаны на предположениях, которые часто являются незначительными. Например, 30% респондентов в Эксперименте 1 дали шансы, большие, чем 50:1, неправильному утверждению о том, что убийства более часты, чем самоубийства. Этих испытуемых ввело в заблуждение то, что примеры убийства гораздо легче припомнить, и они оказались неспособными оценить, что запоминаемость – это несовершенная основа для подобных выводов.

Гиперточность. Чрезмерная уверенность также может проявляться и другими способами. Типичная задача по оценке неопределенных количеств, таких как нормы неудач, состоит в установке верхних и нижних границ таким образом, что существует шанс 98%, что правильное значение лежит между ними. Эксперименты с разнообразными группами людей, выносящими различные виды суждений, показали, что скорее не 2% правильных значений выпадают из 98% границ уверенности, а 20 – 50% (глава 21 и 22). Таким образом, люди думают, что они могут оценить такие значения с гораздо большей точностью, чем это происходит на самом деле. Тверски и Канеман (Tversky и Kahneman, 1974, 1) отнесли такую гиперточность за счет опоры на эвристику привязки и регулирования.

Чрезмерно уверенные эксперты. К сожалению, эксперты, вынужденные выйти за пределы данных и опираться на суждения, могут быть так же склонны к чрезмерной уверенности, как и обыватели. Фишхофф, Словик и Лихтенштейн (Fischhoff, Slovic и Lichtenstein, 1978) повторили свое исследование по дереву ошибок (рис. 3) с профессиональными автомеханиками (в среднем, 15 лет опыта) и обнаружили, что эти эксперты также нечувствительны к упущениям из дерева (см. табл. 5). Хайнс и Ванмарк (Hynes и Vanmarcke, 1976) попросили семерых “всемирноизвестных” инженеров по геотехнике предсказать высоту плотины, которая приведет к обвалу глиняного фундамента, и определить границы уверенности по этой оценке, которые были бы достаточно широки, чтобы иметь 50% шанс включения действительной высоты, приводящей к разрушению. Ни одни из границ, опре-

деленные этими людьми, в действительности, все-таки не включали истинную высоту. Рисунок 4 показывает эти результаты.

Таблица 5. *Нечувствительность экспертов к упущениям в дереве ошибок "машина не заводится"*

Группа	n	Средняя доля нарушений запуска по типу						
		Аккумулятор	Система стартинга	Топливная система	Система зажигания	Двигатель	Злонамеренные действия	Другие
Несокращенное дерево, обычные испытуемые	93	0.264	0.195	0.193	0.144	0.076	0.051	0.078
Несокращенное дерево, эксперты	13	0.410	0.108	0.096	0.248	0.051	0.025	0.060
Сокращенное дерево 1, эксперты	16	0.483	-	0.229	-	0.073	-	0.215 ^a

^a Должно быть 0.441

Источник: (Fischhoff, Slovic и Lichtenstein, 1978).

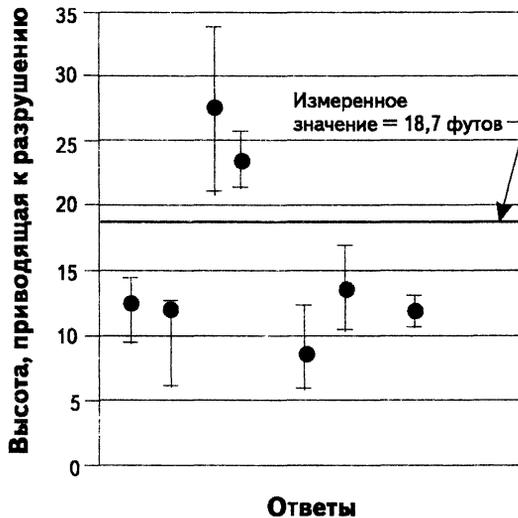


Рис. 4. Пример чрезмерной уверенности в суждениях экспертов, представленный неспособностью ошибочных линий вместить действительное значение. Данные представляют оценки семи "всемирноизвестных" инженеров по геотехнике относительно высоты платформы, при которой она обвалится

Многомиллионное исследование безопасности реакторов (Комиссия по атомному регулированию Соединенных Штатов, 1975), оценивая вероятность плавления ядра в атомном реакторе, использовала ту же процедуру для установления границ уверенности, которая, как было показано в Главах 21 и 22, давала высокую степень чрезмерной уверенности. Фактически, “Комитет Льюиса” заключил свой обзор исследования безопасности реакторов замечанием, что, несмотря на огромные успехи этого исследования, “мы уверены, что ряды ошибок преуменьшены. Мы не можем сказать, насколько. Основания для этого включают неадекватные базы данных, плохую статистическую обработку неконтролируемое распространение неопределенностей в расчетах” (Комиссия по атомному регулированию, США, 1978).

Дальнейшие эпизодические свидетельства чрезмерной уверенности могут быть обнаружены во многих других оценках технического риска (Fischhoff, 1977a). Некоторые распространенные способы, по которым эксперты могут недосмотреть или неправильно вынести суждение о путях, ведущих к катастрофе, показаны в следующем списке:

Неспособность рассмотреть способы, по которым ошибки людей могут воздействовать на технологические системы. Пример: из-за неадекватного обучения и дизайна комнаты контроля, операторы Three Mile Island неоднократно неправильно диагностировали проблемы реактора и предпринимали несоответствующие действия (Sheridan, 1980; Президентская Комиссия, 1979).

Чрезмерная уверенность в текущих научных знаниях. Пример: использование передачи цифровых данных стало повсеместным и бесконтрольным даже до того, как ученые рассмотрели возможность побочных эффектов, которые сегодня сделали ее сомнительным и необратимым благом (Dunlap, 1978).

Неспособность оценить, как технологические системы функционируют в целом. Пример: DC-10 потерпел неудачу в нескольких предыдущих полетах, так как его проектировщики не осознали, что декомпрессия грузового отделения разрушит жизненно важные системы контроля (Hohenemser, 1975).

Медленное обнаружение хронических, совокупных воздействий. Пример: хотя несчастные случаи с шахтерами уже давно были признаны как одна из цен за использование ископаемого растительного топлива, воздействие кислотных дождей на экосистему обнаруживалось медленно.

Неспособность предвидеть ответные действия человека на меры безопасности. Пример: частичная защита, которую дают дамбы и плотины, приносит людям ошибочное ощущение безопасности и способствует разработке затопляемых территорий. Таким образом, хотя наводнения и редки, ущерб от них настолько велик, что среднегодовая потеря в деньгах больше, чем перед тем, как дамбы были построены. (Burton, Kates & White, 1978)

Неспособность предвидеть “нарушения обычной модели”, которые одновременно поражают системы, спроектированные как независимые. Пример: так как электрические кабели, контролирующие множественные системы безопасности реактора на Браунс Ферри, Алабама, были пространственно не отделены, все пять систем экстренного охлаждения ядра были одновременно повреждены одним пожаром (Палата Представителей США, 1975; Jennergren & Keeney, в печати).

Крушение дамбы Тетон в 1976 году дает еще один трагический пример самоуверенности экспертов. Комитет по правительственным операциям приписал эту катастрофу неоправданной уверенности инженеров, которые были абсолютно уверены, что они решили многие серьезные проблемы, возникшие во время строительства (Комитет по правительственным операциям, 1976). Обычно, вероятность неудачи даже не вычисляется для новых дамб, даже хотя 1 из 300 разрушается во время первоначального заполнения резервуара.

Информирование людей о риске

Трудно определенно думать о риске. К сожалению, это совершенно необходимо. Радиационный риск, медицинские побочные эффекты, профессиональные болезни, заражение пищевых продуктов, токсичные химические препараты и неправильное срабатывание различных устройств все больше заполняют наши газеты и наши мысли. Поскольку управление этими рисками жизненно важно для благосостояния индивида и общества, люди в настоящее время настаивают на своем праве играть активную роль в процессе принятия решений. Как результат этого, промоутеры и ответственные лица опасных предприятий сталкиваются с увеличивающимся давлением по вопросу информирования людей о риске, с которым они сталкиваются (см. рис. 5). Например, за последние годы:

Администрация по Продуктам и Лекарствам предписала делать информационные вкладыши для пациентов по лекарствам, отпускаемым по рецептам, количество которых растет.

Департамент жилищного строительства и городского развития начал требовать от продавцов домов, построенных до 1950 года, информирования покупателей о присутствии свинцовых красок.

Предложенный закон об обязательствах федеральной продукции отвел значительное место информированию покупателей и рабочих о риске, с которым они могут, вероятно, столкнуться.

Белый Дом приказал Государственному секретарю по медицине, образованию и благосостоянию разработать общую информационную программу по влиянию радиационного воздействия на здоровье.

Несмотря на эти хорошие намерения, разработка информационной программы может быть довольно сложной. Делать это адекватно – означает найти убедительные способы представления сложного, технического материала, что часто омрачается неопределенностью. Не только из-за иногда слишком ограниченного отведенного времени, но также из-за того, что сообщения должны противопоставлять предвзятое мнение слушателей (и, возможно, неправильное мнение) о рассматриваемом риске его последствиям. Например, в некоторых ситуациях неправильный личный опыт может содействовать ошибочному чувству безопасности, тогда как при других обстоятельствах только обсуждение возможных неблагоприятных последствий может увеличить их кажущуюся угрозу. Более того, как продемонстрировали Росс и Андерсен (Глава 9), убеждения людей часто изменяются медленно и показывают чрезвычайную устойчивость перед лицом опровергающего свидетельства. Далее идет краткий обзор некоторых дополнительных проблем, с которыми сталкиваются информационные программы.



Рис. 5. Рисунок С. Харриса, 1979 The New Yorker Magazine (продукты питания: риски, выгоды)

Важен формат представления

Определенная манера представления риска может иметь первостепенное влияние на восприятие и поведение. Например, деятельность, увеличивающая годовой шанс смерти от 1 на 10.000 до 1,3 на 10.000, вероятно, казалась бы намного более рискованной, если бы была описана вместо этого как дающая 30%-ное увеличение риска смертности. Образцы эффектов формата из литературы представлены ниже.

Деревья ошибок. Разработчики деревьев ошибок, подобного тому, что на рис.3, должны принимать многочисленные осмотрительные решения в отношении того, как организовать и представить разнообразные источники проблемы. Одно такое решение, которое явно дает незначительное различие в ответах, состоит в том, какое количество подробностей должно быть предложено; Фишхофф, Словик и Лихтенштейн (Fischhoff, Slovic и Lichtenstein, 1978) обнаружили сходное восприятие с различным уровнем подробностей. Одно лишь упоминание ветви позволяло людям точно оценить, насколько проблемной будет эта ветвь, описанная подробно. Однако, объединяющие ветви (например, соединяющие пусковую систему и систему зажигания в одну более широкую категорию) или расщепляющие ветви (например, разделяющие систему зажигания на систему зажигания (неисправная катушка и свечи зажигания, см. рис. 3) и систему распределения (неисправный распределитель и проводка)) все же давали разницу в ответах. О данном наборе проблем судили как об объясняющем где-то на 30% больше нарушений, когда он был представлен как две ветви, чем когда он был представлен как одна ветвь.

Ремни безопасности. Вторая демонстрация значения формата презентации идет из изучения отношений к использованию автомобильных ремней безопасности (Slovic, Fischhoff & Lichtenstein, 1978). Рассматривая предыдущее исследование, демонстрирующее критическое значение вероятности вреда в защитном действии пристегивания (Slovic, Fischhoff, Lichtenstein, Corrigan & Combs 1977), Словик, Фишхофф и Лихтенштейн доказывали, что нежелание людей добровольно надевать ремни безопасности может быть вызвано предельно маленькой вероятностью попадания в аварию с летальным исходом во время единственной поездки. Поскольку такие аварии случаются только один раз на 3,5 млн личных поездок, а приводящее к нетрудоспособности повреждение только один раз на 100.000 личных поездок, отказ надеть ремень безопасности может казаться вполне обоснованным. Однако он выглядит менее здравым, если принять во внимание перспективу множественных поездок и рассмотреть существенную вероятность аварии в какой-либо из поездок. За 50 лет вождения (около 40.000 поездок) вероятность смерти увеличивается до 0.01, а вероятность получить, по крайней мере, одно приводящее к нетрудоспособности повреждение до 0.33. В экспериментальном исследовании Словик, Фишхофф и Лихтенштейн по-

казали, что люди, которых просили рассмотреть эту перспективу длиною в жизнь, отвечали в отношении ремней безопасности (и воздушных подушек) более благосклонно, чем это делали люди, которых просили рассмотреть перспективу единичных поездок. Будет ли сохранено и введено в поведение благосклонное отношение к ремням безопасности, вызванное удлинённой временной перспективой, остается под вопросом.

Псевдоуверенность. Согласно “теории перспективы (prospect theory)” (Kahneman & Tversky, 1979b), исходы, которые лишь просто возможны, недооцениваются в сравнении с исходами, которые получены с уверенностью. Как результат, любые защитные действия, которые уменьшают вероятность вреда, скажем, с 0.01 до 0, будут оценены значительно выше, чем действия, уменьшающие вероятность с 0.02 до 0.01.

Тверски и Канеман (Tversky и Kahneman, 1981) отмечают, что умственной репрезентацией защитных действий можно легко манипулировать с целью изменения кажущейся уверенности, с которой они предотвращают вред. Например, страховой полис, который охватывает пожары, но не наводнения, можно было бы представить либо как полную защиту от специфического риска пожаров, либо как сокращение в пределах общей вероятности потери собственности. Теория перспективы предсказывает, что полис будет казаться более привлекательным с точки зрения первой перспективы (названной “псевдоуверенностью”), по которой он предлагает безусловную защиту от ограниченного набора проблем.

Мы проверили это предположение в контексте одного конкретного вида защиты – вакцинации. Было создано две формы “опросного листа по вакцинации”. Форма I (вероятностная защита) описывала болезнь, которая, как ожидалось, охватит 20% населения, и в ней спрашивалось, желают ли люди добровольно привиться вакциной, которая защищает половину из привившихся. Согласно Форме II (псевдоуверенность), существовало две взаимоисключающих и равновероятных разновидности болезни, каждая из которых, вероятно, охватит 10% населения; было сказано, что вакцинация даст полную защиту от одной разновидности и никакой защиты от другой. Участники исследования были набраны по объявлению в студенческой газете Университета штата Орегон. Половина из них получила Форму I, другая – Форму II. После прочтения описания они оценивали вероятность того, что они пройдут вакцинацию в такой ситуации, используя шкалу, ранжированную от 1 (“почти наверняка не буду прививаться”) до 7 (“почти наверняка буду прививаться”).

Хотя обе формы показывали, что вакцинация уменьшала общий риск человека с 20% до 10%, мы ожидали, что она покажется более привлекательной тем, кто получил Форму II (псевдоуверенность), чем тем, кто получил Форму I (вероятностная защита). Результаты подтвердили это предсказание: 57% тех, кто получил Форму II, показали, что они пройдут вакцинацию, в сравнении с 40% тех, кто получил Форму I.

Эффект псевдоуверенности придает первостепенное значение контрасту между сокращением и ликвидацией риска. Как показали Тверски и Канеман, это различие очень трудно оправдать на каких бы то ни было основаниях. Более того, манипуляции уверенностью, по-видимому, значительно вовлечены в разработку и описание других форм защиты (например, медицинское лечение, страхование, деятельность по защите от наводнений и землетрясений).

Привязка. Один из наиболее общих артефактов презентации состоит в тенденции привязки суждений на изначально представленным значениям (Poulton, 1968; Tversky & Kahneman, 1974, 1). В другой разновидности эксперимента, представленного на рис. 2, Лихтенштейн и другие (1978) просили вторую группу людей оценить частоту смертей в США от каждой из 40 причин. Однако, вместо того, чтобы сказать, что около 50.000 людей ежегодно умирают в автомобильных авариях, этим испытуемым сообщили, что около 1.000 ежегодных смертей происходит от электричества. Хотя оба сообщения были точны, предоставление более маленькой цифры уменьшило оценки респондентов по большинству частот. Подобная привязка к первоначальной цифре привела к тому, что оценки двух групп отличались в некоторых случаях в 5 раз.

Фишхофф и МакГрегор (Fischhoff и McGregor, 1980) попросили людей вынести суждение о смертности от разнообразных потенциальных причин, используя один из четырех формально эквивалентных форматов (например, сколько людей выжило на каждого, кто был поражен и умер? Сколько людей умрет из 100.000 тех, кто был поражен?). В таблице 6 представлены их суждения в общем формате и открывает даже более драматичные эффекты влияния вопросительного суждения на восприятие риска. Например, когда люди непосредственно оценивали количество смертей для гриппа (колонка 1), их средний ответ был 393 смерти на 1.000 случаев. Когда им сообщили, что 80.000.000 людей подхватывают грипп в обычный год, и попросили их оценить количество тех, кто умирает (колонка 2), средний ответ респондентов был 4.800, представляющий смертельную норму только как 6 на 100.000 случаев. Это незначительное изменение вопроса изменило оцениваемую норму больше, чем в 60 раз. Подобные противоречия случаются и с другими вопросами и другими видами риска.

Другие эффекты. Многочисленные другие эффекты формата были задокументированы в литературе по рискованному поведению. Например, обнаружилось, что люди оценивают опасные предприятия совершенно иначе, когда они рассматривают их в парах, чем когда судят о них по отдельности (Grether & Plott, 1979; Lichtenstein & Slovic, 1971; 1973). Фишхофф, Словик и Лихтенштейн (Fischhoff, Slovic и Lichtenstein, 1980), Херши, Скоумейкер (Hershey и Schoemaker, 1980) и Скоумейкер и Канрейтер (Schoemaker и Kunreuther, 1979) отметили, что решения о том, покупать ли страховой полис, часто отменяются, когда проблема выглядит как выбор между учас-

Таблица 6. Суждения о смертности при различных ответных моделях, геометрические средние значения.

Болезнь	Норма смертей на 100.000 пораженных				
	Оцененная норма смертности	Оцененное количество тех, кто умрет	Оцененная норма выживания	Оцененное количество тех, кто выживет	Действительная норма смертности
Грипп	393	6	26	511	1
Свинка	44	114	19	4	12
Астма	155	12	14	599	33
Венерические болезни	91	63	8	111	50
Высокое кровяное давление	535	89	17	538	76
Бронхит	162	19	43	2.111	85
Беременность	67	24	13	787	250
Диабет	487	101	52	5.666	800
Туберкулез	852	1.783	188	8.520	1.535
Автомобильные аварии	6.195	3.272	31	6.813	2.500
Инсульт	11.011	4.648	181	24.758	11.765
Сердечный приступ	13.011	3.666	131	27.477	16.250
Рак	10.889	10.475	160	21.749	37.500

Примечания: четырем экспериментальным группам дали следующие инструкции:

(а) Оцените количество смертей: сколько людей умирает на каждые 100.000 пораженных?

(б) Оцените количество тех, кто умер: X людей были поражены, сколько из них умерло?

(в) Оцените количество выживших: Сколько пораженных людей выжило на каждого, кто умер?

(г) Оцените количество тех, кто выжил: Y людей умерло, сколько было поражено, но выжило?

Ответы на вопросы (б), (в) и (г) были пересчитаны в количество смертей на 100.000 пораженных для облегчения сравнения.

Источник: Fischhoff & McGregor, 1980.

тием в рискованном предприятии или потерей меньшей суммы денег. Подобные опции риска, описанные в терминах спасенных жизней, могут быть оценены совершенно иначе, будучи структурированными в терминах потерянных жизней (Tversky & Kahneman, 1981). Дополнительные эффекты формата и контекста можно обнаружить у Fischhoff, Slovic и Lichtenstein (1980), Kahneman и Tversky (1970), Slovic, Fischhoff и Lichtenstein (в печати) и Tversky и Kahneman (1981).

Такие незначительные различия в представлении риска могут иметь заметный эффект в том, как он воспринимается, и это наводит на мысль, что те, кто ответственен за информационные программы, имеет значительные возможности манипулирования восприятием. Более того, поскольку эти эффекты не являются широко известными, люди неумышленно могут манипулировать своим собственным восприятием посредством случайных решений о том, как организовать свои знания.

Сравнение различных видов риска может вводить в заблуждение.

Один из наиболее распространенных подходов к углублению понимания людей состоит в представлении количественных оценок для различных видов риска. Предположительно, опыт, почерпнутый при рассмотрении подобных данных, будет полезным для процесса принятия решений, как в масштабе личности, так и общества. Вилсон (Wilson, 1979) заметил, что нам следует “постараться измерить все наши виды риска количественно... Тогда мы могли бы сравнить эти виды риска и решить, какие из них принять, а какие отвергнуть” (с. 43). Лорд Ротшильд (Lord Rothschild, 1979) добавил: “Нет смысла впадать в панику относительно рисков жизни до тех пор, пока вы не сравните те риски, которые вас беспокоят, с теми, которые нет, но, возможно, должны беспокоить”.

Обычно, подобные рассуждения сопровождаются детально разработанными таблицами и даже “каталогами рисков”, в которых представлены различные индексы смерти или нетрудоспособности для широкого спектра жизненных рисков. Так, Соуби (Sowby, 1965) обеспечил экстенсивные данные по риску в течение 1 часа, когда человек ему подвергается, показывающие, например, что 1 час езды на мотоцикле так же рискован, как и 1 час пребывания в 75-летнем возрасте. Вилсон (Wilson, 1979) разработал таблицу деятельности (например, полет на 1.000 миль на реактивном самолете, 3 часа в угольной шахте), каждая из которых рассматривается как увеличивающая годовой шанс смерти человека на 1 из 1 млн. Вилсон заявил, что “эти сравнения помогают мне оценить риски, и мне представляется, что они также могут помочь в этом и другим. Но важнейшее использование этих сравнений должно помогать нам принимать решения, как нации, чтобы улучшить наше здоровье и уменьшить норму несчастных случаев” (с. 45). Подобным образом Коуэн и Ли (Cohen и Lee, 1979) проранжировали многие виды риска в терминах того, как они сокращают продолжительность жиз-

ни, на основании предположения, что “при некотором приближении, порядок рисков (в этой таблице) должен быть порядком социальных приоритетов. Однако мы видим несколько важных проблем, которые получили очень незначительное внимание..., тогда как некоторые пункты в конце списка, особенно включающие радиацию, привлекают огромное внимание” (с. 720).

Собственно говоря, сравнение рисков не есть процедура процесса принятия решений. Оно не требует выведения особых заключений, скажем, по контрасту между риском езды на мотоцикле и риском престарелого возраста (Fischhoff, Lichtenstein, Slovic, Derby & Keeney, 1981). Более того, даже в качестве помощи интуиции, сравнение рисков имеет ряд присущих ограничений. Например, хотя некоторые люди чувствуют себя информированными, зная, что один взлет и приземление самолета сокращает продолжительность жизни в среднем на 15 минут, другие оказываются совершенно озадачены такой информацией. При приземлении, либо ты преждевременно умрешь (почти наверняка, больше, чем на 15 минут), либо не умрешь. Для многих людей средние значения не охватывают адекватно сущность подобных рисков. В самом деле, МакНейл, Вайпельбаум и Паукер (McNeil, Weichselbaum и Pauker, 1979) обнаружили, что пациенты, сталкивающиеся с перспективой операции рака легких, так же интересовались возможной угрозой смерти во время операции, как и ее вкладом в продолжительность жизни.

Следующее ограничение состоит в том, что краткие суммарные статистические данные могут маскировать важные характеристики риска. Там, где есть неуверенность или несогласие с фактами, представление точечных оценок может внушать чрезмерную уверенность. Поскольку люди особенно интересуются потенциалом катастрофических несчастных случаев (Slovic, Fischhoff & Lichtenstein, 1980), необходимо некоторое указание вероятности и величины предельных потерь. Другими характеристиками, которые воздействуют на отношение людей к риску, но пренебрегаются в статистических изложениях, являются добровольность, контролируруемость, осведомленность, непосредственность последствий, угроза будущим поколениям, легкость сокращения риска и степень справедливости распределения выгоды между теми, кто подвергается риску (Slovic, Fischhoff & Lichtenstein, в печати). Хотя некоторые ошибки, такие как упущение связей неопределенностей, легко исправить, определение того, как оценивать потенциал катастрофы, справедливость и другие важные характеристики, будет требовать серьезных исследовательских усилий.

Выводы

Информирование людей предупреждающими этикетками, вкладышами в упаковку или экстенсивными программами СМИ – это все же только часть гораздо большей проблемы помощи людям в том, как справляться с риском и неопределенностями современной жизни. Мы убеждены, что некоторая

ответственность лежит на наших школах. Учебный план общественных школ должен включать материал, разработанный для того, чтобы обучить людей тому, что наш мир является вероятностным, а не детерминистским, и чтобы помочь изучить стратегии суждений и решений, чтобы иметь дело с этим миром (Beuth-Marom & Dekel, в печати). Эти стратегии так же необходимы для навигации в мире неопределенной информации, как геометрия и тригонометрия – для навигации в мире физических объектов.

Атомная энергия: изучение случаев восприятия риска

Нигде больше вопросы воспринимаемого риска не являются такими яркими или ставки такими высокими, как в полемике вокруг атомной энергии. Этот раздел рассматривает эту полемику в свете только что сделанных выводов.

Общая проблема

Даже перед аварией на Три Майл Айленд атомная индустрия садилась на мели враждебного общественного мнения. Значительное и упорное оппозиционное движение было в ответе за дорогостоящие отсрочки в лицензировании и сооружении новых атомных станций в Соединенных Штатах и за политическую суматоху в нескольких европейских государствах.

Неисправный реактор на Три Майл Айленд стимулировал предсказуемый немедленный подъем в антиядерной лихорадке. Любая попытка планировать роль атомной энергии в энергетическом будущем нации должна считаться с решающими факторами этой оппозиции и предвидеть ее дальнейший курс. Один ключ к разгадке лежит в недавнем исследовании, показывающем, что образы потенциальных ядерных катастроф, которые сформировались в сознании антиядерной общественности, значительно отличаются от оценок, выдвинутых многими техническими экспертами. Мы опишем эти образы и поразмышляем об их происхождении, постоянстве и последствиях.

Основные восприятия

Анкетное исследование людей, противостоящих атомной энергии, показывает, что они судят о своей выгоде, как о достаточно низкой, а о своем риске – как о неприемлемо большом (Fischhoff, Slovic, Lichtenstein, Read & Combs, 1978). Что касается выгодной стороны, то эти люди не видят атомную энергию как жизненно важное звено в удовлетворении основных энергетических потребностей (Pokorny, 1977); скорее, многие рассматривают ее как дополнение к другим источникам энергии, которые сами являются адекватными (или их можно сделать адекватными путем защиты). Что касается риска, то атомная энергия вызывает большее чувство страха, чем почти

любая другая технологическая деятельность (Fischhoff и другие, 1978). Некоторые приписывают эту реакцию боязни невидимого и необратимого радиационного заражения, угрожающего раком и ущербом на генетическом уровне. Однако, использование диагностических рентгеновских лучей, радиационной технологии, которая несет сходный риск, не так страшно. Напротив, этот риск часто недооценивается (Slovic, Lichtenstein & Fischhoff, 1979). Ассоциация атомной энергии с ядерным оружием может объяснить эти различия восприятия. Как результат ее насильственных истоков, к атомной энергии относятся как к технологии, чей риск является неконтролируемым, смертельным и потенциально катастрофическим. Эти характеристики не ассоциируются с применением диагностических рентгеновских лучей.

Когда люди, противостоящие атомной энергии, описывают свои ментальные образы ядерной аварии и ее последствий, они обнаруживают ожидание, что серьезная авария реактора вероятна при их жизни, и в ее результате могут быть сотни тысяч, даже миллионы смертей (Slovic, Fischhoff & Lichtenstein, 1979; Slovic, Lichtenstein & Fischhoff, 1979). Также ожидается, что подобная авария может стать причиной непоправимого ущерба окружающей среде на огромной географической территории. Эти ожидания драматически контрастируют с мнением атомной индустрии, что множественные системы безопасности сделают ущерб минимальным.

Одним неизбежным последствием этого “расхождения в восприятии” является неопределенность и недоверие части общественности, считающей, что риск значительно больше, чем говорят оценки экспертов (Kasper, 1979; Starr & Whipple, 1980). Эксперты, в свою очередь, подвергают сомнению рациональность общественности и осуждают “эмоциональность”, ставящую в тупик технологический прогресс. В результате – резкая и иногда грубая конфронтация.

Признание этого расхождения в восприятии привело к заявлению экспертов, что общественность должна быть “образована” в отношении “реального” риска атомной энергии. Один аналитик общественного мнения (Pokorny, 1977) определил вопрос следующим образом:

Наибольшая проблема, мешающая искушенному суждению по этому вопросу, - это недостаток знаний и фактов. В такой обстановке ужасные рассказы, путаница и иррациональность часто одерживают верх. Только посредством подробного изучения фактов и знаний люди могут узнать, какой выбор является реальным.... (с. 12)

Наша точка зрения состоит в том, что усилия, предпринятые, чтобы уменьшить расхождение в восприятии, сталкиваются с более важными препятствиями. Это заключение основано на двух ключевых аспектах проблемы, техническом и психологическом.

Технические препятствия

Техническая реальность такова, что существует немного определенных фактов, относящихся к вероятности серьезных повреждений реактора. Технологія так нова, а рассматриваемые вероятности так малы, что точные оценки риска не могут быть основаны на эмпирическом наблюдении. Вместо этого, подобные оценки извлекаются из сложных математических моделей и субъективных суждений.

Сложность представления оценок риска заставила многих критиков поставить под сомнение их валидность (Bryan, 1974; Fischhoff, 1977; Primack, 1975). Главный интерес состоит в том, что важные иницирующие события или пути, ведущие к нарушениям, могут быть упущены, что вызовет недооценку риска. Другая проблема оценки надежности проектов реакторов – это сложность принятия в расчет “нарушений общей модели”, в которых мнимо независимые системы, разработанные таким образом, чтобы поддерживать друг друга, выходят из строя из-за одной непредвиденной общей причины. Критик атомной энергии Джон Холдрен (John Holdren) скептически относительно обоснованности оценок редких катастроф и кратко излагает техническую проблему так:

Сообщество экспертов разделилось в отношении возможного реализма вероятностных оценок в пределах от 1 на 10. 000 до 1 на 1 млрд за реакторный год. Я среди тех, кто убежден в невозможности *в принципе* аргументировать цифры, настолько малые, как эти, убедительными теоретическими доводами. Причина, по которой я придерживаюсь такого взгляда, прямая: атомно-энергетические системы так сложны, что вероятность серьезных ошибок, содержащихся в анализе безопасности, так велика, что может сделать бессмысленной крошечную вычисленную вероятность аварии (Holdren, 1976, с. 21).

Психологические препятствия

Общественную боязнь атомной энергии не следует рассматривать как иррациональную. Частично она подпитывается осознанием того, что факты оспариваются, и что эксперты ошибались в прошлом, когда они облучали увеличенные миндалевидные железы или позволяли людям наблюдать испытания ядерных бомб с близкого расстояния. Что можно подвергнуть сомнению, так это степень, в которой фундаментальные способы мышления людей (такие как опора на эвристику доступности) приводят их к искаженным взглядам. Конечно, риск атомной энергии покажется главным кандидатом для предубеждения доступности из-за обширных репортажей в СМИ, посвященных ему, и его ассоциации с яркими, легко воображаемыми опасностями ядерной войны.

Как было отмечено ранее, эвристика доступности подразумевает, что любое обсуждение ядерных аварий может увеличить их воображаемость и,

таким образом, их воспринимаемый риск. Рассмотрим инженера, доказывающего безопасность размещения ядерных отходов в солевых шахтах путем указания на невероятность разнообразных способов случайного выброса радиации (рис. 1). Вместо того чтобы успокаивать аудиторию, эта презентация может заставить их подумать: “Я и не представлял себе, что проблем может быть так много”. В этом случае, опора на запоминаемость и воображаемость может затуманить различие между тем, что отдаленно возможно, и тем, что вероятно. Как жаловался один сторонник использования атомной энергии: “Когда непрофессионалы обсуждают то, что *могло бы* случиться, они иногда даже не беспокоятся о том, чтобы использовать слова “могло бы” (Cohen, 1974, с. 36). Другой аналитик рассмотрел подобную тему в неправильной интерпретации сценариев “наихудшего”:

Часто не было большой разницы, насколько странным или невероятным было предположение в анализе подобного рода, поскольку нужно было всего лишь показать, что некоторый нежелательный эффект может иметь место с уровнем вероятности выше нуля. Оппоненты предложенной операции могли очень просто ее разрушить, применив свое воображение, чтобы представить набор условий, которые хотя и могут быть, как признано, предельно невероятными, способны привести к нежелательным результатам. С преобладанием таких отношений планирование данной ядерной операции становится... опасным (Cohen, 1972, с. 55).

Выводы

Хотя изложенная выше дискуссия определила некоторые возможные источники расхождения восприятия сторонников и противников использования ядерной энергии, она не указала определенно на одну или другую сторону как имеющую наиболее точные оценки общего риска атомной энергии. Эффекты запоминаемости и воображаемости способны как усилить общественные опасения, так и сделать неясной осведомленность экспертов относительно ситуаций, в которых система может выйти из строя. Поскольку о действительном риске можно никогда не узнать с большой точностью, а новая информация имеет тенденцию быть интерпретированной таким образом, чтобы быть совместимой с предыдущими убеждениями человека, расхождение восприятия может оставаться довольно долгое время. Так, для некоторых людей Три Майл Айленд “доказал” возможность катастрофического плавления, тогда как для других – подтвердил их веру в надежность систем множественной безопасности и сдерживания.

Кому решать?

Исследование, описанное в этой главе, демонстрирует, что суждения о риске ошибочны. Оно также показывает, что степень ошибочности часто неожиданно велика, и что ошибочных оценок можно придерживаться с боль-

шой уверенностью. Поскольку даже хорошо проинформированные непрофессионалы испытывают трудности, вынося точное суждение о риске, есть искушение сделать заключение, что общественность следует отстранить от процессов оценки и принятия решений риска в обществе. Подобное действие было бы неправильным по нескольким основаниям. Во-первых, пристальное рассмотрение показывает, что люди все-таки воспринимают некоторые вещи достаточно хорошо, хотя их понимание часто может довольно сильно отличаться от понимания технических экспертов. В ситуациях, где сильно распространено неправильное понимание, след ошибок людей может идти к предубежденному опыту, которому может противостоять образование. В некоторых случаях след сильных опасений людей и их сопротивление убеждениям экспертов может идти от их чувствительности к потенциалу катастрофических несчастных случаев, от их осведомленности о несогласии экспертов в отношении вероятности и величины таких несчастных случаев и от их знаний о серьезных ошибках, сделанных экспертами в прошлом. Даже в сложных случаях, таких как конфликт вокруг атомной энергии, атмосфера доверия и признания того, что и эксперты, и непрофессионалы могут сделать свой вклад, может позволить некоторый обмен информацией и углубление перспектив.

Более того, во многих, если не в большинстве случаев, эффективное управление риском требует сотрудничества большого количества непрофессионалов. Эти люди должны соглашаться на отсутствие некоторых вещей и принимать заместители других; они должны благоразумно голосовать по поводу избирательных мер и за законодателей, которые будут служить им как суррогатные менеджеры риска; они должны подчиняться правилам безопасности и ответственно пользоваться правовой системой. Даже если эксперты были лучшими судьями в отношении риска, чем непрофессионалы, наделение экспертов эксклюзивным правом голоса в управлении риском означало бы замещение краткосрочной действительностью долгосрочных попыток, необходимых для создания информированного гражданского населения.

Результаты, которые мы обсудили, ставят перед не-экспертами важную серию проблем: быть лучше информированными, меньше полагаться на нерассмотренные и не поддержанные суждения, осознавать факторы, которые могут оказывать влияние на суждения о риске и быть более открытыми к новым свидетельствам; если кратко, осознавать потенциал образованности.

Перед экспертами и деятелями политики эти результаты ставят, возможно, более сложную проблему: осознать и принять собственные когнитивные ограничения, попытаться обучать, не пропагандируя, признать законность интересов общественности и каким-то образом разработать способы, посредством которых эти интересы могли бы найти выражение в общественных решениях, не производя в процессе больше жара, чем света.

Часть X
Постскриптум

34. К вопросу о статистической интуиции*

Даниель Канеман и Амос Тверски

Большая часть вышедшей недавно литературы по суждениям и интуитивным размышлениям рассматривала ошибки, предубеждения и заблуждения в разнообразии мыслительных задач (см., например, Einhorn & Hogart, 1981; Hammond, McClelland & Mumpower, 1980; Nisbett & Ross, 1980; Shweder, 1980; Slovic, Fischhoff & Lichtenstein, 1977; Tversky & Kahneman, 1974, 1). Акцент на изучении ошибок является характерным для исследования суждений человека, но оно не уникально в этой области: мы используем иллюзии, чтобы понять принципы нормативного восприятия, и мы рассматриваем память, изучая процесс забывания. Однако ошибки рассуждений уникальны среди когнитивных сбоев в двух существенных отношениях: Они несколько смущают, и их, по-видимому, избегают. Нас не беспокоит наша чувствительность к вертикально-горизонтальной иллюзии или наша неспособность запомнить ряд из более, чем 8 цифр. И наоборот, ошибки рассуждения часто приводят в замешательство – либо от того, что решение, которое мы были не способны найти, оказалось достаточно очевидным в ретроспективе, либо от того, что сделанная нами ошибка осталась привлекательной, хотя мы знаем, что это была ошибка. Многие текущие исследования суждений интересуются проблемами, которые имеют одну или другую из этих характеристик.

Наличие ошибки суждения демонстрируется сравнением ответов людей либо с установленным фактом (например, что две линии одинаковы по длине), либо при помощи принятого правила арифметики, логики или статистики. Однако, не каждый ответ, который оказывается противоречащим установленному факту или принятому правилу, является ошибкой суждения. Противоречие может исходить также от неправильного понимания вопроса испытуемым или от неправильной интерпретации ответа исследователем.

* Эта глава первоначально появилась в *Cognition*, 1982, 11, 123-141. Авторские права ©1981 принадлежат Elsevier Sequoia. Печатается в соответствии с разрешением.

Описание определенного ответа как ошибки суждения, следовательно, включает предположения о коммуникации между экспертом и испытуемым. (Мы вернемся к этому вопросу позже.) Изучающему суждение следует избегать чрезмерно строгих интерпретаций, которые трактуют разумные ответы как ошибки, равно как и снисходительных интерпретаций, которые пытаются рационализировать каждый ответ.

Хотя ошибки суждения являются еще и методом, посредством которого изучаются некоторые когнитивные процессы, этот метод стал значительной частью сообщения. Накопление демонстраций того, как умные люди нарушают элементарные правила логики или статистики, подняло сомнения относительно описательной адекватности рациональных моделей суждения и принятия решений. В течение двух десятилетий после Второй Мировой Войны несколько описательных трактовок реального поведения были основаны на нормативных моделях: теория субъективной ожидаемой пользы в анализах рискованного выбора, исчисления Байеса в исследовании изменений в убеждениях и теория обнаружения сигналов в исследованиях психофизиологических задач. Теоретический анализ этих ситуаций и, в гораздо меньшей степени, экспериментальных результатов, предложил образ людей как эффективных, почти оптимальных в принятии решений. На этом фоне наблюдения элементарных нарушений логических или статистических рассуждений кажутся неожиданными, и эта неожиданность, возможно, поддержала мнение о человеческом интеллекте, которое некоторые авторы критиковали как несправедливо негативное (Cohen, 1979, 1981; Edwards, 1975; Einhorn & Hogart, 1981).

Существуют три связанные причины фокусировки на систематических ошибках и предубеждениях вывода в исследовании рассуждений. Во-первых, они показывают некоторые наши ограничения и предлагают способы улучшения качества нашего мышления. Во-вторых, ошибки и предубеждения часто обнаруживают психологические процессы и эвристические процедуры, которые руководят суждением и выводом. В-третьих, ошибки и заблуждения помогают составить карту интуиции человека, указывая, какие из принципов статистики или логики являются не-интуитивными или контр-интуитивными.

Термины *интуиция* и *интуитивный* использовались в трех различных смыслах. Во-первых, суждение называется интуитивным, если оно достигнуто при помощи неформальной и неструктурированной модели рассуждений, без использования аналитических методов или обдуманых вычислений. Например, большинство психологов следуют интуитивной процедуре в принятии решений относительно размера своих выборок, но применяют аналитические процедуры, чтобы проверить статистическое значение своих результатов. Во-вторых, формальное правило или факт природы является интуитивным, если он совместим с нашей непрофессиональной моделью мира. Так, интуитивно очевидно, что вероятность выиграть приз в лотерею уменьшается с увеличением количества билетов, но контр-интуитив-

но, что есть хоть какой-то шанс, что группа из 23 человек будет включать двух человек, родившихся в один день. В-третьих, о правиле или процедуре говорят, что они являются частью нашего интуитивного репертуара, когда мы применяем правило или следуем процедуре в нашем обычном поведении. Правила грамматики, например, являются частью интуиции носителя языка, и некоторые (хотя и не все) из правил линейной геометрии являются включенными в наше пространственное мышление.

Настоящая глава обращается к некоторым методологическим и концептуальным проблемам, которые поднимаются при попытках составить карту интуиции людей относительно шанса и неопределенности. Мы начинаем с различных проверок статистической интуиции; затем мы обратимся к критике вопросно-ответной парадигмы в исследовании суждений; а закончим обсуждением не-интуитивного характера некоторых статистических законов.

Задачи на статистическую интуицию

Ошибки и предубеждения суждений в условиях неопределенности являются основным источником данных для очертания границ статистической интуиции людей. В этом контексте поучительно различать ошибки применения и ошибки понимания. Неудача в конкретной проблеме называется ошибкой применения, если есть свидетельство того, что люди знают и принимают правило, которое они не применили. Неудача называется ошибкой понимания, если люди не признают валидности правила, которое они нарушили.

Ошибка применения наиболее убедительно демонстрируется, когда человек самопроизвольно или с минимальной подсказкой хватается за голову и восклицает: “Как я мог это упустить?” Хотя многие читатели узнают этот опыт, подобное выражение эмоций не может идти в счет, и должны быть разработаны другие процедуры, демонстрирующие, что люди понимают правило, которое нарушили.

Понимание правила можно проверить (1) выявлением оснований для действий испытуемых или (2) просьбой подтвердить утверждение (1) общего правила или (2) довода за или против конкретного заключения. Комбинация этих характеристик дает четыре процедуры, которые мы сейчас проиллюстрируем и обсудим.

Мы начинаем с неформального примера, в котором понимание правила подтверждается принятием или подтверждением довода. Один из нас представил следующий вопрос игрокам в сквош.

Как вы знаете, играть в сквош можно либо до 9, либо до 15 очков. При сохранении всех остальных правил игры неизменными, если А – игрок лучший, чем В, какая из систем очков даст А лучший шанс на выигрыш?

Хотя все наши испытуемые обладали некоторыми знаниями по статистике, большинство из них сказала, что система очков не имела бы никакой разницы. Затем их попросили рассмотреть довод, что лучший игрок пред-

почел бы более долгую игру, поскольку атипичный исход менее вероятен в большой выборке, чем в маленькой. С очень малым количеством исключений респонденты немедленно приняли довод и отметили, что их первоначальный ответ был ошибкой. Очевидно, наши испытуемые имели какое-то представление о овоздействии размера выборки на ошибки осуществления выборки, но они не смогли раскодировать длину игры в сквош как пример размера выборки. Тот факт, что правильное заключение стало неизбежным, как только была установлена эта связь, показывает, что начальный ответ был ошибкой применения, а не понимания.

Более систематическая попытка диагностики природы ошибки была осуществлена в исследовании феномена, обозначенного как эффект конъюнкции (см. Главу 6). Возможно, самым элементарным принципом теории вероятности является правило конъюнкции, которое утверждает, что вероятность конъюнкции (A&B) не может превысить ни вероятность A, ни вероятность B. Однако, как показывает следующий пример, возможно сконструировать задачи, в которых большинство судей – даже в высокой степени искусственных – утверждает, что конъюнкция событий более вероятна, чем один из его компонентов.

Чтобы вызвать эффект конъюнкции, мы предоставили испытуемым описание людей, подобные помещенному ниже:

Линде 31 год, она не замужем, искренна и умна. Она занимается в основном философией. Будучи студенткой, она серьезно интересовалась вопросами дискриминации и социальной справедливости, а также участвовала в антиядерных демонстрациях.

По одному варианту проблемы, респондентов спрашивали, которое из двух утверждений о Линде было более вероятно: (A) Линда – кассир в банке; (B) Линда – кассир в банке, который активно участвует в феминистском движении. В большой выборке статистически неискушенных студентов последнего курса 86% вынесли суждение, что второе утверждение более вероятно. В выборке выпускников-психологов только 50% допустили эту ошибку. Однако разница между статистически неопытными и опытными респондентами исчезала, когда эти два критических вопроса были внедрены в список из восьми сравнительных утверждений о Линде. Более 80% обеих групп демонстрировали эффект объединения. Сходные результаты были получены внутригрупповом экспериментальном плане, по которому критические категории сравнивались косвенно (см. Главу 6).

Задачи подтверждения правила и подтверждения довода были использованы в попытке определения того, понимают и принимают ли люди правило конъюнкции. Сначала мы представили группе статистически необученных студентов колледжа несколько похожих на правила утверждений, которые они должны были классифицировать как правдивые и ложные. Утверждение: “Вероятность X всегда больше, чем вероятность X и Y” было подтверждено 81% респондентов. Для сравнения, только 6% подтвердило утверждение “Если A более вероятно, чем B, то они не могут произойти оба”.

Эти результаты показывают некоторое понимание правила конъюнкции, хотя подтверждение не единогласное, возможно, из-за абстрактной и необычной формулировки.

Также использовалась процедура подтверждения довода, в ходе которой респондентам дали описание Линды с последующими утверждениями (А) и (В) и попросили проверить, какой из следующих доводов они рассматривают как правильный:

(i) А более вероятно, чем В, потому что вероятность того, что Линда и кассир в банке и активная феминистка, должна быть меньше, чем вероятность того, что она кассир в банке.

(ii) В более вероятно, чем А, потому что Линда похожа на кассира в банке, активно участвующую в феминистском движении, больше, чем на кассира в банке.

Довод (i), благоприятствующий правилу конъюнкции, был подтвержден 83% аспирантов-психологов, но только 43% статистически неопытных студентов. Обширные обсуждения с респондентами подтвердили эту модель. Статистически обученные респонденты немедленно признали валидность правила конъюнкции. Необученные респонденты, с другой стороны, были гораздо менее впечатлены нормативными доводами, и многие остались при своих первоначальных ответах, которые не согласовывались с правилом конъюнкции.

К нашему удивлению, необученные испытуемые не глубоко понимаем правило конъюнкции; они были склонны подтверждать его абстрактно, но не тогда, когда оно противоречило сильному ощущению репрезентативности. С другой стороны, статистически обученные испытуемые признали валидность правила и были способны применить его в особенно ясной проблеме. Однако, статистическая искушенность не предотвращала появления эффекта конъюнкции в менее ясных вариантах той же проблемы. В терминах настоящей трактовки, эффект конъюнкции оказывается ошибкой применения, по крайней мере, для более обученных испытуемых. Для более полного обсуждения этого вопроса, см. Главу 6.

В попытке описать статистическую интуицию людей при различных уровнях обученности Нисбетт, Кранц, Джемсон и Фонг (Nisbett, Krantz, Jepson & Fong, Глава 32, этого издания) использовали процедуру извлечения, согласно которой респондентам требовалось оценить и оправдать определенные заключения и выводы, относящиеся к персонажам кратких рассказов. Исследователи наблюдали большие индивидуальные различия в понимании основных статистических принципов, которые высоко коррелировали с уровнем статистической обученности. Естественно, статистическая интуиция зависит от ума, опыта и образования. Как и в других формах знания, то, что интуитивно для эксперта, часто не-интуитивно для новичка (см., например, Larkin, McDermott, Simon & Simon, 1980). Несмотря на это, некоторые статистические результаты (например, совпадение дней рождения или смена “орлов” и “решек” в игре по подбрасыванию монет) остаются

контр-интуитивными даже для изучающих теорию вероятности (Feller, 1968, с. 85). Далее, существуют свидетельства, что ошибки (например, проигрыш азартного игрока), которые обычно совершаются неопытными респондентами, можно также обнаружить у статистически обученных респондентов при решении более тонких проблем (Tversky & Kahneman, 1971, 2).

Метод выявления причин был также использован (Evans & Wason, 1976; Wason & Evans, 1975) в исследованиях логической интуиции в хорошо известной задаче четырех карточек (Wason, 1966). По стандартному варианту этой задачи, экспериментатор предъявляет четыре карточки, показывая А, Т, 4 и 7, и просит испытуемых показать карточки, которые следует перевернуть, чтобы проверить правило: “Если карта имеет гласную букву с одной стороны, она имеет четное число с другой”. Правильный ответ – необходимо проверить карты А и 7, поскольку нечетное число на одной карте или гласная буква на второй опровергли бы правило. Будучи поразительно неспособными логически рассуждать, большинство испытуемых выбрали для осмотра карты А и 4. Нейзон и Эванс (Nason & Evans) исследовали различные варианты этой задачи и требовали, чтобы их испытуемые приводили причины или доводы своих решений по поводу того, осматривать или нет обратную сторону каждой из четырех карт. Исследователи сделали заключение, что доводы, которыми испытуемые оправдывали свои ответы, были всего лишь рационализацией, а не формулированием правил, которые действительно руководили их решениями.

О другом свидетельстве неадекватного понимания людьми правил верификации сообщали Вазон (Wason, 1969) и Вазон и Джонсон-Лэирд (Wason & Johnson-Laird, 1970). Чтобы обеспечить “терапию”, эти исследователи ставили испытуемых перед последствиями их суждений и обращали внимание испытуемых на непоследовательность их ответов. Эта процедура имела незначительное воздействие на последующее выполнение той же задачи. Взятые вместе, результаты наводили на мысль, что трудности людей в задаче верификации отражают недостаток понимания, а не применения.

Примеры, которые мы пока рассмотрели, включали подтверждение правил или доводов и выявление доводов, оправдывающих конкретный ответ. Мы не обсуждали процедуру описания респондентами релевантного правила, поскольку подобная проверка часто чрезмерно требовательна: мы можем приписать людям понимание правил, которые они не могут должным образом выразить.

Предпочитаемые процедуры для установления ошибок применения требуют сравнения ответов людей в конкретном случае с их суждением о релевантном правиле или доводе (McClelland & Rohrbaugh, 1978; Slovic & Tversky, 1974). Подтвердить ошибку применения можно также и при других планах исследований. Например, Хемилл, Вилсон и Нисбетт (Hamill, Wilson & Nisbett, 1980) показывали испытуемым видеозапись интервью, якобы взятое у тюремного надзирателя. Половине испытуемых было сказано, что мнения надзирателя (очень гуманного или довольно грубого) были

типичны для тюремного персонала, тогда как другим испытуемым сказали, что отношения надзирателя были атипичны, и что он был либо намного более, либо намного менее гуманен, чем большинство его коллег. Затем испытуемые оценивали типичные отношения тюремного персонала по разнообразным вопросам. Неожиданным результатом было то, что мнения, выраженные атипичным надзирателем, имели почти такое же влияние на обобщение, что и мнения, приписанные типичному члену группы. Очевидно, в этой модели суждения что-то не правильно, хотя невозможно описать любое конкретное суждение как ошибочное, и маловероятно, что многие из испытуемых осознали бы, что они не находились под влиянием информации о типичности надзирателя (Nisbett & Wilson, 1977). В этом случае и в других внутригрупповых исследованиях, явно благоразумно будет заключить, что была допущена ошибка применения, если межгрупповое сравнение выдает результат, который большинство людей рассматривает как непрочный.

Мы определили ошибку применения как ответ, нарушающий валидное правило, которое индивид понимает и принимает. Однако часто сложно определить природу ошибки, так как различные проверки понимания и принятия правила могут давать различные результаты. Более того, одно и то же правило может нарушаться в одном контексте проблемы и не нарушаться в другом. Задача верификации обеспечивает поразительный пример: испытуемые, которые неправильно проверили правило “Если карточка имеет гласную букву с одной стороны, она имеет четное число с другой” не имели никаких сложностей с проверкой формально эквивалентного правила: “Если письмо с печатью, то на нем пятицентовая марка” (см. Johnson-Laird, Legrenzi & Sonino-Legrenzi, 1972; Johnson-Laird & Wason, 1977; Wason & Shapiro, 1971).

Эти результаты иллюстрируют типичную модель в изучении рассуждений. Оказывается, что люди не обладают общим валидным правилом для проверки условных утверждений, иначе они решили бы задачу карточек. С другой стороны, они не слепы в отношении правильного правила, иначе они также не смогли бы справиться с задачей марки. Утверждение, что люди не обладают правильной интуицией, является, строго говоря, правильным, если под применением правила имеется в виду то, что ему всегда следуют. С другой стороны, это утверждение может вводить в заблуждение, поскольку оно может навести на мысль о более общем дефиците, чем тот, который фактически наблюдается.

Некоторые заключения ранних исследований репрезентативности, оказывается, имеют сходный статус. Было продемонстрировано, что многие взрослые не имеют общей валидной интуиции, соответствующей закону больших чисел, роли базовых значений в выводе Байеса или принципам регрессивного прогнозирования. Но не обязательно, что каждая задача, для которой они будут релевантны, будет решаться неправильно, или что правила не могут оказаться неустойчивыми в особых контекстах.

Свойства, которые делают формально эквивалентные проблемы легкими или трудными для решения, оказываются связанными с умственными моделями, или схемами, которые эти проблемы вызывают (Rumelhart, 1979). Например, оказывается легче увидеть релевантность “не- q ” в “ p подразумевает q ” в схеме контроля качества (Они забыли наклеить марку на письмо, скрепленное печатью?), чем в схеме подтверждения (Подразумевают ли отрицание заключения отрицание гипотезы?). Оказывается, что процесс действительного рассуждения схематично или содержательно ограничен таким образом, что различные операции или правила вывода доступны в различных контекстах (Hayes & Simon, 1977). Следовательно, рассуждение человека не может быть адекватно описано в терминах содержательно-независимых формальных правил.

Проблема выявления границ статистической или логической интуиции далее усложняется возможностью достичь в высокой степени непредвиденных заключений серией высоко интуитивных шагов. Именно этот метод с большим успехом использовал Сократ, чтобы убедить своих неопытных учеников в том, что они всегда знали истину, которую он лишь заставил их обнаружить. Следует ли любые заключения, которых можно достичь серией интуитивных шагов, рассматривать как интуитивные? Брейн (Braine, 1978) обсуждал этот вопрос в контексте дедуктивных рассуждений и предложил в качестве проверки непосредственность: утверждение интуитивно только тогда, когда его истинность является непосредственно данной, и если она подтверждается одним шагом.

Вопрос сократовых намеков точно не трактовался в контексте суждений в ситуациях неопределенности, и нет правил, которые отличают справедливые проверки интуиции от изобретательных загадок, с одной стороны, и сократовых инструкций – с другой. Представьте, например, как Сократ мог учить своего ученика давать правильный ответ на следующий вопрос:

“В какой больнице, – большой или маленькой, – будут чаще отмечаться дни, когда более 60% новорожденных будут мальчиками?”

Это сложный вопрос для студентов Стэнфорда (Kahneman & Tversky, 1972, с. 441, 3), но правильный ответ может быть извлечен в серии легких шагов, возможно таких:

“Не согласитесь ли вы, что дети, рожденные в конкретной больнице за конкретный день, могут рассматриваться как выборка?”

“Совершенно верно. А теперь, одинаково ли вы уверены в результатах большой выборки и маленькой?”

“В самом деле. Не согласитесь ли вы, что ваша уверенность больше в выборке, которая с меньшей вероятностью будет ошибочной?”

“Конечно, вы всегда это знали. Не скажете ли вы мне сейчас, какова пропорция мальчиков в совокупности детей, которую вы рассматриваете как ближайшую к идеалу истины?”

“Мы опять согласны. Разве это не означает тогда, что день, когда родилось более 60% мальчиков, значительно удален от идеального?”

“Таким образом, если вы уверены в выборке, разве вам не следует ожидать, что она скорее обнаружит правду, чем ошибку?” И т.д.

Сократова процедура – это очень неуклюжий способ направления респондента к желаемому ответу, но есть и более тонкие способы достижения той же цели. Фишхофф, Словик и Лихтенштейн (Fischhoff, Slovic и Lichtenstein, 1979) показали, что испытуемые становятся чувствительными к базовым значениям и надежности свидетельства, когда они сталкиваются с последовательными проблемами, отличающимися только по этим критическим переменным. Хотя эти исследователи не получили эффекта размера выборки даже внутригрупповом экспериментальном плане, подобные эффекты были получены Эвансом и Дюзуаром (Evans и Dusoir, 1977) и Бар-Хиллел (Bar-Hillel, 1979) с более ясной формулировкой и более предельными исходами по выборке.

Намек, обеспеченный параллельными проблемами, может привести к тому, что испытуемые присвоят переменной значение, являющееся иррелевантным правильному ответу: Фишхофф и Бар-Хиллел (1980а) продемонстрировали, что респонденты были чувствительны к иррелевантной информации о базовом значении, если это была лишь единственная переменная, различающая набор проблем. В самом деле, испытуемые склонны верить, что любое свойство данных, которое систематически изменяется, является релевантным правильному ответу. Внутригрупповые проекты связаны с многозначительными проблемами интерпретации в нескольких областях психологических исследований (Poulton, 1975). В исследованиях интуиции они обязаны вызывать эффект, который они намереваются проверить.

К вопросу об ограничениях вопросно-ответной парадигмы

В предыдущем разделе мы указали на возможность того, что внутригрупповые планы и сократовы намеки могут подталкивать исследуемой интуиции. В действительности проблема намного шире. Большинство исследований по суждениям в ситуации неопределенности и по индуктивным выводам проводилось в разговорной парадигме, при которой испытуемым предъявляется информация, и их просят ответить на вопросы или оценить значения, устно или письменно. В этом разделе мы обсудим некоторые сложности и ограничения, связанные с этой вопросно-ответной парадигмой.

Использование коротких вопросников, заполняемых случайно мотивированными испытуемыми, часто критикуется на том основании, что испытуемые действовали бы иначе, если бы отнеслись к ситуации более серьезно. Однако свидетельства показывают, что ошибки рассуждения и выбора, которые были первоначально установлены в гипотетических вопросах, не ликвидируются предъявлением более основательных стимулов (Grether, 1979; Grether & Plott, 1979; Lichtenstein & Slovic, 1971, 1973; Tversky & Kahneman, 1981). Гипотетические вопросы являются подходящими тогда,

когда люди способны предсказать, как бы они отвечали в более реалистичной обстановке, и когда у них нет никаких мотивов для лжи в своих ответах. Это не говорит о том, что поощрения и стимулы не воздействуют на суждения. Скорее, мы утверждаем, что ошибки рассуждения и выбора не исчезают с присутствием поощрений. Ни газеты, ни исследование прошлых политических и военных решений не поддерживают оптимистического взгляда, что рациональность преобладает, когда ставки высоки (Janis, 1972; Janis & Mann, 1977; Jervis, 1975).

Возможно, более серьезный интерес в отношении вопросно-ответной парадигмы состоит в том, что мы не можем безопасно предположить, что “экспериментальные разговоры”, в которых испытуемые получают сообщения и отвечают на вопросы, будут имитировать выводы, которые люди делают при обычном взаимодействии со средой. Хотя некоторые суждения в повседневной жизни являются ответом на точные вопросы, многие не являются таковыми. Более того, разговорные эксперименты во многом отличны от обычного социального взаимодействия.

Интерпретируя ответы испытуемых, экспериментаторы находятся в искушении предположить, (i) что вопросы только лишь выявляют у испытуемых открытые выражения мыслей, которые возникнут у них самопроизвольно, и (ii) что вся информация, данная испытуемому, включена в экспериментальное сообщение. С точки зрения испытуемого ситуация выглядит несколько иначе. Во-первых, вопрос, который задает экспериментатор, может возникать не самопроизвольно в ситуации, которую эксперт имитирует. Во-вторых, испытуемого обычно интересуют многие вопросы, которые экспериментатор никогда и не думал задавать, такие как: Существует ли правильный ответ на этот вопрос? Ожидает ли экспериментатор, что я его обнаружу? Вероятно ли вообще, что очевидный ответ правилен? Дает ли вопрос какие-либо намеки на ожидаемый ответ? Что определило отбор информации, предоставленной мне? Может ли часть ее быть иррелевантной, чтобы ввести меня в заблуждение, или вся она релевантна? Единственный открытый ответ, наблюдаемый экспериментатором, определяется частично ответами испытуемого на эту группу подразумеваемых вопросов. И экспериментальное сообщение является лишь одним из источников информации, которую использует испытуемый для генерации как открытых, так и скрытых ответов (Orne, 1973).

Вслед за лекциями Грайса Вильяма Джеймса в 1967 году (Grice, 1975), большой массив литературы по философии, лингвистике и психолингвистике имел дело с вкладом принципа кооперативности в значение изречений (для справок см. Clark & Clark, 1977). Согласно этому принципу, слушатель в разговоре имеет право предположить, что говорящий старается быть “информативным, правдивым, релевантным и понятным” (Clark & Clark, 1977, с. 560). Грайс составил список нескольких принципов, которым обычно будет следовать кооперативный говорящий. Например, принцип количества запрещает говорящему говорить то, что слушатель уже знает или мог бы легко понять из

контекста или оставшейся части сообщения. Именно по этому принципу утверждение “Джон пытался убрать в доме” выражает, что попытка была безуспешной: Слушатель может предположить, что успешная попытка была бы описана более простым предложением: “Джон убрал в доме”.

Испытуемые приходят на эксперимент с опытом кооперативности в разговоре, полученным в течение всей жизни. Они обычно ожидают увидеть кооперативного экспериментатора, хотя это ожидание часто ошибочно. Предположение кооперативности оказывает много еле уловимых эффектов на интерпретацию испытуемых предоставляемой им информации. В особенности, оно делает исключительно сложным для экспериментатора исследование эффектов “иррелевантной” информации. Поскольку представление иррелевантной информации нарушает правила разговора, вероятно, испытуемые ищут релевантность в любом экспериментальном сообщении. Например, Тейлор и Крокер (Taylor и Crocker, 1979) комментировали факт, что впечатления испытуемых о человеке находились под влиянием утверждений, которые истинны для всех, например, “Марк скромнен в отношениях со своими профессорами”. Но вывод испытуемых о том, что Марк необычайно скромнен, может быть оправдан убеждением, что кооперативный экспериментатор не включил бы полное избыточное утверждение в описание личности. Сходные вопросы поднимаются в других исследованиях (например, Kahneman & Tversky, 1973; Nisbett, Zukier & Lemley, 1981), которые исследовали воздействие иррелевантной и ненужной информации.

Роль предположений, заложенных в вопросе, была проиллюстрирована Лофтусом и Пальмером (Loftus и Palmer, 1974), которые показали, что свидетели давали более высокую оценку скорости машины, когда их спрашивали “Как быстро ехала машина, когда она разбила вдребезги другую машину?”, чем когда вопрос был “Как быстро ехала машина, когда она ударила другую машину?”. Использование в вопросе слова *разбила* подразумевает, что задающий вопрос, если он откровенен и кооперативен, убежден, что машина ехала быстро.

Нормативный анализ подобного вывода может быть разделен на две стадии проблемы. (i) Должен ли свидетель находиться под влиянием вопроса, формируя личное мнение относительно скорости машины? (ii) Должен ли свидетель находиться под влиянием вопроса, формулируя общественную оценку? Ответ на (i) должен быть положительным, если вопрос выражает новую информацию. Ответ на (ii) менее ясен. С одной стороны, кажется неподходящим для ответа, если он вторит информации, содержащейся в вопросе. С другой стороны, от кооперативного свидетеля ожидают, что он даст наилучшую возможную оценку, отвечая на вопрос о количестве. Что же делать свидетелю, если на эту оценку только что повлиял вопрос? Следует ли ему ответить: “Перед тем, как вы спросили меня, я думал, что ...”? Какими бы ни были нормативные достоинства случая, свидетельства показывают, что люди часто неспособны изолировать прошлые мнения от настоящих или оценить вес факторов, оказавших воздействие на их взгляды (Fischhoff,

1977b; Goethals & Reckman, 1973; Nisbett & Wilson, 1977; Ross & Lepper, 1980).

Наше исследование привязки (Tversky & Kahneman, 1974, 1) далее иллюстрирует силу едва уловимых предположений. В одном исследовании мы просили группу испытуемых оценить вероятность того, что население Турции было больше, чем 5 млн, а другую группу мы попросили оценить вероятность того, что население Турции было меньше, чем 65 млн. Следуя этому заданию, две группы записали свои наилучшие догадки относительно населения Турции; средние оценки были 17 и 35 млн, соответственно для групп, которым представили низкие и высокие привязки. Эти ответы могут быть также рационализированы предположением, что значения, которые появляются в вопросах о вероятности, не слишком далеки от правильных.

Мы доказывали, что эффекты предположения иногда могут быть оправданы, поскольку нет ясного разграничения между предположением и информацией. Однако важно отметить, что люди не принимают предположения, *потому что* так нужно. Во-первых, они обычно не знают, что на них повлияло предположение (Loftus, 1979; Nisbett & Wilson, 1977). Во-вторых, сходные эффекты предположения наблюдаются даже тогда, когда респонденты не могут основательно верить, что привязка, данная им, несет информацию. Испытуемые, которым требовалось дать оценки количества, регулируя вверх или вниз от случайно генерированного значения, демонстрировали сильное проявление эффектов привязки (Tversky & Kahneman, 1974, 1). Но неспособность предполагать, как таковая, является беспокойством, а явная неспособность отвергать – неинформативные сообщения.

Когда испытуемым требуется указать свой ответ, выбрав ответ из списка или построив распределение вероятности по данному набору альтернатив, выбор категорий экспериментатором может быть информативным. Лофтус (Loftus, 1979) показал, что респонденты сообщали о гораздо большем количестве случаев головной боли в неделю, когда шкала ответов была выражена как 1-5, 5-10, 10-15 и т.д., чем когда шкала была выражена как 1-3, 3-5, 5-7 и т.д. В этом случае шкала могла законно влиять на границы того, что должны называть головной болью. Даже когда такие повторные интерпретации не являются возможными, можно ожидать, что испытуемые предпочтут среднюю часть области распространения в своих оценках количества и построят субъективное вероятностное распределение, в котором каждая категория определена неявной вероятностью (Olson, 1976; Parducci, 1965).

Предположения, подразумеваемые опросником, могут также вносить вклад в результаты, наблюдавшиеся Фишхоффом, Словиком и Лихтенштейном (1978), которые просили неопытных испытуемых и опытных автомехаников оценить вероятность различных неполадок, которые могут быть причиной нарушений в запуске машины. Они обнаружили, что оцененная вероятность категории “все другие проблемы” была достаточно нечувствительна к полноте списка и едва увеличилась, когда основной фактор (например, вся электрическая система) был удален из списка.

Даже едва уловимые и косвенные ключи к разгадке могут быть эффективны. В недавнем исследовании мы дали испытуемым следующую информацию: “Мистер А – кавказец, 33 лет. Он весит 190 фунтов”. Одну группу испытуемых попросили угадать его рост. Другие испытуемые также пытались догадаться о его росте, после того, как сначала пытались догадаться о размере его талии. Средняя оценка была значительно выше в первой группе, примерно на один дюйм. Мы подозреваем, что испытуемые, сначала оценивавшие размер талии приписали большую часть веса мистера А его обхвату, чем это сделали испытуемые, которые только лишь угадывали его рост.

Мы заключаем, что разговорный аспект исследований суждений заслуживает более тщательного рассмотрения, чем он получил в прошлых исследованиях, включая наше собственное. Мы не можем всегда предполагать, что люди сделают или должны сделать тот же вывод, наблюдая факт и слушая сообщение об этом же факте, потому что правила разговора, которые регулируют коммуникацию между людьми, не применимы к информации, которая получена из естественного наблюдения. Часто трудно задавать вопросы, не давая (полезных или вводящих в заблуждение) ключей относительно правильного ответа и не выражая информации об ожидаемом ответе. Обсуждение родственного нормативного вопроса, касающегося интерпретации-свидетельства, включено в исследование Бар-Хиллел и Фолк (Bar-Hillel и Falk, 1980).

Естественно, вероятно, что факторы предубеждения, упомянутые нами, будут оказывать наибольшее воздействие в ситуациях высокой неопределенности. Испытуемые не будут придавать много значения интерпретациям разговорного отношения экспериментатора, если они будут противоречить уверенному знанию правильного ответа на вопрос. Однако, в области, где было проведено большинство исследований суждений, изменения разговорного контекста могут воздействовать на процесс рассуждения так же, как и на наблюдаемый ответ.

Ошибки суждения: положительный и отрицательный анализ

Часто полезно различать положительную и отрицательную стороны ошибок суждения. Положительный анализ концентрируется на факторах, которые производили частично неправильный ответ; отрицательный анализ объясняет, почему не был дан правильный ответ. Например, положительный анализ ошибки ребенка в задаче Пиаже пытается обозначить факторы, которые определяют ответ ребенка, такие как относительная высота или площадь поверхности двух контейнеров. Отрицательный анализ того же поведения сфокусировался бы на препятствиях, которые помешали ребенку получить и понять сохранение объема. В исследовании суждений в ситуациях неопределенности положительный анализ касается эвристики, которую используют люди в суждениях, оценках и предсказаниях. Отрицательный анализ связан со сложностями понимания и применения элементар-

ных правил рассуждения. В случае ошибки понимания отрицательный анализ фокусируется на препятствиях, которые мешают людям открыть релевантное правило самим или принять простые его объяснения. Отрицательный анализ ошибки применения стремится идентифицировать способы, при которых кодирование проблем может скрывать релевантность правила, которое известно и принято.

В общем, положительный анализ ошибки наиболее полезен тогда, когда одна эвристика объясняет суждения в разнящемся наборе проблем, где нарушаются различные нормативные правила. Соответственно, отрицательный анализ будет наиболее показательным, когда люди последовательно нарушают правило в различных проблемах, но совершают ошибки, которые нельзя отнести к одной эвристике. Тогда становится уместным вопрос, почему люди не могут выгучить правило, если рутинные наблюдения повседневных событий предлагает достаточно возможностей для этого. Также становится уместным вопрос, почему люди сопротивляются правилу, если их не убеждают простые валидные доводы. Сложности изучения статистических принципов на повседневном опыте обсуждались несколькими авторами, наиболее полно Айнхорном, Хогартом (Einhorn и Hogart, 1978), Л.Р.Голдбергом (L. R. Goldberg, 1968b), Нисбеттом и Россом (Nisbett и Ross, 1980). Неудачи научения обычно исходят из недоступности необходимого кодирования релевантных примеров или к отсутствию корректирующей обратной связи по ошибочным суждениям. Сопротивление принятию правила обычно приписывается его не-интуитивной природе. В качестве примера мы обратимся теперь к анализу причин сопротивления правилу регрессивного прогнозирования.

Исследования интуитивного прогнозирования обеспечили много свидетельств преобладания тенденции делать прогнозы, которые являются радикальными или недостаточно регрессивными. (Недавний обзор этой литературы см. в Главе 15). В более ранних статьях мы предлагали положительный анализ этого эффекта как проявление репрезентативной эвристики (Kahneman & Tversky, 1973, 4; 1979a, 30). Однако, как мы увидим, есть основания для того, чтобы обратиться к отрицательному анализу для более понятной трактовки.

Отрицательный анализ представляет особый интерес в случае с ошибками понимания, при которых люди находят правильное правило не-интуитивным или даже контр-интуитивным. Как засвидетельствует большинство преподавателей элементарной статистики, студенты находят концепцию регрессии очень сложной для понимания и применения, несмотря на жизненный опыт, в котором крайние прогнозы было часто слишком крайними. Спортивные тренеры и учителя, например, знакомы с проявлениями регрессии к среднему: за исключительными достижениями чаще идет разочарование, а за неудачами – улучшение.

Более того, когда регрессия критериальной переменной по предсказывающей переменной в действительности линейна, и когда условное распреде-

ление критерия (для фиксированных значений предсказывающей переменной) симметрично, правило регрессивного прогнозирования может быть защищено непреодолимым доводом: имеет смысл сделать одинаковый прогноз для всех случаев, которые имеют одинаковое значение предсказывающей переменной, и имеет смысл выбрать этот прогноз так, чтобы средняя величина и медиана критериального значения, для всех случаев, которые делят одинаковое предсказанное значение Y , равнялось Y . Это правило, однако, противостоит другим видам интуиции, некоторые из которых обсуждаются в следующих абзацах.

1. “Оптимальное правило прогнозирования должно, по крайней мере, позволять, если не гарантировать, совершенно точное прогнозирование для целого набора случаев”. Принцип регрессивного прогнозирования нарушает это, по-видимому, благоразумное требование. Он выдает набор предсказанных значений, который имеет меньшее различие, чем соответствующий набор действительных значений критерия и, таким образом, исключает возможность набора совершенно точных прогнозов. В самом деле, правило регрессии гарантирует, что ошибка будет допускаться в каждой паре коррелирующих наблюдений: мы никогда не сможем найти сына, чей рост был правильно предсказан по росту его отца, и также позволял точное прогнозирование роста отца, за исключением, когда оба значения находятся в середине распределения роста. Странно, что правило прогнозирования, гарантирующее ошибку, должно превратиться в оптимальное.

2. “Связь между наблюдением и основанным на нем прогнозированием должна быть симметрична”. По-видимому, благоразумно ожидать, что, если B предсказано по знаниям об A , тогда A должно быть соответствующим прогнозом, когда известно B . Регрессивное прогнозирование, конечно же, нарушает эту симметрию, поскольку прогнозирование двух переменных друг от друга регулируется не одним и тем же уравнением регрессии. Родственная асимметрия встречается при сравнении регрессивного прогнозирования с действительными значениями критериальной переменной. Регрессивное прогнозирование не подвержено предубеждению, в том смысле, что среднее критериальное значение, во всех случаях, для которых было предсказано частное значение Y , ожидается как Y . Однако, если мы рассмотрим все случаи, для которых критериальное значение было Y , окажется, что среднее значение их предсказанных показателей будет лежать между Y и групповым средним значением. Эти асимметрии являются загадочными и контринтуитивными для сообразительных, но статистически неопытных людей.

Асимметрии регрессивного прогнозирования особенно проблематичны, когда первоначальное наблюдение и критерий генерированы одним и тем же процессом и априорно не отличаются, как в случае повторного осуществления выборки из той же совокупности или в случае параллельных форм одного теста. Единственной моделью прогнозирования, которая удовлетворяет симметрии в подобных ситуациях, является правило отождествления, при котором показатели по второй форме предсказываются как такие же,

что и при первоначальном наблюдении. Принцип регрессивного прогнозирования представляет различие, для которого нет никаких очевидных оснований: как можно предсказать знак разницы между двумя значениями, выбранными из одной совокупности, как только становится известным одно из этих значений?

3. “Любой системный эффект должен иметь причину”. Разница между первоначальными наблюдениями и действительными критериальными значениями является фактом, который можно наблюдать при любом типе рассеивания. Однако он оказывается эффектом без причины. Например, в ситуации повторного теста знание о том, что первоначальные показатели были высоки, влечет за собой прогнозирование, что последующие будут ниже, но первое наблюдение не является причиной того, что второе будет ниже. Появление беспричинного эффекта нарушает вескую интуицию. В самом деле, пониманию регрессии сильно мешает тот факт, что любому примеру регрессии, с которым случайно сталкиваешься, вероятно, будет дано причинное объяснение. Например, в контексте умелого выполнения регрессия от первоначального теста к последующему обычно приписывается напряженному усердию после первой неудачи и чрезмерной уверенности, следующей за первоначальным успехом. Часто трудно осознать, что выполняющие будут регрессировать, даже не зная результатов, а только из-за непреодолимой ненадежности своего выполнения. Регрессия первого выполнения по второму также удивительна, потому что ей нельзя дать простого причинного объяснения.

Мы сделали набросок отрицательного анализа сложностей, которые испытывают люди в понимании и применении концепции регрессивного прогнозирования. Мы предполагаем, что у людей есть сильная интуиция относительно статистического прогнозирования, и что некоторые нормативно правильные принципы являются контр-интуитивными именно потому, что они нарушают их существующую интуицию. С этой точки зрения, “принципы”, которые люди принимают, представляют значительные убеждения, а не только лишь рационализации, и они играют существенную роль в замедлении изучения верных правил. Однако эти убеждения часто являются противоречивыми и, следовательно, непонятными. Например, невозможно построить не-выраженное совместное распределение роста отцов и (первых) сыновей так, чтобы средний рост отца был бы объективным предсказывающим фактором роста сына, а рост сына был бы объективным предсказывающим фактором роста его отца.

В заключении, мы предположили, что некоторые ошибки и предубеждения в суждениях в ситуациях неопределенности требуют двойственного анализа: положительной стороны, которая объясняет выбор частного ошибочного ответа в терминах эвристики, и отрицательной стороны, объясняющей, почему правильное правило не было выучено. Хотя эти два анализа не являются несовместимыми, они имеют тенденцию придавать большое значение различным аспектам изучаемого феномена. Попытка интегрировать

положительные и отрицательные стороны, похоже, обогатит теоретический анализ индуктивного рассуждения.

Резюме

В этой главе мы обратились к трем группам методологических и концептуальных проблем в области суждений в ситуациях неопределенности. Во-первых, мы сделали различие между ошибками применения и ошибками понимания и обсудили различные методы изучения статистической интуиции. Во-вторых, мы сделали обзор некоторых ограничений вопросно-ответной парадигмы исследования суждений и исследовали эффекты мысленных предположений, сократовых намеков и правил разговора. В-третьих, мы обсудили роль положительных и отрицательных объяснений ошибок суждения.

Соображения, выдвинутые в этой главе, усложняют эмпирический и теоретический анализ суждений в ситуациях неопределенности; они также предлагают новые направления будущих исследований. Мы надеемся, что более глубокая оценка концептуальных и методологических проблем, связанных с изучением статистической интуиции, приведет к лучшему пониманию сложностей, тонкостей и ограничений индуктивного рассуждения человека.

35. Варианты неопределенности*

Даниель Канеман и Амос Тверски

Понятие неопределенности в философии, статистике и теории принятия решений обычно трактуют для всех форм неопределенности в терминах единичного измерения вероятности или степени убежденности. Недавние психологические исследования принятия решений в условиях неопределенности часто следовали этой традиции и фокусировались на соответствии интуитивных суждений стандартной логике вероятности (Einhorn & Hogart, 1981; Nisbett & Ross, 1980; Slovic, Fischhoff & Lichtenstein, 1977; Глава 34). Всесторонняя психологическая перспектива неопределенности, однако, открывает разнообразие процессов и опыта, которые включают такие основные механизмы, как привыкание к неоднократному стимулированию в отдельном нейроне, и такие сложные виды деятельности, как оценка научных гипотез.

В этой главе мы сделаем набросок из ряда наблюдений, которые обычно рассматриваются при психологическом анализе принятия решений в условиях неопределенности. Здесь будут обсуждаться два уровня ответов на неопределенность. Сначала мы обсудим некоторые основные процессы ожидания и удивления в восприятии, которые могут быть рассмотрены как предвестники субъективной вероятности. Затем мы обратимся к феноменологическому рассмотрению, в котором мы отличаем внутренние от внешних атрибуций неопределенности и кратко опишем четыре модели суждения, которые могут использовать люди, оценивая неопределенность.

Элементарные формы вероятности

Неопределенность – это факт, с которым все формы жизни должны быть готовы бороться. На всех уровнях биологической сложности существует неопределенность относительно значения знаков или стимулов и относитель-

* Эта глава впервые появилась в *Cognition*, 1982, 11. 143-157. Авторские права © 1981 принадлежат Elsevier Sequoia. Печатается в соответствии с разрешением.

но возможных последствий действий. На всех уровнях действие должно предприниматься до того, как прояснена неопределенность, и должен быть достигнут надлежащий баланс между высшим уровнем специфической готовности к событиям, которые произойдут с наибольшей вероятностью, и общей способностью отвечать соответствующим образом, когда случается непредвиденное. Поскольку фокус настоящей трактовки скорее находится на убеждении, чем на действии, мы не будем обсуждать замечательные процессы, когда низшие организмы распределяют свое ответное усилие в соответствии с вероятностью подкрепления (Herrnstein, 1970). Наш главный интерес в этом разделе относится к неопределенности восприятия.

Перцептивные ожидания

Перед событием существуют ожидания. После него может быть удивление. Удивление, главным образом, изучалось психофизиологическими методами, и было измерено разнообразными индикаторами ориентировочного ответа (Lynn, 1966; Sokolov 1969) и компонентом P300 потенциалов, связанным с событием (Duncan-Johnson & Donchin, 1977; Donchin, Ritter & McCallum, 1978). Ожидания изучались во многих контекстах и широким разнообразием методов.

Наше обсуждение перцептивных ожиданий будет организовано по схеме, показанной на Рисунке 1, которая различает три основных типа ожиданий. Первое основное различие отделяет активные от пассивных ожиданий: активное ожидание занимает сознание и ограниченный объем внимания; напротив, пассивное ожидание автоматически и не требует усилий, и лучше описывается как диспозиция, чем как деятельность (Posner, 1978).

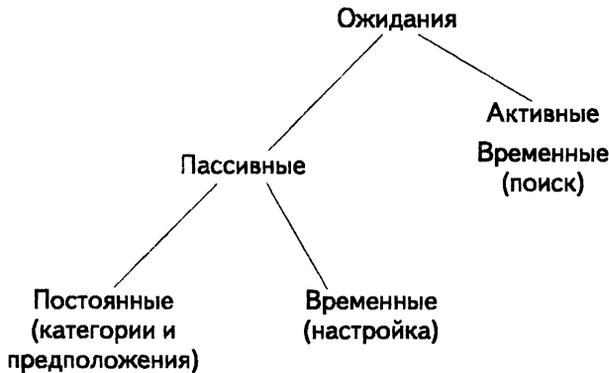


Рис. 1. Перцептивные ожидания

Некоторые ожидания относительно постоянны. Длительные ожидания о ковариациях признаков определяют категории восприятия, которые мы используем, чтобы организовать и закодировать опыт (Broadbent, 1971). Специфические ожидания об объектах, например, что комнаты и окна, вероятно, будут прямоугольными, функционируют, как постоянные предположения, которые помогают определять интерпретацию неясных стимулов (Ittelson & Kilpatrick, 1951). Мы хронически лучше подготовлены к одним событиям, чем к другим, как проиллюстрировало стойкое влияние прошлой частоты на порог узнавания слов (Broadbent, 1967; Morton, 1969). В самом деле, ожидания иногда дают галлюцинаторный опыт, который люди не могут отличить от действительного, как при эффекте фонематического восстановления. Так, вся сенсорная информация, соответствующая “s” в слове “legislature”, может быть удалена из записи слова и заменена кашлем или каким-либо другим естественным звуком. Испытуемые, которым представляют эту запись, предельно убеждены, что они слышат иллюзорную фонему (Warren, 1970).

Пассивные и временные ожидания опосредуют огромные эффекты контекста при узнавании (Foss & Blank, 1980) и несколько вариантов эффекта настройки (Posner, 1978). Например, включение буквы в предупреждающий сигнал, который представляет испытание, облегчает ответ на эту букву в ускоренной задаче на соответствие, даже когда случайности устроены так, что предупреждающий сигнал не выражает никакой валидной информации о цели. Познер (Posner, 1978) задокументировал некоторые важные различия между пассивным ожиданием, установленным неинформативным предупреждающим сигналом, и активным ожиданием, выработанным, когда цель фактически предсказуема, хотя и неполно, по этому сигналу. Пассивное ожидание дает выгоду (например, более быстрый ответ), когда оно подтверждается, но оно не препятствует ответу по целям, которые не были настроены. Напротив, сигнал, который является причиной того, что испытуемый готовится к определенной цели, также снижает ответ по непредвиденным целям. На языке теории вероятности, активное ожидание придерживается правила комплементарности: высокая степень готовности к определенному событию достигается за счет недостаточной подготовки к другим событиям. Пассивная настройка связана с не-комплементарной моделью выгод без потерь.

Пассивное ожидание и сознательное предвидение могут конфликтовать, и есть свидетельства того, что пассивный процесс оказывает большее влияние на интерпретацию неясных стимулов. Эпштейн и Рок (Epstein и Rock, 1960) противопоставляли два типа ожиданий, используя рисунок, в котором левосторонний и правосторонний профили были объединены, чтобы сформировать модель обратимой фигуро-фоновой организации. Наблюдатели составного рисунка видели только один из профилей, который обозначал общий контур. Построив два профиля, которые можно было объединить таким образом, Эпштейн и Рок представили профили отдельно, регулярно

чередую в ряде проб, создав сознательное ожидание, что за одним всегда будет следовать другой. Затем в первый раз был представлен составной рисунок, и отмечалось, какое лицо видели в нем испытуемые. В соответствии с эффектом настройки, наблюдатели почти всегда видели профиль, который был показан в предыдущей пробе, а не тот, который они сознательно ожидали увидеть.

Также сообщалось об аналогичной демонстрации конфликта между различными уровнями ожидания, в которых компонент Р300 электроэнцефалограммы (ЭЭГ) был главной зависимой переменной. Р300 – это положительное отклонение в ЭЭГ, которое происходит где-то через 300 мсек после представления любого стимула, который наблюдатель трактует как значительный или релевантный задаче. Многие подробные исследования демонстрируют тесную связь между предшествующей вероятностью событий и величиной отклонений Р300, которую она вызывает (Donchin, Ritter & McCallum, 1978). Когда испытуемому предъявляют серию Бернулли, часто повторяемые события вызывают у него меньший Р300, чем редкие. Более того, ряд повторений одного и того же события связан со стойким уменьшением Р300, предполагая дальнейшее увеличение субъективной вероятности дальнейших повторений. Наоборот, сознательное ожидание повторений последовательно уменьшается во время долгого ряда, по хорошо известному заблуждению азартного игрока. Очевидно, наблюдатель может быть подготовленным, или “настроенным” для одного события, сознательно ожидая другое – и может показать физиологическое свидетельство удивления на событие, которое было сознательно предсказано. Таким образом, есть ощущение, по которому индивид в некотором смысле может иметь конфликтующие вероятности для одного и того же события в одно и то же время. Эти наблюдения предполагают образ ума, подобный бюрократии (Dennet, 1979), в котором различные части имеют доступ к различным данным, придают им различное значение и придерживаются различных взглядов на ситуацию.

Восприятие как пари

Ожидания, которые выработались в течение жизненного визуального опыта, имеют глубокое влияние на восприятие и поразительно недоступны сознательному знанию или намерению. Наиболее известные демонстрации этих фактов были разработаны транзакционалистами, изучающими восприятие (Ittelson & Kilpatrick, 1951; Kilpatrick, 1961). Наблюдателей известной искаженной комнаты и вращающегося окна вынуждают приобретать визуальный опыт, который противоречит как их общим знаниям, так и их специфическому знакомству с объектами иллюзии. Так, друзья могут показаться человеку великанами или крошечными существами, которые изменяют свой размер, продвигаясь вдоль стены искаженной комнаты, а бумажная салфетка разрезать вращающееся окно. Эти поразительные эффекты обеспечиваются за счет господствующего предположения, что комнаты

и окна прямоугольны. Хотя наблюдатель отлично знает, что это предположение неприменимо в данном случае, это знание не имеет никакого значительного эффекта на сознательное восприятие. Модели действительности, которые строились годами, не могут быть исправлены по требованию для частного случая. Эти наблюдения снова подтверждают, что наблюдатель может одновременно иметь конфликтующие взгляды на одно и то же событие.

Мы отмечали, что перцептивные ожидания определяют то, что мы “видим” в неясных стимулах. В самом деле, трансакционалисты интерпретировали восприятие, как пари на реальность (Kilpatrick, 1961). Значительным аспектом таких перцептивных выборов является сильная приверженность выбранной интерпретации. В нашем опыте не содержится никаких показателей двусмысленности стимулов, и даже если перцептивные интерпретации колеблются с течением времени, как с кубом Некера, они стремятся быть достаточно определенными в любой конкретный момент. Подавление неопределенности и двусмысленности в восприятии наводит на мысль, что мы можем быть биологически запрограммированы, чтобы действовать по наилучшему перцептивному пари, как если бы это пари не включало бы никакого риска ошибки. Значительное различие между сознательным опытом восприятия и мыслью состоит в том, что последняя может представлять сомнение и неопределенность, тогда как первое обычно нет.

Хотя подавление неопределенности отличает перцептивные пари от сознательного суждения о неопределенных событиях, обработка неопределенности на этих двух уровнях может быть сходной в других отношениях. Два поразительных наблюдения трансакционалистского исследования предлагают гипотезы, которые оказываются применимы к сознательным убеждениям. Первая состоит в том, что перестроенный образ среды имеет тенденцию быть внутренним, отражая обычные влияния и зависимости среди признаков обстановки и стимулов. Так, если объект представлен в условиях, которые делают неясными как его размер, так и расстояние до него, выбранная перцептивная интерпретация изберет размер и расстояние, которые относятся к размеру на сетчатке глаза стандартным образом: если объект воспринимается большим, тогда он также оказывается удаленнее, чем если бы он казался маленьким (Ittelson & Kilpatrick, 1951).

Второе наблюдение состоит в том, перцептивная конструкция оказывается иерархическим процессом, в котором решения о глобальных признаках обстановки принуждают и доминируют над решениями в отношении объектов, содержащихся в ней. Искаженная комната обеспечивает наилучший пример. То, что мы видим – это не компромисс между двумя крайними взглядами: люди нормального роста в искаженной комнате или люди необычного роста в нормальной комнате. Последний взгляд просто преобладает над первым, как если бы форма комнаты была вычислена перед тем, как началась обработка людей в ней. Можно ли будет показать, что сходные правила действуют, например, в построении сценариев будущих событий – это проблема, которая заслуживает изучения.

Феноменология неопределенности

Предыдущий раздел пытался показать, что правила, управляющие перцептивными ожиданиями, отличаются от правил теории вероятности. Настоящий раздел расширяет этот анализ до опыта сомнений и неопределенности, который, предположительно, отражают суждения субъективной вероятности. Как мы увидим, понятие вероятности относится в естественном языке к нескольким различным состояниям сознания, к которым правила стандартного исчисления вероятности не могут быть применимы в равной степени.

Чтобы оценить сложность ожиданий, рассмотрим одно из их проявлений: удивление, которое мы переживаем, когда ожидание нарушается. Представьте, что монету нужно подбросить 40 раз. Какое количество “решек” вы ожидаете? Если вы предположите, что монета без дефектов, вы, вероятно, заявите, что результат 20-20 более вероятен, чем какой либо другой, и все же вы будете более удивлены этим исходом, чем результатом 22 “решек” и 18 “орлов”. Показана ли “истинная” субъективная вероятность этих двух событий рассмотренным суждением об их относительной вероятности или невольной реакцией удивления, которую они вызовут?

Одной возможной интерпретацией является то, что пример иллюстрирует конфликт между двумя подходами к суждению о вероятности: суждение о том, что наиболее вероятный результат 20-20, происходит от знания правил шансов, но такие исходы, как 22-18 или 17-23, более вероятно, находясь на другом уровне, где вероятность определяется репрезентативностью. Немного неравный исход представляет как правильность монеты, так и случайность подбрасывания, которое вовсе не представлено равным результатом. С этой позиции, большая психологическая реальность ожиданий, основанная на репрезентативности, проявляет себя в реакции удивления.

Возможна несколько иная интерпретация, которая фокусируется на кодировании возможных исходов. Как мы увидим, часто в разговоре является подходящим расширить определение X до “X или что-то вроде этого”. Если самопроизвольное кодирование событий следует подобным правилам, результаты, подобные 22-18 или 17-23, будут самопроизвольно закодированы как “почти равное разделение”, тогда как результату 20-20 присвоят отличительный код “точно равное разделение”. Человек, который пытается судить об относительной вероятности событий, рассмотрит точные утверждения о результатах и отметит, что 20-20 более вероятно, чем, скажем, 22-18. Но реакция удивления может быть определена естественным кодированием событий. Событие 22-18 будет тогда относительно неувидительным, потому что оно закодировано как приблизительно равный результат, который, в самом деле, более вероятен, чем точно равный результат.

Роль кодирования событий очевидна в интерпретации неопределенных утверждений, таких как “Я оцениваю, что...” или иногда “Я думаю, что...”. Неопределенные утверждения являются классом речевых актов, который характеризуется специфическими условиями искренности и проверками

валидности. Рассмотрим, например, прогноз: “Я думаю, что цена на золото будет выше на 50% через 6 месяцев, чем сейчас”. Взятое буквально, это точечное прогнозирование, которому следует присвоить очень маленькую вероятность подтверждения. Но прогнозирование не предназначено для того, чтобы его понимали буквально. Точечные прогнозы обычно понимаются как сравнительные утверждения или утверждения об области распространения, в которую, как ожидается, попадет результат, например, “Я думаю, что повышение цены на золото будет ближе к 50%, чем к X% или Y%”. Говорящий и слушающий обычно ожидают согласия по мысленно подразумеваемым значениям X и Y. Например, прогнозирующий, цитируемый выше, будет считаться замечательно точным, если цена на золото в действительности поднимется на 53% за следующие 6 месяцев, хотя прогноз не был строго точным. Таким образом, говорящий, который отстаивает численное прогнозирование, скорее придерживается области распространения, чем конкретной точки в этой области. Говорящий также придерживается предположения, что значение в равной степени вероятно быть как выше оценки, так и ниже ее, исключая случаи, когда характер прогноза делает это невозможным. Так, человек, который говорит “Я думаю, что цена на золото поднимется на 50% за следующие 6 месяцев”, будет считаться намеренно вводящим в заблуждение, если он также думал, но не говорил, что действительное значение было бы более вероятно выше этой оценки, чем ниже ее.

Значительно то, что условия искренности, связанные с прогнозированием, не требуют, чтобы предсказываемые значения (или область распространения) считались высоко вероятными, но лишь чтобы они считались более вероятными, чем сравнительные значения (или область распространения). Например, мужчина, который утверждает “Я думаю, что Билли Джон выиграет золотую медаль по прыжкам в высоту на следующих Олимпийских Играх”, не будет считаться лгуном, если он предпочтет поставить против этого предположения, чем на него, но ему запрещено добавлять “а шансы Джека Смолла даже еще выше”. Таким образом, естественный язык позволяет играть привилегированную роль наилучшей догадке, а идентификация благоприятной догадки выражает информацию об альтернативах, с которыми она может справедливо сравниваться. Упоминание любимого атлета показывает, что он должен сравниваться с другими атлетами, а не с дизъюнкцией возможных победителей. Одним последствием этого правила является то, что иногда можно “предсказать” событие, которое рассматривается как менее вероятное, чем его дополнение, если дополнение естественно закодировано как дизъюнкция.

Сходное ограничение относится к выражениям уверенности. Утверждение уверенности выражает неопределенность человека в прогнозировании, оценке или выводе, которые он уже сделал. Так, будет естественным спросить “Насколько вы уверены в том, что вы правы?”, но будет аномальным спросить “Насколько вы уверены в том, что вы ошибаетесь?”. Уверенность – это субъективная вероятность или степень убеждения, связанная с тем, что, как мы “думаем”, случится.

Обычный язык также обеспечивает огромное количество выражений, чтобы говорить о событиях, которые могут произойти, хотя нам не обязательно “думать”, что они произойдут. Так, люди оценивают шансы кандидатов, риск различных видов деятельности, дают шансовые результаты футбольных матчей и понимают утверждения синоптиков о вероятности дождя. Теперь мы обратимся к более подробному анализу состояний неопределенности, которые могут выражать подобные утверждения, следуя схеме, показанной на Рисунке 2. Два уровня рисунка, атрибуции неопределенности и варианты неопределенности, обсуждаются в следующих разделах.

Атрибуция неопределенности

Первостепенное различие, показанное на рис. 2, относится к двум локусам, к которым может быть приписана неопределенность: внешний мир или наше состояние знания. Например, мы приписываем причинным системам в реальном мире неопределенность, связанную с подбрасыванием монеты, вытягиванием карт из колоды, результатами футбольного матча и поведением вулкана Св. Хеленса. Эти причинные системы имеют диспозиции производить различные события, а мы судим о вероятности этих событий, оценивая относительную силу конкурирующих диспозиций. Наоборот, такие утверждения, как “Я думаю, что Монблан – это самая высокая гора в Европе” или “Я надеюсь, что написал ее имя правильно”, отражают неопределенность, которая приписывается скорее уму человека, а не горе или женщине. (Howell и Burnett, 1978, применили термины *внутренняя неопределенность* и *внешняя неопределенность*, соответственно, к событиям, которые субъект может и не может контролировать.)

Наше различие между незнанием и внешней неопределенностью тесно связано с более общим различием между внешними и внутренними атрибутами опыта. Цвет, размер и текстура, например, обычно ощущаются как свойства, которые принадлежат внешним объектам, но боль, чувства и воспоминания приписываются скорее чувствующему субъекту, чем проявляющемуся объекту.

Относительно атрибуции неопределенности можно иногда сделать вывод, исходя из простого лингвистического теста: подходящей ли при описании оценки неопределенности будет фраза “*вероятность составляет...*”? Или нужно сказать “*моя вероятность составляет...*”? В противоположность позиции Байеса, который трактует все вероятности как субъективные и личные, естественный язык отмечает различие между внешней и внутренней неопределенностью. Так, будет законным говорить о “лучшей оценке вероятности смены режима в Саудовской Аравии в следующем году”, но будет ненормальным говорить, что “наилучшей оценкой вероятности того, что Нил – это самая длинная река в мире, будет...”. Наилучшие оценки вероятности принадлежат общественной сфере. Выражения о частном неведении – нет.

Этот тест не всегда отличает внутреннюю неопределенность от внешней. Например, можно говорить о вероятности того, что Марлоу написал “Гамлета”, хотя эта вероятность приписывается, скорее, нашему неведению, чем силе склонности Марлоу к написанию пьес. Использование “термина вероятность” в этом примере оправдывается существованием общественной части знания, которая уменьшает, но не ликвидирует неопределенность относительно авторства “Гамлета”. Не всем нужно иметь доступ к этому знанию, но оцениваемая вероятность относится к разумному или согласованному выводу на основе имеющихся свидетельств. Однако, в примере с Нилом общественный массив свидетельств, конечно же, включает правильный ответ, и неведение может быть только частным.

Атрибуция неопределенности события к диспозиции или к неведению зависит, помимо всего прочего, от времени. Неопределенность прошлых событий, вероятно, будет ощущаться как неведение, особенно если истина известна кому-либо еще, тогда как неопределенность будущего, более естественно, приписывается диспозициям релевантной системы. В самом деле, отмечалось, что люди демонстрируют различные отношения к результату подбрасывания монет, в зависимости от того, подброшена уже монета или еще нет (Rothbart & Snyder, 1970).

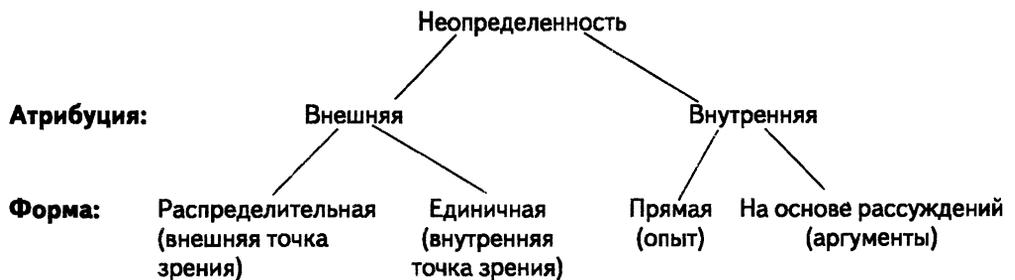


Рис. 2. Варианты неопределенности

Варианты неопределенности

Второй уровень на Рисунке 2 различает четыре прототипных варианта неопределенности, идентифицированной характером данных, которые может рассматривать человек, выносящий суждение, оценивая вероятность. Внешняя неопределенность может быть оценена двумя способами: (i) согласно распределительной форме, когда рассматриваемый случай принимается как пример класса подобных случаев, для которых относительная частота известна или может быть оценена; (ii) согласно единичной форме, когда вероятности оцениваются из наличных особенностей частного рассматриваемого случая. Эти две формы суждения иллюстрируются следующей правдивой историей.

Группа, имевшая дело с разработкой учебного плана высшей школы по мышлению в условиях неопределенности, проводила заседание по планированию. Был поднят вопрос о времени, которое потребуется, чтобы завершить первый вариант учебника. Участников обсуждения попросили оценить это значение как можно более реалистично; семь оценок ранжировались от 18 месяцев до 3 лет. Тогда руководитель группы обратился к одному из участников, работнику сферы образования со значительным опытом в проблемах разработки учебных планов со следующим вопросом: “Каков был опыт других групп, которые пытались написать учебник и разработать учебный план в новой области, где не существовало никаких предшествующих учебных курсов? Сколько им потребовалось времени, чтобы закончить учебник, начиная с той стадии, в которой находится наш проект?” Приводящий в уныние смысл ответа оказался даже более неожиданным для эксперта, который его давал, чем для остальных участников: “Большинство групп, о которых я могу вспомнить, потерпели неудачу и так и не закончили учебник. У тех, кто преуспел в этом деле, время завершения составляет от 5 до 9 лет, со средним значением 7 лет”.

Последующая обнаружила обнаружило, что все участники произвели свои первоначальные оценки согласно единичной форме, строя планы и сценарии с некоторым допуском безопасных резервов для непредвиденных случайностей. Из-за эффектов привязки (Kahneman & Tversky, 1974, 1) оценка, полученная при добавлении безопасных резервов к текущим планам, наиболее вероятно, будет в высокой степени оптимистичной. Примечательным аспектом этого эпизода было то, что релевантные данные распределения не были использованы самопроизвольно, хотя они были в наличии у одного из экспертов в личном знании, и могли быть оценены достаточно точно несколькими другими участниками.

Наш пример иллюстрирует применение распределительной и единичной формы рассуждений в прогнозировании непрерывной переменной: времени, требуемого на завершение проекта. Данные распределения, в этом случае, состояли из знания об относительной частоте различных сроков завершения. Конечно же, подобное рассуждение можно применить, чтобы оценить вероятность дискретного результата, такого как провал проекта. Относительная частота такого результата в релевантном классе обеспечивает основу для распределительной оценки вероятности, а другая информация о частном случае, использованная в единичной форме, может нести отпечаток склонности к провалу или к успеху. Существует много примеров, в которых к одному и тому же вопросу можно подойти как с позиции единичной, так и распределительной формы.

Сравните следующие примеры:

1. “Есть шанс, что ты застанешь Джона дома, если позвонишь завтра утром. Он говорил, что предпочитает работать дома”.
2. “Есть шанс, что ты застанешь Джона дома, если позвонишь завтра утром. Он часто был дома, когда я звонил ему”.

Утверждение 1 допускает только единичное суждение о вероятности, что Джон будет дома. Утверждение 2 могло бы поддержать как распределительную, так и единичную оценки. Относительная частота сходных событий, когда Джон был утром дома, обеспечивает естественную оценку вероятности, что он будет там завтра, но это выражение также наделило Джона склонностью проводить утро дома, в большей степени, чем это сделало Утверждение 1.

Мы предположили (Kahneman & Tversky, 1979a, 30), что люди обычно предпочитают единичную форму, согласно которой они принимают “внутреннюю точку зрения” причинной системы, который немедленно выдает исход, в большей степени “внешнюю точку зрения”, которая связывает рассматриваемый случай со схемой осуществления выборки. Наш пример планирования иллюстрирует это предпочтение единичной модели. Он также иллюстрирует еще один эффект, который, как мы подозреваем, достаточно общий: распределительная форма суждения с большей вероятностью, чем единичная, выдает точные оценки значений и разумную оценку вероятности.

Теперь мы обратимся к различию между формами оценки внутренней неопределенности, которые иллюстрируют следующие примеры:

3. “Я полагаю, что Нью-Йорк севернее Рима, но я не уверен”.
4. “Я думаю, ее зовут Дорис, но я не уверен”.

Неопределенность, выраженная этими утверждениями, полностью внутренняя: утверждения отражают (частичное) неведение, а не диспозицию внешних объектов. Несомненно, говорить о некоторой возможности Нью-Йорка быть расположенным севернее Рима (между прочим, это не так), или о некоторой возможности Линды запоминаться как Дорис, несколько неестественно.

Эти два утверждения различаются по характеру свидетельства, на котором они основаны. Утверждение 3 могло бы отразить процесс тщательной проверки и взвешивания свидетельств и доводов (например, Нью-Йорк намного холоднее, чем Рим, Рим находится в центре Италии и т.д.). Утверждение 4 имеет иной характер. Выраженная уверенность основана на интроспективном суждении о силе ассоциации. Как это часто случается, когда мы проверяем орфографию по тому, “правильно ли выглядит” слово, уверенность опирается на непроанализированный опыт. В исследованиях психофизики и памяти уверенность, связанная с суждениями, значительно коррелирует с точностью: люди с большей вероятностью уверены, когда они правы, чем когда неправы, хотя их оценки о собственной правоте плохо калиброваны (см. Главу 22).

Как и в случае внешней неопределенности, внутренняя неопределенность, связанная с определенным вопросом, иногда может быть оценена как аргументированной формой, так и интроспективной формой. Например, к вопросу относительно возраста кинозвезды можно подойти интроспективно, пытаясь найти ответ, который звучит похоже, или согласно аргументированной форме, пытаясь извлечь ответ из других знаний.

Мы не хотели бы предположить, что любой опыт неопределенности можно отнести к одному из четырех вариантов Рисунка 2. Несомненно, существует много смешанных и неопределенных случаев. Мы видели, что неопределенность в данной проблеме может быть приписана внешним диспозициям, неведению человека или комбинации того и другого, и что она может быть оценена согласно единичной форме, распределительной форме или смеси этих форм. Целью нашего трактования было выдвинуть на первый план некоторые значительные измерения в разнообразии опыта неопределенности, а не предложить исчерпывающую взаимоисключающую классификацию этого опыта. Попытка классификации экспериментальных операций в измерении субъективной вероятности описана у Хоуэлл и Бернетт (Howell и Burnett, 1978).

Обсуждение

Хотя язык вероятности может быть использован для выражения любой формы неопределенности, законы теории вероятности не применяются ко всем вариантам неопределенности с одинаковой силой. Наиболее вероятно, эти законы принимаются и удовлетворяются в интуитивных суждениях, когда внешняя неопределенность оценивается согласно распределительной, или частотной, форме. Например, комплементарность субъективной вероятности является очень непреодолимой тогда, когда мы обращаемся к статистическим данным о погоде, чтобы оценить вероятность того, что будет дождь 12 апреля следующего года: релевантный набор прошлых апрельских дней вполне можно разделить на дождливые дни и дни, когда дождя не было.

Комплементарность менее непреодолима в других вариантах. Когда неопределенность оценивается в терминах склонности возможности, доводов или уверенности, менее очевидно, что вероятность следует увеличивать до единицы, — даже если определенно известно, что одна из альтернатив правильна. Например, можно подвергнуть сомнению, почему степень убежденности в утверждении, что Нью-Йорк севернее Рима, и степень убежденности в утверждении, что Нью-Йорк южнее Рима, должны в сумме давать то же значение, что и степени убежденности для любой другой пары комплементарных утверждений. В самом деле, некоторые авторы (например, Cohen, 1977; Shafer, 1976) предполагали, что комплементарность не следует применять к степеням убежденности. В частности, Шафер оспаривал комплементарность убежденности на том основании, что существуют ситуации, в которых две взаимоисключающие и исчерпывающие гипотезы обе имеют существенную поддержку, и другие ситуации, в которых ни одна из гипотез не имеет особой поддержки. Сходные вопросы могут быть подняты в отношении необходимости комплементарности в ощущении уверенности и в оценке конфликтных склонностей.

Варианты неопределенности могут различаться по уверенности, с которой они оцениваются. Представьте, что канцелярская кнопка была подро-

шена четыре раза и упала дважды на кончик и дважды на головку. Имея эти данные, наблюдатели определяют вероятность 0.5 событию, что кнопка упадет на головку при следующем подбрасывании. Они также определяют вероятность 0.5 событию, что подброшенная монета покажет “решку”, но они выражают гораздо большую уверенность в своем суждении о монете, чем о кнопке. Как иллюстрирует этот пример, вполне возможно определить различную степень уверенности одному и тому же суждению о возможности. Уверенность в отношении вероятностей важна, поскольку она контролирует решения. Есть свидетельства (Ellsberg, 1961; Raiffa, 1961), что люди предпочитают ставить на события, которые имеют известные вероятности, такие как подбрасывание монеты, а не события, которые связаны с комбинацией внешней неопределенности и неведения, такие как подбрасывание канцелярской кнопки.

Существуют естественные связи между концепциями вероятности, выдвинутыми различными школами, изучающими эту тему, и формами неопределенности, которые мы обсудили. Так, частотная, или объективная, интерпретация вероятности ограничивает концепцию до внешней неопределенности, генерируемой процессом осуществления выборки. Наоборот, школа Байеса, или личностная школа, трактует всю неопределенность как незнание. Согласно школе Байеса, предпочтения являются основой убеждений, а вероятности являются производными от предпочтений между пари. Однако с психологической точки зрения эта эвристика пари кажется нереалистичной. Обсуждения в этой области всегда были острыми из-за существования различных видов интуиции, которые являются индивидуально вызванными и взаимно несовместимыми, и из-за того, что нет согласия по критерию для разрешения нормативных споров, когда эти виды интуиции конфликтуют. Психологический анализ, возможно, мог бы внести вклад в нормативное обсуждение, обеспечив адекватное описание видов интуиции, из которых черпают свою привлекательность различные позиции.

Литература

Abelson, R.P. Modes of resolution of belief dilemmas. *Journal of Conflict Resolution*, 1959,3, 343-352. (8)

Abelson, R.P., Aronson, E., McGuire, W.J., Newcomb, T.M., Rosenberg, M.J., & Tannenbaum, P.H. (Eds.). *Theories of cognitive consistency: A sourcebook*. Chicago: Rand-McNally, 1968. (15)

Adams, J. K. A confidence scale defined in terms of expected percentages. *American Journal of Psychology*, 1957, 70, 432-436. (20, 22)

Adams, J.K., & Adams, P. A. Realism of confidence judgments. *Psychological Review*, 1961,68, 33-45. (22,31)

Adams, P.A., & Adams, J.K. Training in confidence judgments. *American Journal of Psychology*, 1958, 71, 747-751. (22, 31)

Adler, A. Individual psychology. In C. Murchinson (Ed.), *Psychologies of 1930*. Worcester, Mass.: Clark University Press, 1930. (16)

Adorno, T., Frenkel-Brunswik, E., Levinson, D., & Sanford, N. *The authoritarian personality*, New York: Harper, 1950. (15)

Ajzen, I. Intuitive theories of events and the effects of base-rate information on prediction, *Journal of Personality and Social Psychology*, 1977, 35, 303-314. (10)

Alberoni, F. Contribution to the study of subjective probability, Part I. *Journal of General Psychology*, 1962, 66, 241-264. (3)

Alexander, S.A.H. Sex, arguments, and social engagements in marital and premarital relations. Unpublished master's thesis. University of Missouri, Kansas City, 1971. (28)

Alker, H.A. Is personality situationally specific or intrapsychically consistent? *Journal of Personality*, 1972, 40, 1-16. (15)

Allen, J.G. Breast surgery, principles and practice. In C.A. Moyer (Ed.), *Surgery: Principles and Practice*. Philadelphia: Lippincott, 1965. (18)

Allport, G.W. *The nature of prejudice*. Reading, Mass.: Addison-Wesley, 1954. (15)

- Allport, G.W. Traits revisited. *American Psychologist*, 1966, 21, 1-10. (15)
- Alpert, W., & Raiffa, H.A progress report on the training of probability assessors. Unpublished manuscript, 1969. (1, 22, 33)
- Anderson, C.A., Lepper, M.R., & Ross, L. The perseverance of social theories: The role of explanation in the persistence of discredited information. *Journal of Personality and Social Psychology*, 1980, 39, 1037-1049. (9, 15)
- Anderson, N.H. Information integration theory: A brief survey. In D.H. Krantz, R.C. Atkinson, R.D. Luce, and P. Suppes (Eds.), *Contemporary Developments in Mathematical Psychology* (Vol. 2). San Francisco: Freeman, 1974. (27)
- Arkes, H.R., Wortmann, R.C., Saville, P.D., & Harkness, A.R. Hindsight bias among physicians weighing the likelihood of diagnoses. *Journal of Applied Psychology*, 1981, 66, 252-254. (31)
- Armeliuss, K. Task predictability and performance as determinants of confidence in multiple-cue judgments. *Scandinavian Journal of Psychology*, 1979, 20, 19-25. (31)
- Ashmore, R.D., & Collins, B.E. Studies in forced compliance X: Attitude change and commitment to maintain publicly a counter-attitudinal position. *Psychological Reports*, 1968, 22, 1229-1234. (9)
- Bach, R. Nothing by chance. *The American Way*, 1973, 6, 32-38. (33)
- Bacon, F. *The new organon and related writings*. New York: Liberal Arts Press, 1960. (Originally published in 1620.) (9)
- Barclay, S., & Peterson, C.R. Two methods for assessing probability distributions (Tech. Rep. 73-1). McLean, Va.: Decisions and Designs, Inc., 1973. (22)
- Bar-Hillel, M. On the subjective probability of compound events. *Organizational Behavior and Human Performance*, 1973, 9, 396-406. (1, 23, 30)
- Bar-Hillel, M. Similarity and probability. *Organizational Behavior and Human Performance*, 1974, H, 277-282. (5)
- Bar-Hillel, M. The role of sample size in sample evaluation. *Organizational Behavior and Human Performance*, 1979, 24, 245-257. (5, 34)
- Bar-Hillel, M. The base-rate fallacy in probability judgments. *Acta Psychologica*, 1980, 44, 211-233. (10, 34) (a)
- Bar-Hillel, M. What features make samples seem representative? *Journal of Experimental Psychology: Human Perception and Performance*, 1980, 6, 578-589. (5, 6) (b)
- Bar-Hillel, M., & Falk, R. Some teasers concerning conditional probabilities. Unpublished manuscript, 1980. (34)
- Bar-Hillel, M., & Fischhoff, B. When do base rates affect predictions? *Journal of Personality and Social Psychology*, 1981, 41, 671-680. (10)
- Barraclough, G. *Mandarins and Nazis*. *New York Review of Books*, 1972, 19, 37-42. (23)
- Bartlett, F.C. *Remembering*. Cambridge: Cambridge University Press, 1932. (12)
- Beaver, W.H. Financial ratios as predictors of failure. In *Empirical research*

in accounting: Selected studies. Chicago: University of Chicago, Graduate School of Business, Institute of Professional Accounting, 1966. (28)

Becker, B. W., & Greenberg, M. G. Probability estimates by respondents: Does weighting improve accuracy? *Journal of Marketing Research*, 1978, 15, 482-486. (31)

Becker, C. Everyman his own historian. *American Historical Review*, 1935, 40, 221-236. (23)

Beckman, L. Effects of students' performance on teachers' and observers' attributions of causality. *Journal of Educational Psychology*, 1970, 61, 75-82. (9)

Bem, D. J. Self-perception: An alternative interpretation of cognitive dissonance phenomena. *Psychological Review*, 1967, 74, 183-200. (7, 32)

Bem, D. J., & Allen, A. On predicting some of the people some of the time: The search for cross-situational consistencies in behavior. *Psychological Review*, 1974, 51, 506-520. (15)

Bem, D. J., & Funder, D. C. Predicting more of the people more of the time: Assessing the personality of situations. *Psychological Review*, 1978, 85, 485-501. (15)

Bem, D. J., & Lord, C. G. Template matching: A proposal for probing the ecological validity of experimental settings in social psychology. *Journal of Personality and Social Psychology*, 1979, 37, 833-846. (15)

Bem, D. J., & McConnell, H. K. Testing the self-perception explanation of dissonance phenomena: On the salience of premanipulation attitudes. *Journal of Personality and Social Psychology*, 1970, 14, 23-31. (32)

Benson, L. *Toward the scientific study of history: Selected essays*. Philadelphia; Lippincott, 1972. (23)

Bentham, J. *Handbook of political fallacies*. Baltimore: Johns Hopkins Press, 1952. (Originally published in 1824.) (32)

Ben Zvi, A. Hindsight and foresight: A conceptual framework for the analysis of surprise attacks. *World Politics*, 1976, 28, 381-395. (23)

Beyth, R. Man as an intuitive statistician: On erroneous intuitions concerning the description and prediction of events. Unpublished master's thesis. The Hebrew University, Jerusalem, 1972. (In Hebrew) (4)

Beyth-Marom, R., & Dekel, S. Thinking under uncertainty: A textbook for junior high school students. In press. (In Hebrew) (31, 33)

Bierbrauer, G. Effect of set, perspective, and temporal factors in attribution. Unpublished doctoral dissertation, Stanford University, 1973. (9)

Block, M. A., & Reynolds, W. How vital is mammography in the diagnosis and management of breast carcinoma? *Archives of Surgery*, 1974, 108, 588-591. (18)

Borgida, E. Scientific deduction-Evidence is not necessarily informative: A reply to Wells and Harvey. *Journal of Personality and Social Psychology*, 1978, 36, 477-482. (10)

Borgida, E., & Brekke, N. The base-rate fallacy in attribution and prediction.

In J.H. Harvey, W.J. Ickes, and R.F. Kidd (Eds.), *New Directions in Attribution Research* (Vol. 3). Hillsdale, N.J.: Erlbaum, 1981. (10)

Borgida, E., & Nisbett, R. The differential impact of abstract vs. concrete information on decisions. *Journal of Applied Social Psychology*, 1977, 7, 258-271. (7)

Bousfield, W.A., & Sedgewick, C.H. An analysis of sequences of restricted associative responses. *Journal of General Psychology*, 1944, 30, 149-165. (11)

Bowman, E. H. Consistency and optimality in managerial decision making. *Management Science*, 1963, 9, 310-321. (28)

Bradley, G. W. Self-serving biases in the attribution process: A reexamination of the fact or fiction question. *Journal of Personality and Social Psychology*, 1978, 36, 56-71. (9)

Braine, M.D.S. On the relation between the natural logic of reasoning and standard logic. *Psychological Review*, 1978, 85, 1-21. (34)

Bransford, J.D., & Franks, J.J. The abstraction of linguistic ideas. *Cognitive Psychology*, 1971, 2, 331-350. (6)

Brier, G.W. Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 1950, 75, 1-3. (22)

Broadbent, D.E. Word-frequency effect and response bias. *Psychological Review*, 1967, 74, 1-15. (35)

Broadbent, D.E. *Decision and Stress*. London: Academic Press, 1971. (35)

Brown, E.A.R. The tyranny of a construct: Feudalism and historians of medieval Europe. *American Historical Review*, 1974, 79, 1063-1088. (23)

Brown, T.A. An experiment in probabilistic forecasting (Report R-944-ARPA). Santa Monica: The RAND Corp., 1973. (22)

Brown, T.A., & Shuford, E.H. Quantifying uncertainty into numerical probabilities for the reporting of intelligence (Report R-1185-ARPA). Santa Monica: The RAND Corp., 1973. (22)

Brucker, H. *Communication is power: Unchanging values in a changing journalism*. New York: Oxford University Press, 1973. (33)

Bruner, J.S. Going beyond the information given. In H. Gruber, et al. (Eds.), *Contemporary approaches to cognition*. Cambridge, Mass.: Harvard University Press, 1957. (15) (a)

Bruner, J.S. On perceptual readiness. *Psychological Review*, 1957, 64, 123-152. (15) (b)

Bruner, J.S. The act of discovery. *Harvard Educational Review*, 1961, 31, 21-32. (12)

Bruner, J.S., Postman, L., & Rodrigues, J. Expectations and the perception of color. *American Journal of Psychology*, 1951, 64, 216-227. (15)

Bruner, J.S., & Potter, M.C. Interference in visual recognition. *Science*, 1964, 144, 424-425. (11)

Brunswik, E. Organismic achievement and environmental probability. *Psychological Review*, 1943, 50, 255-272. (19)

Bryan, W. B. Testimony before the Subcommittee on State Energy Policy,

Committee on Planning, Land Use, and Energy, California State Assembly, February 1, 1974. (33)

Burton, I., Kates, R.W., & White, G.F. The environment as hazard. New York: Oxford University Press, 1978. (33)

Byrne, R. R. Correlation of thermography, xeromammography and biopsy in a community hospital. *Wisconsin Medical Journal*, 1974, 73, 35-37. (18)

Cambridge, R.M., & Shreckengost, R. C. Are you sure? The subjective probability assessment test. Unpublished manuscript. Langley, Va.: Office of Training, Central Intelligence Agency, 1978. (22)

Campbell, D.T. Blind variation and selective retention in creative thought as in other knowledge processes. *Psychological Review*, 1960, 67, 380-400. (19)

Campbell, D.T. Reforms as experiments. *American Psychologist*, 1969, 24, 409-429. (4)

Campbell, D. T. "Degrees of freedom" and the case study. *Comparative Political Studies*, 1975, 8, 178-193. (23)

Carlsson, G. Random walk effects in behavioral data. *Behavioral Science*, 1972, 17, 430-437. (23)

Carr, E.H. What is history? London: Macmillan, 1961. (23)

Carver, R.P. A critical review of mathagenic behaviors and the effect of questions upon the retention of prose materials. *Journal of Reading Behavior*, 1972, 4, 93-119. (12)

Cass, J., & Birnbaum, M. Comparative guide to American colleges. New York: Harper & Row, 1968. (28)

Casscells, W., Schoenberger, A., & Grayboys, T. Interpretation by physicians of clinical laboratory results. *New England Journal of Medicine*, 1978, 299, 999-1000. (10, 18)

Castellan, N.J., Jr. Decision making with multiple probabilistic cues. In N. J. Castellan, D.P. Pisoni, & G.R. Potts (Eds.), *Cognitive Theory*, 2. Hillsdale, N.J.: Erlbaum, 1977. (19)

Cavanaugh, J.C., & Borkowski, J.G. Searching for meta-memory-memory connections. *Developmental Psychology*, 1980, 16, 441-453. (31)

Chan, S. The intelligence of stupidity: Understanding failures in strategic warning. *American Political Science Review*, 1979, 73, 171-180. (23)

Chapman, L.J. Illusory correlation in observational report. *Journal of Verbal Learning and Verbal Behavior*, 1967, 6, 151-155. (15)

Chapman, L.J., & Chapman, J.P. Genesis of popular but erroneous psychodiagnostic observations. *Journal of Abnormal Psychology*, 1967, 73, 193-204. (15)

Chapman, L.J., & Chapman, J.P. Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology*, 1969, 74, 271-280. (1, 15, 17)

Choo, G.T.G. Training and generalization in assessing probabilities for discrete events (Tech. Rep. 76-5). Uxbridge, England: Brunel Institute of Organizational and Social Studies, 1976. (22)

Christensen-Szalanski, J.J.J., & Bushyhead, J.B. Physicians' use of probabilistic information in a real clinical setting. *Journal of Experimental Psychology: Human Perception and Performance*, 1981, 7, 928-935. (22, 31)

Clark, H. H., & Clark, E. V. *Psychology and Language*. New York: Harcourt Brace Jovanovich, 1977. (34)

Clark, R.L., Copeland, M.M., Egan, R.L., Gallagher, H. S., Geller, H., Lindsay, J. P., Robbins, L.C., & White, E.C. Reproducibility of the technic of mammography (Egan) for cancer of the breast. *American Journal of Surgery*, 1965, 109, 127-133. (18)

Clark, R.L., & Robbins, L.C. Mammography (Egan) in cancer of the breast. *The American Journal of Surgery*, 1965, 209, 125-126. (18)

Clarke, F.R. Confidence ratings, second-choice responses, and confusion matrices in intelligibility tests. *Journal of the Acoustical Society of America*, 1960, 32, 35-46. (22, 31)

Claudy, J.G. A comparison of five variable weighting procedures. *Educational and Psychological Measurement*, 1972, 32, 311-322. (28)

Cocozza, J.J., & Steadman, H.J. Prediction in psychiatry: An example of misplaced confidence in experts. *Social Problems*, 1978, 25, 265-276. (31)

Cohen, B., & Lee, I. A catalog of risks. *Health Physics*, 1979, 36, 707-722. (33)

Cohen, B.L. Perspectives on the nuclear debate. *Bulletin of the Atomic Scientists*, 1974, 30, 25-39. (33)

Cohen, J. The statistical power of abnormal-social psychological research. *Journal of Abnormal and Social Psychology*, 1962, 65, 145-153. (2, 3)

Cohen, J. *Statistical power analysis in the behavioral sciences*. New York: Academic Press, 1969, (2)

Cohen, J., Boyle, L.E., & Chesnick, E.I. Patterns of preference in locating targets. *Occupational Psychology*, 1969, 43, 129-144. (24)

Cohen, J., Chesnick, E.I., & Haran, D. A confirmation of the inertial-f effect in sequential choice and decision. *British Journal of Psychology*, 1972, 63, 41-46. (1)

Cohen, J., & Hansel, C.E. . *Risk and gambling*. New York: Philosophical Library, 1956. (3)

Cohen, J., & Hansel, C.E.M. The nature of decisions in gambling. *Acta Psychologica*, 1958, 13, 357-370. (24)

Cohen, J.J. A case for benefit-risk analysis. In H.J. Otway (Ed.), *Risk vs. benefit: Solution or dream (Report LA-4860-MS)*. Los Alamos Scientific Laboratory, February 1972. (Available from the National Technical Information Service.) (33)

Cohen, L.J. *The probable and the provable*. Oxford: Clarendon Press, 1977. (35)

Cohen, L.J. On the psychology of prediction: Whose is the fallacy? *Cognition*, 1979, 7, 385-407. (34)

Cohen, L.J. Can human irrationality be experimentally demonstrated? *The Behavioral and Brain Sciences*, 1981, 4, 317-331. (34)

Cohn, H.E. Mammography in its proper perspective. *Surgery, Gynecology and Obstetrics*, 1972, 334, 97-98. (18)

Coles, R. Shrinking history. *New York Review of Books*, Part I, February 22, 1973, pp. 20, 15-21; Part II, March 8, 1973, pp. 20, 25-29. (23)

Collins, B.E., & Hoyt, M.F. Personal responsibility for consequences: An integration and extension of the forced-compliance literature. *Journal of Experimental Social Psychology*, 1972, 8, 558-593. (9)

Combs, B., & Slovic, P. Causes of death: Biased newspaper coverage and biased judgments. *Journalism Quarterly*, 1979, 56, 837-843, 849. (33)

Commager, H.S. *The nature and study of history*. Columbus, Ohio: Merrill, 1965. (23)

Committee on Government Operations. *Teton Dam Disaster*. Washington, D.C.: U.S. Government Printing Office, 1976. (33)

Cooke, W.E. Forecasts and verifications in Western Australia. *Monthly Weather Review*, 1906, 34, 23-24. (22) (a)

Cooke, W.E. Weighting forecasts. *Monthly Weather Review*, 1906, 34, 274-475. (22) (b)

Cox, D.R. Two further applications of a model for binary regression. *Biometrika*, 1958, 45, 562-565. (22)

Cronbach, L. J. Beyond the two disciplines of scientific psychology. *American Psychologist*, 1975, 30, 116-127. (19)

Grouse, T. *The boys on the bus*. New York: Random House, 1974. (7)

D'Andrade, R.G. Trait psychology and componential analysis. *American Anthropologist*, 1965, 67, 215-228. (13)

Darley, J.M., & Latane, B. Bystander intervention in emergencies: Diffusion in responsibility. *Journal of Personality and Social Psychology*, 1968, 8, 377-383. (7,10)

Darlington, R.B. Reduced-variance regression. *Psychological Bulletin*, 1978, 85, 1238-1255. (28)

Davis, W.L., & Davis, D.E. Internal-external control and attribution of responsibility for success and failure. *Journal of Personality*, 1972, 40, 123-136. (9)

Dawes, R.M. A case study of graduate admissions: Application of three principles of human decision making. *American Psychologist*, 1971, 26, 180-188. (28)

Dawes, R.M. Graduate admissions criteria and future success. *Science*, 1975, 187, 721-723. (28) (a)

Dawes, R.M. The mind, the model, and the task. In F. Restle, R.M. Shiffrin, N.J. Castellan, H.R. Lindman & D.F. Pisoni (Eds.), *Cognitive theory*, Hillsdale, N.J.: Erlbaum, 1975. (19) (b)

Dawes, R.M. Shallow Psychology. In J.S. Carroll & J.W. Payne (Eds.), *Cognition and social behavior*. Hillsdale, N.J.: Erlbaum, 1976. (13, 28, 31)

Dawes, R.M. Confidence in intellectual judgments vs. confidence in perceptual judgments. In E.D. Lantermann & H. Feger (Eds.), *Similarity and Choice: Papers in honor of Clyde Coombs*. Bern: Hans Huber, 1980. (31)

Dawes, R.M., & Corrigan, B. Linear models in decision making. *Psychological Bulletin*, 1974, 81, 95-106. (28)

Dawid, A. P. The well-calibrated Bayesian. *Journal of the American Statistical Association*, in press. (22)

Deacon, E.B. A discriminant analysis of predictors of business failure. *Journal of Accounting Research*, 1972, 10, 167-179. (28)

DeCharms, R.C. *Personal causation: The internal affective determinants of behavior*. New York: Academic Press, 1968. (12,16)

Deci, E. Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology*, 1971, 18, 105-115. (9)

De Finetti, B. La prevision: Ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincare*, 1937, 7, 1-68. English translation in H.E. Kyburg, Jr., & H.E. Smokier (Eds.), *Studies in subjective probability*. New York: Wiley, 1964. (22)

De Finetti, B. Probability: Interpretations. In D.E. Sills (Ed.), *International Encyclopedia of the Social Sciences* (Vol. 12). New York: Macmillan, 1968. Pp. 496-504. (1)

Degler, C.N. Why historians change their minds. *Pacific Historical Review*, 1976, 48, 167-189. (23)

DeGowin, E., & DeGowin, R. *Bedside diagnostic examination* (2nd ed.). New York: Macmillan, 1969. (18)

DeGroot, A.D. *Het denken van den schaker* [Thought and choice in chess]. The Hague, The Netherlands: Mouton, 1965. (28)

Del Regato, J.A. Diagnosis, treatment and prognosis. In L.V. Ackerman (Ed.), *Cancer* (4th Ed.) St. Louis: Mosby, 1970. (18)

DeLuca, J.T. A statistical comparison study of patients undergoing breast biopsy at a community hospital over a 16-year period. *Radiology*, 1974, 222, 315-318. (18)

Dennett, D.C. *Brainstorms*. Hassocks, England: Harvester, 1979. (35)

DeSmet, A.A., Fryback, D.G., & Thornbury, J.R. A second look at the utility of radiographic skull examination for trauma. *American Journal of Radiology*, 1979, 132, 95-99. (22)

Detmer, D.E., Fryback, D.G., & Gassner, K. Heuristics and biases in medical decision-making. *Journal of Medical Education*, 1978, 53, 682-683. (31)

Diaconis, P. Statistical problems in ESP research. *Science*, 1978, 201, 131-136. (23)

Dodson, J. D. Simulation system design for a TEAS simulation research facility. (Tech. Rep. No. AFCRL-1112, PRC R-194). Los Angeles: Planning Research Corporation, November, 1961. (26)

Donchin, E., Ritter, W., & McCallum, W.C. Cognitive psychophysiology: The endogenous components of the ERP. In E. Callaway, P. Teuting, & S.H. Koslow (Eds.), *Event-related brain potentials in man*. New York: Academic Press, 1978. (35)

Dowie, J. On the efficiency and equity of betting markets. *Economica*, 1976, 43, 139-150. (22, 31)

Dowie, J. Personal communication. November 1980. (22)

Dreman, D. *Contrarian investment strategy*. New York: Random House, 1979. (23, 31)

Duncan-Johnson, C.C., & Donchin, E. On quantifying surprise: The variation of event-related potentials with subjective probability. *Psychophysiology*, 1977, 14, 456-467. (35)

Dunlap, T.R. Science as a guide in regulating technology: The case of DDT in the United States. *Social Studies of Science*, 1978, 8, 265-285. (33)

Duval, S., & Hensley, V. Extensions of objective self-awareness theory: The focus of attention-causal attribution hypothesis. In J.H. Harvey, W.L. Ickes, & R.F. Kidd (Eds.), *New directions in attribution research* (Vol. 1). Hillsdale, N.J.: Erlbaum, 1976. (9)

Duval, S., & Wicklund, R.A. *A theory of objective self-awareness*. New York: Academic Press, 1972. (9)

Edwards, A.E. *Experimental design in psychological research*. New York: Holt, Rinehart & Winston, 1960. (20)

Edwards, D.D., & Edwards, J.S. Marriage: Direct and continuous measurement. *Bulletin of the Psychonomic Society*, 1977, 20, 187-188. (28)

Edwards, W. Optimal strategies for seeking information: Models for statistics, choice reaction times, and human information processing. *Journal of Mathematical Psychology*, 1965, 2, 312-329. (25)

Edwards, W. Nonconservative probabilistic information processing systems. (Report ESD-TR-66-404). U.S. Air Force, AF Systems Command. Electronic Systems Division, Decision Sciences Laboratory, 1966. (25, 26)

Edwards, W. Conservatism in human information processing. In B. Kleinmuntz (Ed.), *Formal representation of human judgment*. New York: Wiley, 1968, 17-52. (I, 2, 3, 8)

Edwards, W. Bayesian and regression models of human information processing – a myopic perspective. *Organizational Behavior and Human Performance*, 1971, 6, 639-648. (19)

Edwards, W. Comment. *Journal of the American Statistical Association*, 1975, 70, 291-293. (34)

Edwards, W. Technology for director dubious: Evaluation and decision in public contexts. In K. R. Hammond (Ed.), *Judgment and decision in public policy formation*. Boulder, Colo.: Westview Press, 1978. (28)

Edwards, W., Lindman, H., & Phillips, L.D. Emerging technologies for making decisions. *New directions in psychology* (Vol. 2). New York: Holt, Rinehart & Winston, 1965. (25)

Edwards, W., Lindman, H., & Savage, L.J. Bayesian statistical inference for psychological research. *Psychological Review*, 1963, 70, 193-242. (25)

Egan, R.L. Fundamentals of mammographic diagnoses of benign and malignant diseases. *Oncology*, 1969, 23, 126-148. (18)

- Egan, R.L. Contributions of mammography in the detection of early breast cancer. *Cancer*, 1971, 28, 1555-1557. (18)
- Egan, R.L. *Mammography* (2nd Ed.). Springfield, 111.: Charles C. Thomas, 1972. (18)
- Einhorn, H.J. Expert measurement and mechanical combination. *Organizational Behavior and Human Performance*, 1972, 7, 86-106. (28)
- Einhorn, H.J. Decision errors and fallible judgment: Implications for social policy. In K.R. Hammond (Ed.), *Judgment and decision in public policy formation*. Denver, Colo.: Westview Press, 1978. (23)
- Einhorn, H.J. Personal communication, January 1979. (28)
- Einhorn, H.J., & Hogarth, R.M. Unit weighting schemas for decision making. *Organizational Behavior and Human Performance*, 1975, 13, 171-192. (28)
- Einhorn, H.J., & Hogarth, R.M. Confidence in judgment: Persistence in the illusion of validity. *Psychological Review*, 1978, 85, 395-416. (9, 19, 28, 34)
- Einhorn, H.J., & Hogarth, R.M. Behavioral decision theory: Processes of judgment and choice. *Annual Review of Psychology*, 1981, 32, 53-88. (34, 35)
- Einhorn, H.J., Kleinmuntz, D.N., & Kleinmuntz, B. Linear regression and process tracing models of judgment. *Psychological Review*, 1979, 86, 465-485. (19)
- Ellis, R.J., & Holmes, J.G. Focus of attention in social interaction: The impact on self evaluations. Unpublished manuscript. Ontario, Canada: University of Waterloo, 1979. (9)
- Ettsberg, D. Risk, ambiguity, and the Savage axioms. *Quarterly Journal of Economics*, 1961, 75, 643-699. (35)
- Epstein, W., & Rock, I. Perceptual set as an artifact of recency. *American Journal of Psychology*, 1960, 73, 214-228. (35)
- Erlick, D.E. Human estimates of statistical relatedness. *Psychonomic Science*, 1966, 5, 365-366. (15)
- Erlick, D.E., & Mills, R.G. Perceptual quantification of conditional dependency. *Journal of Experimental Psychology*, 1967, 73, 9-14. (15)
- Estes, W.K. Probability learning. In A.W. Melton (Ed.), *Categories of human learning*. New York: Academic Press, 1964. (2, 3)
- Estes, W.K. The cognitive side of probability learning. *Psychological Review*, 1976, 83, 37-64. (19)
- Evans, J.St.B.T., & Dusoir, A.E. Proportionality and sample size as factors in intuitive statistical judgment. *Acta Psychologica*, 1977, 41, 129-137. (34)
- Evans, J.St.B.T., & Wason, P.C. Rationalization in a reasoning task. *British Journal of Psychology*, 1976, 67, 486-497. (34)
- Fabricand, B. P. *Horse sense*. New York: David McKay, 1965. (22)
- Fama, E. F. Random walks in stock market prices. *Financial Analysts Journal*, 1965, 21, 55-60. (23)
- Feather, N.T. Attribution of responsibility and valence of success and failure in relation to initial confidence and task performance. *Journal of Personality and Social Psychology*, 1969, 13, 129-144. (9)

Feller, W. An introduction to probability theory and its applications (3rd ed., Vol. 1). New York: Wiley, 1968. (3, 23, 34)

Ferrell, W.R., & McGoey, P.J. A model of calibration for subjective probabilities. *Organizational Behavior and Human Performance*, 1980, 26, 32-53. (22, 31)

Festinger, L. A theory of cognitive dissonance. Stanford: Stanford University Press, 1957. (9, 15)

Festinger, L. Conflict, decision and dissonance. Stanford: Stanford University Press, 1964. (15)

Feyerabend, P. Against method. New York: NLB (Schocken), 1975. (23)

Fischer, D. H. Historians' fallacies. New York: Harper & Row, 1970. (23, 31)

Fischhoff, B. Hindsight " foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, 1975, 1, 288-299. (23, 30, 31)

Fischhoff, B. Attribution theory and judgment under uncertainty. In N.H. Harvey, W.J. Ickes, & R.F. Kidd (Eds.), *New directions in attribution research*. Hillsdale, N.J.: Erlbaum, 1976. (13)

Fischhoff, B. Cost-benefit analysis and the art of motorcycle maintenance. *Policy Sciences*, 1977, 8, 177-202. (33) (a)

Fischhoff, B. Perceived informativeness of facts. *Journal of Experimental Psychology: Human Perception and Performance*, 1977, 3, 349-358. (23, 31, 34) (b)

Fischhoff, B. Intuitive use of formal models, A comment on Morrison's "Quantitative Models in History." *History and Theory*, 1978, 17, 207-210. (23)

Fischhoff, B. Latitude and latitudes. How much credit do people deserve? In G. Ungson & D. Braunstein (Eds.), *New directions in decision making*. New York: Kent, in press. (31) (a)

Fischhoff, B. No man is a discipline. In J. Harvey (Ed.), *Cognition, social behavior and the environment*. Hillsdale, N.J.: Erlbaum, in press. (23) (b)

Fischhoff, B., & Bar-Hillel, M. Focusing techniques as aids to inference (Decision Research Report 80-9). Eugene, Oreg.: Decision Research, 1980. (10, 34)

Fischhoff, B., & Beyth, R. "I knew it would happen" - Remembered probabilities of once-future things. *Organizational Behavior and Human Performance*, 1975, 13, 1-16. (22, 23, 31)

Fischhoff, B., Goitein, B., & Shapira, Z. The experienced utility of expected utility approaches. In N. Feather (Ed.), *Expectancy, incentive and action*. Hillsdale, N.J.: Erlbaum, in press. (23)

Fischhoff, B., Lichtenstein, S., Slovic, P., Derby, S., & Keeney, R. Acceptable risk. New York: Cambridge University Press, 1981. (33)

Fischhoff, B., & MacGregor, D. Judged lethality (Decision Research Report 80-4). Eugene, Oreg.: Decision Research, 1980. (33)

Fischhoff, B., & Slovic, P. A little learning...: Confidence in multicue judgment. In R. Nickerson (Ed.), *Attention and Performance VIII*. Hillsdale, N.J.: Erlbaum, 1980, (22, 23, 31)

Fischhoff, B., Slovic, P., & Lichtenstein, S. Knowing with certainty; The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance*, 1977, 3, 552-564. (22, 23, 31, 33)

Fischhoff, B., Slovic, P., & Lichtenstein, S. Fault trees: Sensitivity of estimated failure probabilities to problem representation. *Journal of Experimental Psychology: Human Perception and Performance*, 1978, 4, 330-334. (19, 23, 33, 34)

Fischhoff, B., Slovic, P., & Lichtenstein, S. Subjective sensitivity analysis. *Organizational Behavior and Human Performance*, 1979, 23, 339-359. (10, 34)

Fischhoff, B., Slovic, P., & Lichtenstein, S. Knowing what you want: Measuring labile values. In T. Watlsten (Ed.), *Cognitive processes in choice and decision behavior*. Hillsdale, N.J.: Erlbaum, 1980. (19, 33)

Fischhoff, B., Slovic, P., Lichtenstein, S., Read, S., & Combs, B. How safe is safe enough? A psychometric study of attitudes towards technological risks and benefits. *Policy Sciences*, 1978, 9, 127-152. (33)

Fiske, S.T., Taylor, S.E., Etcoff, N., & Laufer, J. Imaging, empathy and causal attribution. *Journal of Experimental Social Psychology*, 1979, 15, 356-377. (13)

Fitch, G. Effects of self-esteem, perceived performance, and chance on causal attributions. *Journal of Personality and Social Psychology*, 1970, 16, 311-315. (9)

Florovsky, G. The study of the past. In R. H. Nash (Ed.), *Ideas of history* (Vol. 2). New York: Dutton, 1969. (23)

Forrester, J. W. *World dynamics*. Cambridge: Wright-Allen, 1971. (11)

Foss, D. J., & Blank, M. A. Identifying the speech codes. *Cognitive Psychology*, 1980, 12, 1-31. (35)

Freize, I., & Weiner, B. Cue utilization and attributional judgments for success and failure. *Journal of Personality*, 1971, 39, 591-606. (9)

Friedman, A.K., Ashovitz, S.I., Berger, S.M., Dodd, M.D., Fisher, M.S., Lapayowker, M.S., Moore, J.P., Parlee, D.E., Stein, G.N., & Pendergrass, E.P. A cooperative evaluation of mammography in seven teaching hospitals. *Radiology*, 1966, 86, 886-891. (18)

Fryback, D.G. Use of radiologists' subjective probability estimates in a medical decision making problem. (Michigan Mathematical Psychology Program, Report No. 74-14). Ann Arbor: University of Michigan, 1974. (22)

Furby, L. Interpreting regression toward the mean in developmental research. *Developmental Psychology*, 1973, 8, 172-179. (23)

Galbraith, R.C., & Underwood, B.J. Perceived frequency of concrete and abstract words. *Memory and Cognition*, 1973, 1, 56-60. (1)

Gallic, W.B. *Philosophy and the historical understanding*. London; Chatto & Windus, 1964. (23)

Garcia, J., McGowan, B.K., & Greene, K.F. Sensory quality and integration: Constraints on conditioning. In A.H. Black & W.F. Prokasy (Eds.), *Classical conditioning II: Current research and theory*. New York: Appleton-Century-Crofts, 1972. (15)

Gardiner, P.C., & Edwards, W. Public values: Multi-attribute-utility measurement for social decision making. In M.F. Kaplan & S. Schwartz, (Eds.), *Human judgment and decision processes*. New York: Academic Press, 1975. (28)

Garner, W.R. Good patterns have few alternatives. *American Scientist*, 1970, 58(1), 34-43. (3)

Gershon-Cohen, J., & Borden, A.G.G. Biopsy and mammography. *New York State Journal of Medicine*, 1964, 2751-2756. (18)

Gettys, C.F., & Willke, T.A. The application of Bayes' theorem when the true data state is uncertain. *Organizational Behavior and Human Performance*, 1969, 4, 125-141. (26)

Ginosar, Z., & Trope, Y. The effects of base rates and individuating information on judgments about another person. *Journal of Experimental Social Psychology*, 1980, 16, 228-242. (10)

Glanzer, M., & Clark, W.H. Accuracy of perceptual recall: An analysis of organization. *Journal of Verbal Learning and Verbal Behavior*, 1963, 1, 289-299. (3)

Goethals, G.R., & Reckman, R.F. The perception of consistency in attitudes. *Journal of Experimental Social Psychology*, 1973, 9, 491-501, (32, 34)

Goffman, E. *Interaction ritual*. New York: Anchor, 1967. (16)

Gold, R.H. The accuracy of mammography in the diagnosis of benign and malignant disease. *Oncology*, 1969, 23, 159-163. (18)

Goldberg, L.R. The effectiveness of clinician's judgments: The diagnosis of organic brain damage from the Bender-Gestalt Test. *Journal of Consulting Psychology*, 1959, 23, 25-33. (20)

Goldberg, L.R. Diagnosticians vs. diagnostic signs: The diagnosis of psychosis vs. neurosis from the MMPI. *Psychological Monographs*, 1965, 79. (28)

Goldberg, L.R. Seer over sign: The first "good" example? *Journal of Experimental Research in Personality*, 1968, 3, 168-171. (28) (a)

Goldberg, L.R. Simple models or simple processes? Some research on clinical judgments. *American Psychologist*, 1968, 23, 483-496. (23, 34) (b)

Goldberg, L.R. Man vs. model of man: A rationale, plus some evidence, for a method of improving on clinical inferences. *Psychological Bulletin*, 1970, 73, 422-432. (23, 28)

Goldberg, L.R. Parameters of personality inventory construction and utilization comparison of prediction strategies and tactics. *Multivariate Behavioral Research Monographs*, 1972, 72, 1-59. (28)

Goldberg, L.R. Man vs. model of man: Just how conflicting is that evidence? *Organizational Behavior and Human Performance*, 1976, 16, 13-22. (28)

Goldberg, P. Are women prejudiced against women? *Trans-Action*, April 1968, 28-30. (15)

Golding, S.L., & Rorer, L.G. Illusory correlation and the learning of clinical judgment. *Journal of Abnormal Psychology*, 1972, 80, 249-260. (15)

Goldman, A.I. Epistemics: The regulative theory of cognition. *Journal of Philosophy*, 1978, 75, 509-523. (32)

- Goodfellow, L. D. A psychological interpretation of the results of the Zenith radio experiments in telepathy. *Journal of Experimental Psychology*, 1938, 23, 601-631 (3)
- Gray, C.W. Predicting with intuitive correlations. *Psychonomic Science*, 1968, 11, 41-42. (4)
- Green, B.F., Jr. Parameter sensitivity in multivariate methods. *Multivariate Behavioral Research*, 1977, 3, 263. (28)
- Green, D.M., & Swets, J.A. *Signal detection theory and psychophysics*. New York: Wiley, 1966. (22)
- Greenwald, A. G. The totalitarian ego; Fabrication and revision of personal history. Unpublished manuscript, 1978. (12)
- Greenwald, A.G., & Albert, R.D. Acceptance and recall of improvised arguments. *Journal of Personality and Social Psychology*, 1968, 8, 31-34. (12)
- Greenwald, A.G., & Sakumura, J.S. Attitude and selective learning: Where are the phenomena of yesteryear? *Journal of Personality and Social Psychology*, 1957, 7, 387-397. (12)
- Grether, D.M. Bayes' rule as a descriptive model: The representativeness heuristic (Social Science Working Paper No. 245). Pasadena: California Institute of Technology, 1979. (34)
- Grether, D.M., & Plott, C.R. Economic theory of choice and the preference reversal phenomenon. *American Economic Review*, 1979, 69, 623-638. (19, 33,34)
- Grice, H.P. Logic and conversation. In D. Davidson and G. Harman (Eds.), *The logic of grammar*. Encino: Dickenson, 1975. (34)
- Gross, A. Evaluation of the target person in a social influence situation. Unpublished doctoral dissertation, Stanford University, 1966. (9)
- Gruneberg, M.M., Morris, P.E., & Sykes, R.N. (Eds.), *Practical aspects of memory*. London: Academic Press, 1978. (31)
- Guilford, J.P. *Psychometric methods*. New York: McGraw-Hill, 1954. (4)
- Hacking, I. *The emergence of probability*. London: Cambridge University Press, 1975. (32)
- Hamill, R., Wilson, T.D., & Nisbett, R.E. Insensitivity to sample bias: Generalizing from atypical cases. *Journal of Personality and Social Psychology*, 1980, 39, 578-589. (9, 32, 34)
- Hamilton, D.L. A cognitive attributional analysis of stereotyping. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 12). New York: Academic Press, 1979. (9)
- Hamilton, D.L., & Rose, T.R. Illusory correlation and the maintenance of stereotype beliefs. Unpublished manuscript, University of California at Santa Barbara, 1978. (13)
- Hamlin, R.M. The clinician as judge: Implications of a series of studies. *Journal of Consulting Psychology*, 1954, 18, 233-238. (20)
- Hammerton, M. A case of radical probability estimation. *Journal of Experimental Psychology*, 1973, 202, 252-254. (10)

Hammond, K.R. Probabilistic functioning and the clinical method. *Psychological Review*, 1955, 62, 255-262. (28)

Hammond, K.R. Toward increasing competence of thought in public policy formation. In K.R. Hammond (Ed.), *Judgment and decision in public policy formation*. Boulder, Colo.: Westview Press, 1978. (19)

Hammond, K.R., & Adelman, L. Science, values, and human judgment. *Science*, 1976, 194, 389-396. (28)

Hammond, K.R., McClelland, G. H., & Mumpower, J. *Human judgment and decision making: Theories, methods and procedures*. New York: Praeger, 1980. (34)

Hansen, R.D., & Donoghue, J.M. The power of consensus: Information derived from one's own and others' behavior. *Journal of Personality and Social Psychology*, 1977, 35, 294-302. (10)

Hansen, R.D., & Lowe, C.A. Distinctiveness and consensus: The influence of behavioral information on actors' and observers' attributions. *Journal of Personality and Social Psychology*, 1976, 34, 425-433. (10)

Hart, J.T. Memory and the memory-monitoring process. *Journal of Verbal Learning and Verbal Behavior*, 1967, 6, 686-691. (11)

Hartshorne, H., & May, M.A. *Studies in the nature of character* (Vol. 1), *Studies in deceit*. New York: Macmillan, 1928. (15)

Hastie, R., & Kumar, P.A. Person memory: Personality traits as organizing principles in memory for behavior. *Journal of Personality and Social Psychology*, 1979, 37, 25-38. (9)

Hathaway, S.R. Clinical intuition and inferential accuracy. *Journal of Personality*, 1956, 24, 223-250. (20)

Hayes, J.R., & Simon, H.A. Psychological differences among problem isomorphs. In N.J. Castellan, Jr., D.B. Pisoni, & G.R. Potts (Eds.), *Cognitive theory*. Hillsdale, N.J.: Erlbaum, 1977. Vol. 2. (34)

Hazard, T.H., & Peterson, C.R. Odds versus probabilities for categorical events (Tech. Rep. (73-2). McLean, Va.: Decisions and Designs, Inc., 1973. (22)

Hedges, L.V., & Olkin, I. Vote counting methods on research synthesis. *Psychological Bulletin*, 1960, 88, 359-369. (31)

Heider, F. Social perception and phenomenal causality. *Psychological Review*, 1944, 51, 358-373. (9)

Heider, F. *The psychology of interpersonal relations*. New York: Wiley, 1958. (8, 9)

Hendrick, I. The discussion of the "Instinct to Master." *Psychoanalytic Quarterly*, 1943, 12, 561-565. (16)

Henrion, M. *Assessing probabilities: A review*. Pittsburgh, Pa.: Carnegie-Mellon University, Department of Engineering and Public Policy, 1980. (31)

Henslin, J.M. Craps and magic. *American Journal of Sociology*, 1967, 73, 316-330. (16)

Herrnstein, R.J. On the law of effect. *Journal of the Experimental Analysis of Behavior*, 1970, 13, 243-266. (35)

Hershey, J.C., & Schoemaker, P.J.H. Risk taking and problem context in the domain of losses: An expected utility analysis. *Journal of Risk and Insurance*, 1980, 47, 111-132. (33)

Hession, E., & McCarthy, E. Human performance in assessing subjective probability distributions. Unpublished manuscript, September, 1974. (Available from Department of Business Administration, University College Dublin, Belfield, Dublin 4, Ireland.) (22)

Hexter, J.H. *The history primer*. New York: Basic Books, 1971. (23)

Hoerl, A.E., & Fallin, H.K. Reliability of subjective evaluations in a high incentive situation. *Journal of the Royal Statistical Society (Series A)*, 1974, 337, 227-230. (22)

Hoffman, P.J. The paramorphic representation of clinical judgment. *Psychological Bulletin*, 1960, 57, 116-131. (23, 28)

Hogarth, R.M. Cognitive processes and the assessment of subjective probability distributions. *Journal of the American Statistical Association*, 1975, 70, 271-289. (19, 31)

Hohenemser, K.H. The failsafe risk. *Environment*, 1975, 17(1), 6-10. (33)

Holdren, J.P. The nuclear controversy and the limitations of decision making by experts. *Bulletin of the Atomic Scientists*, 1976, 32, 20-22, (33)

Holmes, D. S. Dimensions of projection. *Psychological Bulletin*, 1968, 69, 248-268. (9)

Holt, R.R. Yet another look at clinical and statistical prediction. *American Psychologist*, 1970, 25, 337-339. (28)

Horowitz, L.M., Norman, S.A., & Day, R.S. Availability and associative symmetry. *Psychological Review*, 1966, 73, 1-15. (11)

Hovland, C.I. Reconciling conflicting results derived from experimental and survey studies of attitude change. *American Psychologist*, 1959, 14, 8-17, (8)

Hovland, C.I., Mandell, W., Campbell, E.H., & Brock, T., Luchins, A.S., Cohen, A.R., McGuire, W.J., Janis, I.L., Feierabend, R.L., & Anderson, N.H. The effects of "commitment" on opinion change following communication. In C.I. Hovland et al., (Eds.), *The order of presentation in persuasion*. New Haven: Yale University Press, 1957. (9)

Howard, J.W., & Dawes, R.M. Linear prediction of marital happiness. *Personality and Social Psychology Bulletin*, 1976, 2, 478-480. (28)

Howell, W.C., & Burnett, S.A. Uncertainty measurement: A cognitive taxonomy. *Organizational Behavior and Human Performance*, 1978, 22, 45-68. (31, 35)

Howell, W.C., Gettys, C., & Martin, D. On the allocation of inference functions in decision systems. *Organizational Behavior and Human Performance*, 1971, 6, 132-149. (26)

Huntsberger, D.V. *Elements of statistics* (2nd ed.). Boston: Allyn & Bacon, 1967. (15)

Hynes, M., & Vanmarcke, E. Reliability of embankment performance predictions. *Proceedings of the ASCE Engineering Mechanics Division Specialty*

Conference. Waterloo, Ontario, Canada: University of Waterloo Press, 1976. (31, 33)

Ittelson, W.H., & Kilpatrick, F.P. Experiments in perception. *Scientific American*, 1951, 185, 50-55. (35)

Janis, I. *Victims of groupthink*. Boston: Houghton Mifflin, 1972. (8, 23, 34)

Janis, I.L., & Mann, L. *Decision making*. New York: Free Press, 1977. (34)

Jenkins, H.M., & Ward, W.C. Judgment of contingency between responses and outcomes. *Psychological Monographs*, 1965, 79. (11, 15, 19)

Jennergren, L.P., & Keeney, R.L. Risk assessment. In *Handbook of applied systems analysis*. Laxenberg, Austria: IIASA, in press. (31, 33)

Jennings, D.L., Lepper, M.R., & Ross, L. Persistence of impressions of personal persuasiveness: Perseverance of erroneous self assessments outside the debriefing para-digm. Unpublished manuscript, Stanford University, 1980. (9, 15)

Jervis, R. *Perception and misperception in international relations*. Princeton: Princeton University Press, 1975. (8, 34)

Johnson, D.M. *Systematic introduction to the psychology of thinking*. New York: Harper & Row, 1972. (4)

Johnson, T.J., Feigenbaum, R., & Weiby, M. Some determinants and consequences of the teacher's perception of causation, *Journal of Experimental Psychology*, 1964, 55, 237-246. (9)

Johnson-Laird, P.N., Legrenzi, P., & Sonino-Legrenzi, M. Reasoning and a sense of reality. *British Journal of Psychology*, 1972, 63, 395-400. (6, 34)

Johnson-Laird, P.N., & Wason, P.C. A theoretical analysis of insight into a reasoning task. In P.N. Johnson-Laird and P.C. Wason (Eds.), *Thinking*. Cambridge: Cambridge University Press, 1977. (6, 34)

Jones, E.E., & Davis, K.E. From acts to dispositions: The attribution process in person perception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2). New York: Academic Press, 1965. (9, 13)

Jones, E.E., Davis, K.E., & Gergen, K. Role playing variations and their informational value for person perception. *Journal of Abnormal and Social Psychology*, 1961, 63, 302-310. (9)

Jones, E.E., & DeCharms, R. Changes in social perception as a function of the personal relevance of behavior. *Sociometry*, 1957, 20, 75-85. (9)

Jones, E.E., & Harris, V.A. The attribution of attitudes. *Journal of Experimental Social Psychology*, 1967, 3, 1-24. (9)

Jones, E.E., Kanouse, D.E., Kelley, H.H., Nisbett, R.E., Valins, S., & Weiner, B. *Attribution: Perceiving the causes of behavior*. Morristown, N.J.: General Learning Press, 1971. (8, 9, 15)

Jones, E.E., & Nisbett, R.E. The actor and the observer: Divergent perceptions of the causes of behavior. In E.E. Jones, D. Kanouse, H.H. Kelley, R.E. Nisbett, S. Valins, & B. Weiner, *Attribution: Perceiving the causes of behavior*. Morristown, N.J.: General Learning Press, 1971. (9, 11, 12)

Jones, J.M. *Prejudice and racism*. Reading, Mass.: Addison-Wesley, 1972. (15)

Jones, M.R. From probability learning to sequential processing: A critical

review. *Psychological Bulletin*, 1971, 76, 153-185. (3)

Kadane, J. Personal communication, November 1980. (22)

Kahneman, D., & Tversky, A. On prediction and judgment. *ORI Research Monograph*, 1972, 12(4). (10) (a)

Kahneman, D., & Tversky, A. Subjective probability: A judgment of representativeness. *Cognitive Psychology*, 1972, 3, 430-454. (1, 4, 5, 6, 11, 23, 32, 34) (b)

Kahneman, D., & Tversky, A. On the psychology of prediction. *Psychological Review*, 1973, 80, 237-251. (1, 5, 6, 7, 8, 9, 10, 13, 15, 18, 23, 27, 30, 34)

Kahneman, D., & Tversky, A. Intuitive prediction: Biases and corrective procedures. *TIMS Studies in Management Science*, 1979, 12, 313-327. (8, 33, 34, 35) (a)

Kahneman, D., & Tversky, A. Prospect theory: An analysis of decisions under risk. *Econometrica*, 1979, 47, 263-291. (19, 33) (b)

Kasper, R. G. "Real" vs. perceived risk: Implications for policy. In G.T. Goodman & W.D. Rowe (Eds.), *Energy risk management*. London: Academic Press, 1979. (33)

Kassin, S.M. Base rates and prediction: The role of sample size. *Personality and Social Psychology Bulletin*, 1979, 5, 210-213. (10) (a)

Kassin, S. M. Consensus information, prediction, and causal attribution: A review of the literature and issues. *Journal of Personality and Social Psychology*, 1979, 37, 1966-1981. (10) (b)

Kates, R. W. Hazard and choice perception in flood plain management (Research Paper No. 78). Chicago: University of Chicago, Department of Geography, 1962. (23, 33)

Katz, D., & Allport, F. *Students' attitudes*. Syracuse: Craftsman Press, 1931. (9)

Kaufman, M.T. *The New York Times*, October 18, 1973. (7)

Kelley, H. H. Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska Symposium on Motivation: 1967*. Lincoln: University of Nebraska Press, 1967, 192-241. (7, 9, 13, 15)

Kelley, H.H. Attribution in social interaction. In E.E. Jones, D. Kanouse, H.H. Kelley, R.E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior*. Morristown, N.J.: General Learning Press, 1971. (9, 15)

Kelley, H.H. The process of causal attribution. *American Psychologist*, 1973, 28, 107-128. (9, 15)

Kelley, H.H., & Stahelski, A.J. Social interaction basis of cooperators' and competitors' beliefs about others. *Journal of Personality and Social Psychology*, 1970, 16, 66-91. (9)

Kelly, L. Evaluation of the interview as a selection technique. In *Proceedings of the 1953 Invitational Conference on Testing Problems*. Princeton, N.J.: Educational Testing Service, 1954. (28)

Kelman, H.C. The human use of human subjects: The problem of deception in social psychological experiments. In A.G. Miller (Ed.), *The social psychology of psychological research*. New York: Macmillan, 1972. (9)

- Kidd, J.B. The utilization of subjective probabilities in production planning. *Acta Psychologica*, 1970, 34, 338-347. (30)
- Kilpatrick, F.P. *Explorations in transactional psychology*. New York: New York University Press, 1961. (35)
- King, J.F., Zechmeister, E.B., & Shaughnessy, J.J. Judgment of knowing: The influence of retrieval practice. *American Journal of Psychology*, in press. (31)
- Koltuv, B. Some characteristics of intrajudge trait intercorrelations. *Psychological Monographs*, 1962, 76. (15)
- Koriat, A., Lichtenstein, S., & Fischhoff, B. Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory*, 1980, 6, 107-118. (22, 23, 31)
- Kostlan, A.A method for the empirical study of psychodiagnosis. *Journal of Consulting Psychology*, 1954, 18, 83-88. (20)
- Krantz, D.H. Measurement structures and psychological laws. *Science*, 1972, 175, 1427-1435. (28)
- Krantz, D.H., Luce, R.D., Suppes, P., & Tversky, A. *Foundations of measurement (Vol. 1)*. New York: Academic Press, 1971. (28)
- Kruskal, W., & Mosteller, F. Representative sampling I: Non-scientific literature. *International Statistical Review*, 1979, 47, 13-24. (6) (a)
- Kruskal, W., & Mosteller, F. Representative sampling II: Scientific literature excluding statistics. *International Statistical Review*, 1979, 47, 111-127. (6) (b)
- Kuhn, T. *The structure of scientific revolution*. Chicago: University of Chicago Press, 1962. (8, 23)
- Kushner, R. Breast cancer- "The night I found out." *San Francisco Chronicle*, March 24, 1976. (18)
- Lakatos, I. Falsification and scientific research programmes. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of scientific knowledge*. Cambridge: Cambridge University Press, 1970. (23)
- Langer, E.J., Taylor, S.E., Fiske, S.T., & Chanowitz, B. Stigma, staring and discomfort: A novel stimulus hypothesis. *Journal of Experimental Social Psychology*, 1976, 72, 451-463. (13)
- Lanir, Z. *Critical reevaluation of the strategic intelligence methodology*. Tel Aviv: Center for Strategic Studies, Tel Aviv University, 1978. (23)
- Larkin, J., McDermott, J., Simon, D.P., & Simon, H.A. Expert and novice performance in solving physics problems. *Science*, 1980, 208, 1335-1342. (34)
- Larson, J.R., & Reenan, A.M. The equivalence interval as a measure of uncertainty. *Organizational Behavior and Human Performance*, 1979, 23, 49-55. (22, 31)
- Lepper, M.R., & Greene, D. Turning play into work: Effects of adult surveillance and extrinsic rewards on children's intrinsic motivation. *Journal of Personality and Social Psychology*, 1975, 31, 479-486. (9)
- Lepper, M.R., & Greene, D. *The hidden costs of reward*. Hillsdale, N.J.: Erlbaum, 1978. (9)
- Lepper, M.R., Greene, D., & Nisbett, R.E. Undermining children's intrinsic

interest with extrinsic reward: A test of the overjustification hypothesis. *Journal of Personality and Social Psychology*, 1973, 28, 129-137. (9)

Lepper, M.R., Ross, L., & Lau, R. Persistence of inaccurate and discredited personal impressions: A field demonstration of attributional perseverance. Unpublished manuscript, Stanford University, 1979. (9)

Lesnick, G.J. Mammography: A word of caution. *New York State journal of Medicine*, 1966, 2005-2008. (18)

Libby, R. Man versus model of man: Some conflicting evidence. *Organizational Behavior and Human Performance*, 1976, 16, 1-12. (28)

Lichtenstein, S., & Fischhoff, B. Do those who know more also know more about how much they know? The calibration of probability judgments. *Organizational Behavior and Human Performance*, 1977, 20, 159-183. (22, 31)

Lichtenstein, S., & Fischhoff, B. How well do probability experts assess probability? (Decision Research Report 80-5). Eugene, Oreg.: Decision Research, 1980. (22) (a)

Lichtenstein, S., & Fischhoff, B. Training for calibration. *Organizational Behavior and Human Performance*, 1980, 26, 149-171. (22, 31) (b)

Lichtenstein, S., & Fischhoff, B. The effects of gender and instructions on Calibration (Decision Research Report 81-5). Eugene, Oreg.: Decision Research 1981. (22)

Lichtenstein, S., Fischhoff, B., & Phillips, L.D. Calibration of probabilities: The state of the art. In H. Jungermann & G. deZeeuw (Eds.), *Decision making and change in human affairs*. Amsterdam: D. Reidel, 1977. (33)

Lichtenstein, S., & Slovic, P. Reversal of preferences between bids and choices in gambling decisions. *Journal of Experimental Psychology*, 1971, 89, 46-55. (19, 33, 34)

Lichtenstein, S., & Slovic, P. Response-induced reversals of preference in gambling; An extended replication in Las Vegas. *Journal of Experimental Psychology*, 1973, 101, 16-20. (33, 34)

Lichtenstein, S., Slovic, P., Fischhoff, B., Layman, M., & Combs, B. Judged frequency of lethal events. *Journal of Experimental Psychology. Human Learning and Memory*, 1978, 4, 551-578. (33)

Lindman, H.G., & Edwards, W. Supplementary report: Learning the gambler's fallacy. *Journal of Experimental Psychology*, 1961, 62, 630. (23)

Loftus, E.F. *Eyewitness testimony*. Cambridge, Mass.: Harvard University Press, 1979. (34)

Loftus, E.F., & Palmer, J.C. Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior*, 1974, 16, 585-589. (34)

Lord, C., Lepper, M.R., & Ross, L. Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, 1979, 37, 2098-2110. (9, 15)

Ludke, R.L., Stauss, F.F., & Gustafson, D. H. Comparison of five methods for estimating subjective probability distributions. *Organizational Behavior and Human Performance*, 1977, 19, 162-179. (31)

Lusted, L. B. Introduction to medical decision making. Springfield, 111.: Charles C Thomas, 1968. (18)

Lusted, L.B. A study of the efficacy of diagnostic radiologic procedures: Final report on diagnostic efficacy. Chicago: Efficacy Study Committee of the American College of Radiology, 1977. (22,31)

Lusted, L.B., Bell, R.S., Edwards, W., Roberts, H.V., & Wallace, D.L. Evaluating the efficacy of radiologic procedures by Bayesian methods: A progress report. In K.J. Snapper (Ed.). Models in metrics for decision makers. Washington, D.C.: Information Resources Press, 1977. (18)

Lynn, R. Attention, arousal and the orientation reaction. Oxford: Pergamon, 1966. (35)

Lyon, D., & Slovic, P. Dominance of accuracy information and neglect of base rates in probability estimation. *Acta Psychologica*, 1976, 40, 287-298. (10, 18)

Lyons, A.S. Aids to the surgeon by mammography. *Mt. Sinai Journal of Medicine*, New York, 1975, 42, 223-231. (18)

Machover, K. Personality projection in the drawing of the human figure. Springfield, 111.: Charles C Thomas, 1949. (17)

Malpass, R.S. Effects of attitude on learning and memory. The influence of instruction-induced sets. *Journal of Experimental Social Psychology*, 1969, 5, 441-453. (12)

Manis, M., Dovalina, I., Avis, N.E., & Cardoze, S. Base rates can affect individual predictions. *Journal of Personality and Social Psychology*, 1980, 38, 231-248. (10)

Markman, E.M. Realizing that you don't understand: Elementary school children's awareness of inconsistencies. *Child Development*, 1979, 50, 643-655. (31)

Markovic, M. Social determinism and freedom. In H.E. Kiefer & M.K. Munitz, (Eds.), *Mind, science and history*. Albany, N.Y.: State University of New York Press, 1970. (23)

Marquardt, D.W., & Snee, R.D. Ridge regression in practice. *American Statistician*, 1975, 29(1), 3-19. (28)

Marwick, A. *The nature of history*. London: Macmillan, 1970. (23)

Mayzner, M.S., & Tresselt, M.E. Tables of single-letter and bigram frequency counts for various word-length and letter-position combinations. *Psychonomic Monograph Supplements*, 1965, 1, 13-32. (11)

McArthur, C. Analyzing the clinical process. *Journal of Counseling Psychology*, 1954, 1, 203-207. (20)

McArthur, L.Z. The how and what of why: Some determinants and consequences of causal attribution. *Journal of Personality and Social Psychology*, 1972, 22, 171-193. (7, 9)

McArthur, L.Z. The lesser influence of consensus than distinctiveness information on causal attributions: A test of the person-thing hypothesis. *Journal of Personality and Social Psychology*, 1976, 33, 733-742. (9)

McArthur, L.Z., & Post, D. Figural emphasis and person perception. *Journal of Experimental Social Psychology*, 1977, 13, 520-535. (9)

McArthur, L.Z., & Solomon, L.K. Perceptions of an aggressive encounter as a function of the victim's salience and the perceiver's arousal. *Journal of Personality and Social Psychology*, 1978, 36, 1278-1290. (9)

McCauley, C., & Stitt, C.L. An individual and quantitative measure of stereotypes. *Journal of Personality and Social Psychology*, 1978, 36, 929-940. (10)

McClelland, G., & Rohrbaugh, J. Who accepts the Pareto axiom? The role of utility and equity in arbitration decisions. *Behavioral Science*, 1978, 23, 446-456. (34)

McClow, M.V., & Williams, A.C. Mammographic examinations (4030): Ten-year clinical experience in a community medical center. *Annals of Surgery*, 1973, 177, 616-619. (18)

McGrath, P. . Radioactive waste management: Potentials and hazards from a risk point of view (Report EURFNR 1204 [KFK 1992]). Karlsruhe, West Germany: U.S.-EURATOM Fast Reactor Exchange Programme, 1974. (33)

McNeil, B.J., Weichselbaum, R., & Pauker, S.G. Fallacy of the five-year survival in lung cancer. *New England Journal of Medicine*, 1978, 299, 1397-1401. (33)

Meehl, P.E. Clinical versus statistical prediction: A theoretical analysis and a review of the evidence. Minneapolis: University of Minnesota Press, 1954. (28)

Meehl, P.E. When shall we use our heads instead of the formula? *Journal of Counseling Psychology*, 1957, 4, 268-273. (20)

Meehl, P.E. Seer over sign: The first good example. *Journal of Experimental Research in Personality*, 1965, 1, 27-32. (28)

Meehl, P.E. Nuisance variables and the ex post facto design. In M. Radner & S. Winokur (Eds.), *Minnesota studies in the philosophy of science*. Minneapolis: University of Minnesota Press, 1970. (23)

Meehl, P.E., & Rosen, A. Antecedent probability and the efficacy of psychometric signs, patterns or cutting scores. *Psychological Bulletin*, 1955, 52, 194-216. (10)

Merton, R.K. The self-fulfilling prophecy. *Antioch Review*, 1948, 8, 193-210. (15)

Mervis, C.B., & Rosch, E. Categorization of natural objects. *Annual Review of Psychology*, 1981, 32, 89-115. (6)

Michotte, A. *The perception of causality*. New York: Basic Books, 1963. (19)

Milgram, S. Behavioral study of obedience. *Journal of Abnormal and Social Psychology*, 1963, 67, 371-378. (7, 9)

Mill, J.S. *A system of logic, ratiocinative and inductive*. Toronto: University of Toronto Press, 1974 (Originally published in 1843). (32)

Miller, A.G. *The social psychology of psychological research*. New York: Macmillan, 1972. (9)

Miller, A.G., Gillen, B., Schenker, C., & Radlove, S. Perception of obedience to authority. *Proceedings of the 81st Annual Convention of the American Psychological Association*, 1973, 8, 127-128. (7)

Miller, D.T. What constitutes a self-serving attributional bias? *Journal of*

Personality and Social Psychology, 1978, 36, 1221-1223. (9)

Miller, D.T., & Ross, M. Self-serving biases in the attribution of causality: Fact or fiction? *Psychological Bulletin*, 1975, 82, 213-225. (9)

Mischel, W. *Personality and assessment*. New York: Wiley, 1968. (4, 15)

Mischel, W. Continuity and change in personality. *American Psychologist*, 1969, 24, 1012-1018. (15)

Mischel, W. On the interface of cognition and personality: Beyond the person-situation debate. *American Psychologist*, 1979, 34, 740-754. (6)

Mischel, W., & Gilligan, C. Delay of gratification, motivation for the prohibited gratification, and responses to temptation. *Journal of Personality and Social Psychology*, 1964, 64, 411-417. (15)

Mitchell, T.R., & Kalb, L.S. Outcome knowledge and content as determinants of supervisor attributions and judgments of responsibility about a subordinate's poor performance, *Journal of Applied Psychology*, in press. (31)

Moore, P.G. The manager's struggle with uncertainty. *Journal of the Royal Statistical Society (Series A)*, 1977, 140, 129-165. (31)

Morris, P. A. Decision analysis expert use. *Management Science*, 1974, 20, 1233-1241. (31)

Morrison, R.J. Franklin D. Roosevelt and the Supreme Court: An example of the use of probability theory in political history. *History and Theory*, 1977, 16, 137-146. (23)

Morton, J. Interaction of information in word recognition. *Psychological Review*, 1969, 76, 165-178. (35)

Moskowitz, H., & Bullers, W.I. Modified PERT versus fractile assessment of subjective probability distributions (Paper No. 675). *Purdue University*, 1978. (22)

Murphy, A.H. Scalar and vector partitions of the probability score (Part I). Two-state situation. *Journal of Applied Meteorology*, 1972, 31, 273-282. (22)

Murphy, A.H. A new vector partition of the probability score. *Journal of Applied Meteorology*, 1973, 12, 595-600. (22)

Murphy, A.H. Personal communication, August 1980. (22)

Murphy, A.H., & Winkler, R.L. Forecasters and probability forecasts: Some current problems. *Bulletin of the American Meteorological Society*, 1971, 52, 239-247. (22)

Murphy, A.H., & Winkler, R.L. Subjective probability forecasting experiments in meteorology: Some preliminary results. *Bulletin of the American Meteorological Society*, 1974, 55, 1206-1216. (22, 31)

Murphy, A.H., & Winkler, R.L. Can weather forecasters formulate reliable probability forecasts of precipitation and temperature? *National Weather Digest*, 1977, 2, 2-9. (22, 31)(a)

Murphy, A.H., & Winkler, R.L. The use of credible intervals in temperature forecasting: Some experimental results. In H. Jungermann & G. deZeeuw (Eds.), *Decision making and change in human affairs*. Amsterdam: D. Reidel, 1977. (22)(b)

Naatanen, R., & Summala, H. Road-user behavior and traffic accidents. Amsterdam: North-Holland, 1975. (33)

Nickerson, R.S., & McGoldrick, C.C., Jr. Confidence ratings and level of performance on a judgmental task. *Perceptual Motor Skills*, 1965, 20, 311-316 (22, 31)

Nisbett, R.E., & Borgida, E. Attribution and the psychology of prediction. *Journal of Personality and Social Psychology*, 1975, 32, 932-943. (7, 10, 27, 32)

Nisbett, R.E., Borgida, E., Crandall, R., & Reed, H. Popular induction: Information is not necessarily informative. In J. S. Carroll and J. W. Payne (Eds.), *Cognition and social behavior*. Hillsdale, N.J.: Erlbaum, 1976. (19, 28)

Nisbett, R.E., & Ross, L. *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, N.J.: Prentice-Hall, 1980. (6, 9, 10, 15, 32, 34, 35)

Nisbett, R.E., & Schachter, S. Cognitive manipulations of pain. *Journal of Experimental Social Psychology*, 1966, 2, 227-236. (7)

Nisbett, R.E., & Wilson, T.D. Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 1977, 84, 231-259. (19, 32, 34)

Nisbett, R.E., Zukier, H., & Lemley, R.E. The dilution effect: Nondiagnostic information weakens the implications of diagnostic information. *Cognitive Psychology*, 1981, 13, 248-277. (34)

Norman, W.T., & Goldberg, L.R. Raters, ratees, and randomness in personality structure. *Journal of Personality and Social Psychology*, 1966, 4, 681-691. (15)

Nowell-Smith, P. H. Historical explanation. In H. E. Kiefer & M. K. Munitz (Eds.), *Mind, science and history*. Albany, N.Y.: State University of New York Press, 1970, (23)

O'Leary, M.K., Coplin, W.D., Shapiro, H.B., & Dean, D. The quest for relevance. *International Studies Quarterly*, 1974, IS, 211-237. (23)

Olson, C.. Some apparent violations of the representativeness heuristic in human judgment. *Journal of Experimental Psychology: Human Perception and Performance*, 1976, 2, 599-608. (5, 34)

Orne, M.T. Demand characteristics and their implications for real life: The importance of quasi-controls. In A.G. Miller (Ed.), *The social psychology of psychological research*. New York: Macmillan, 1972. (9)

Orne, M.T. Communication by the total experimental situation: Why it is important, how it is evaluated, and its significance for the ecological validity of findings. In P. Pliner, L. Krames, & T. Alloway (Eds.), *Communication and affect*. New York: Academic Press, 1973. (34)

Oskamp, S. The relationship of clinical experience and training methods to several criteria of clinical prediction. *Psychological Monographs*, 1962, 76, (28, Whole No. 547). (20, 22, 31)

Overall, J.E. Classical statistical hypothesis testing within the context of Bayesian theory. *Psychological Bulletin*, 1969, 71, 285-292. (2)

Parducci, A. Category judgment: A range-frequency model. *Psychological Review*, 1965, 72, 407-418. (34)

Passini, F.T., & Norman, W.T. A universal conception of personality structure? *Journal of Personality and Social Psychology*, 1966, 4, 44-49. (15)

Patterson, C.H. Diagnostic accuracy or diagnostic stereotypy? *Journal of Consulting Psychology*, 1955, 19, 483-485. (20)

Payne, J.W. Information processing theory: Some concepts and methods applied to decision research. In T. Wallsten (Ed.), *Cognitive processes in choice and decision behavior*. Hillsdale, N.J.: Erlbaum, 1980. (19)

Pennington, D.C., Rutter, D.R., McKenna, K., & Morley, I.E. Estimating the outcome of a pregnancy test: Women's judgments in foresight and hindsight. *British Journal of Social and Clinical Psychology*, 1980, 79, 317-323. (31)

Peterson, C.R. (Ed.). Special issue: Cascaded inference. *Organizational Behavior and Human Performance*, 1973, 10, 315-432. (27)

Peterson, C.R., & Beach, L.R. Man as an intuitive statistician. *Psychological Bulletin*, 1967, 68(1), 29-46. (3, 15, 31)

Peterson, C.R., DuCharme, W.M., & Edwards, W. Sampling distributions and probability revisions. *Journal of Experimental Psychology*, 1968, 76, 236-243. (3)

Peterson, C.R., Schneider, R.J., & Miller, A.J. Sample size and the revision of subjective probabilities. *Journal of Experimental Psychology*, 1965, 69, 522-527. (25)

Phillips, L.D. Some components of probabilistic inference (Tech. Rep. No. 1). Human Performance Center, University of Michigan, 1966. (11)

Phillips, L.D., & Edwards, W. Conservatism in simple probability inference tasks. *Journal of Experimental Psychology*, 1966, 72, 346-357. (25)

Phillips, L.D., Hays, W.L., & Edwards, W. Conservatism in complex probabilistic inference. *IEEE Transactions on Human Factors in Electronics*, 1966, HFE-7, 7-18. (25)

Phillips, L.D., & Wright, G.N. Cultural differences in viewing uncertainty and assessing probabilities. In H. Jungermann & G. deZeeuw (Eds.), *Decision making and change in human affairs*. Amsterdam: D. Reidel, 1977. (22, 31)

Pickhardt, R.C., & Wallace, J.B. A study of the performance of subjective probability assessors. *Decision Sciences*, 1974, 5, 347-363. (22, 31)

Pitz, G.F. Subjective probability distributions for imperfectly known quantities. In L.W. Gregg (Ed.), *Knowledge and cognition*. New York: Wiley, 1974. (22, 31)

Pitz, G.F., Downing, L., & Reinhold, H. Sequential effects in the revision of subjective probabilities. *Canadian Journal of Psychology*, 1967, 21, 381-393. (3)

Pokorny, G. Energy development: Attitudes and beliefs at the regional/national levels. Cambridge, Mass.: Cambridge Reports, 1977. (33)

Polefka, J. The perception and evaluation of responses to social influences. Unpublished doctoral dissertation, Stanford University, 1965. (9)

Pollack, L., & Decker, L.R. Confidence ratings, message receptions, and the receiver operating characteristics. *Journal of the Acoustical Society of America*, 1958, 30, 286-292. (22)

Posner, M.I. *Chronometric explorations of mind*. Hillsdale, N.J.: Erlbaum, 1978. (35)

Posner, M.I., & Keele, S.W. On the genesis of abstract ideas. *Journal of Experimental Psychology*, 1968, 77, 353-363. (6)

Poulton, E.C. The new psychophysics: Six models for magnitude estimation. *Psychological Bulletin*, 1968, 69, 1-19. (33)

Poulton, E.C. Range effects in experiments with people. *American Journal of Psychology*, 1975, 88, 3-32. (34)

Pratt, J.W. Personal communication, October 1975. (22)

President's Commission on the Accident at Three Mile Island. Report. Washington, D.C.: U.S. Government Printing Office, 1979. (33)

Primack, J. Nuclear reactor safety: An introduction to the issues. *Bulletin of the Atomic Scientists*, 1975, 31, 15-17. (33)

Pryor, J.B., & Kriss, M. The cognitive dynamics of salience in the attribution process. *Journal of Personality and Social Psychology*, 1977, 35, 49-55. (13)

Raiffa, H. Risk, ambiguity and the Savage axioms: Comment. *Quarterly Journal of Economics*, 1961, 75, 690-694. (35)

Raiffa, H. Assessments of probabilities. Unpublished manuscript, Harvard University, 1969. (22)

Rather, D., & Heskowitz, M. *The camera never blinks*. New York: Ballantine Books, 1977. (12)

Regan, D.T., & Totten, J. Empathy and attribution: Turning observers into actors. *Journal of Personality and Social Psychology*, 1975, 32, 850-856. (9)

Remus, W.E., & Jenicke, L.O. Unit and random linear models in decision making. *Multivariate Behavioral Research*, 1978, 13, 215-221. (28)

Rethans, A. An investigation of consumer perceptions of product hazards. Doctoral dissertation, University of Oregon, 1979. (33)

Rhoads, J. E. *Textbook of surgery*. Philadelphia: Lippincott, 1969. (18)

Rogers, T.B., Kuiper, N.A., & Kirker, W.S. Self-reference and the encoding of personal information. *Journal of Personality and Social Psychology*, 1977, 35, 677-688. (12)

Root, H. E. Probability statements in weather forecasting. *Journal of Applied Meteorology*, 1962, 1, 163-168. (22, 31)

Rosato, F., Thomas, J., & Rosato, E. Operative management of nonpalpable lesions detected by mammography. *Surgery, Gynecology, and Obstetrics*, 1973, 337, 491-493. (18)

Rosch, E. Cognitive reference points. *Cognitive Psychology*, 1975, 7, 532-547. (6, 8)

Rosch, E. Principles of categorization. In E. Rosch and B.B. Floyd (Eds.), *Cognition and categorization*. Hillsdale, N.J.: Erlbaum, 1978. (6)

Rosch, E., & Mervis, C.B. Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 1975, 7, 573-605. (6)

Rosenberg, J. A question of ethics: The DNA controversy. *American Educator*, 1978, 2, 27-30. (33)

Rosenhan, D.L. On being sane in insane places. *Science*, 1973, 279, 250-258. (15)

Rosenthal, R., & Jacobson, L. *Pygmalion in the classroom: Teacher*

expectation and pupils' intellectual development. New York: Holt, Rinehart & Winston, 1968. (9, 15)

Ross, L. The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 10). New York: Academic Press, 1977, 174-177. (8, 9, 13, 15, 19, 30, 32)

Ross, L. Some afterthoughts on the intuitive psychologist. In L. Berkowitz (Ed.), *Cognitive theories in social psychology*. New York: Academic Press, 1978. (15)

Ross, L., Amabile, T.M., & Steinmetz, J.L. Social roles, social control, and biases in social perception processes. *Journal of Personality and Social Psychology*, 1977, 35, 485-494. (9)

Ross, L., Bierbrauer, G., & Polly, S. Attribution of educational outcomes by professional and non-professional instructors. *Journal of Personality and Social Psychology*, 1974, 29, 609-618. (9)

Ross, L., Greene, D., & House, P. The false consensus phenomenon: An attributional bias in self perception and social perception processes. *Journal of Experimental Social Psychology*, 1977, 13, 279-301. (9)

Ross, L., & Lepper, M.R. The perseverance of beliefs: Empirical and normative considerations. In R.A. Shweder (Ed.), *New directions for methodology of behavioral sciences: Fallible judgment in behavioral research*. San Francisco: Jossey-Bass, 1980. (9, 34)

Ross, L., Lepper, M.R., & Hubbard, M. Perseverance in self perception and social perception: Biased attributional processes in the debriefing paradigm. *Journal of Personality and Social Psychology*, 1975, 32, 880-892. (9, 15)

Ross, L., Lepper, M.R., Strack, F., & Steinmetz, J.L. Social explanation and social expectation: The effects of real and hypothetical explanations upon subjective likelihood. *Journal of Personality and Social Psychology*, 1977, 35, 817-829. (9,15)

Ross, M., & Sicoly, F. Egocentric biases in availability and attribution. *Journal of Personality and Social Psychology*, 1979, 37, 322-336. (13)

Rothbart, M., Fulero, S., Jensen, C., Howard, J., & Birrell, P. From individual to group impressions: Availability heuristics in stereotype formation. *Journal of Experimental Social Psychology*, 1978, 14, 237-255. (13)

Rothbart, M., & Snyder, M. Confidence in the prediction and postdiction of an uncertain event. *Canadian Journal of Behavioral Science*, 1970, 2, 38-43. (35)

Rothschild, N. Coming to grips with risk. Address presented on BBC television, November 1978. (Reprinted in the *Wall Street Journal*, May 13, 1979.) (33)

Rumelhart, D. E. Schemata: The building blocks of cognition. In R. Spiro, B. Bruce, and W. Brewer (Eds.), *Theoretical issues in reading comprehension*. Hillsdale, N.J.: Erlbaum, 1979. (34)

Russell, B. *Philosophy*. New York: Norton, 1927. (7)

Russo, J.E. The value of unit price information. *Journal of Marketing Research*, 1977, 11, 193-201. (19)

Sanders, F. The evaluation of subjective probability forecasts (Scientific Report No. 5). Cambridge: Massachusetts Institute of Technology, Department of Meteorology, 1958. (22)

Savage, L.J. The foundations of statistics. New York: Wiley, 1954. (1, 25)

Savage, L.J. Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association*, 1971, 66, 336, 783-801. (22)

Sawyer, J. Measurement and prediction, clinical and statistical. *Psychological Bulletin*, 1966, 66, 178-200. (28)

Schachter, S., & Singer, J.E. Cognitive, social, and physiological determinants of emotional state. *Psychological Review*, 1962, 69, 379-399. (7)

Schaefer, R.E., & Borcharding, K. The assessment of subjective probability distribution: A training experiment. *Acta Psychologica*, 1973, 37, 117-129. (22, 31)

Schmidt, F.L. The relative efficiency of regression and simple unit predictor weights in applied differential psychology. *Educational and Psychological Measurement*, 1971, 31, 669-714. (28)

Schoemaker, P.J.H., & Kunreuther, H.C. An experimental study of insurance decisions. *Journal of Risk and Insurance*, 1979, 46, 603-618. (33)

Schum, D. Current developments in research and cascaded inference processes. In T. Wallsten (Ed.), *Cognitive processes in choice and decision behavior*. Hillsdale, N.J.: Erlbaum, 1980. (19)

Seaver, D.A., von Winterfeldt, D., & Edwards, W. Eliciting subjective probability distributions on continuous variables. *Organizational Behavior and Human Performance*, 1978, 21, 379-391. (22, 31)

Selvidge, J. Experimental comparison of different methods for assessing the extremes of probability distributions by the fractile method. (Management Science Report Series, Report 75-13). Boulder, Colo. Graduate School of Business Administration, University of Colorado, 1975. (22)

Selvidge, J. Assessing the extremes of probability distributions by the fractile method. *Decision Sciences*, 1980, II, 493-502. (31)

Shafer, G. A mathematical theory of evidence. Princeton, N.J.: Princeton University Press, 1976. (6, 35)

Shapiro, S., Strax, P., & Venet, L. Periodic breast cancer screening. *Archives of Environmental Health*, 1967, 15, 547-553. (18)

Shapiro, S., Strax, P., & Venet, L. Periodic breast cancer screening in reducing mortality from breast cancer. *Journal of American Medical Association*, 1971, 215, 1777-1785. (18)

Sheridan, T.B. Human error in nuclear power plants. *Technology Review*, February 1980, 23-33. (33)

Shweder, R.A. (Ed.), *New directions for methodology of behavioral sciences: Fallible judgment in behavioral research*. San Francisco: Jossey-Bass, 1980. (34)

Shweder, R.A., & D'Andrade, R.G. The systematic distortion hypothesis. In R.A. Shweder (Ed.), *Fallible judgment in behavioral research*. San Francisco: Jossey-Bass, 1980. (6)

Sieber, J.E. Effects of decision importance on ability to generate warranted subjective uncertainty. *Journal of Personality and Social Psychology*, 1974, 30, 688-694. (22,31)

Siegel, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956. (5)

Siegler, R.S. The origins of scientific reasoning. In R.S. Siegler (Ed.), *Children's thinking: What develops?* Hillsdale, N.J.: Erlbaum, 1979. (19)

Silverman, I. Motives underlying the behavior of the subject in the psychological experiment. Paper presented at the meeting of the American Psychological Association, Chicago, September 1965. (9)

Simon, H. *Models of man: Social and rational*. New York: Wiley, 1957. (13)

Simon, H.A. Information-processing theory of human problem solving. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes* (Vol. 5). Hillsdale, N.J.: Erlbaum, 1978.

(19) Simon, H.A., & Chase, W.G. Skill in chess. *American Scientist*, 1973, 63, 394-403. (28)

Simon, H.A., & Hayes, J.R. The understanding process: Problem isomorphs. *Cognitive Psychology*, 1976, 8, 165-190. (19)

Simon, H. A., & Newell, A. Human problem solving: The state of the theory in 1970. *American Psychologist*, 1971, 26, 145-159. (19)

Slovic, P. Cue consistency and cue utilization in judgment. *American Journal of Psychology*, 1966, 79, 427-434. (4)

Slovic, P. From Shakespeare to Simon: Speculations – and some evidence – about man's ability to process information. *Oregon Research Institute Research Bulletin*, 1972, 12(2). Available from Decision Research, Eugene, Oregon (8, 22) (a)

Slovic, P. Limitations of the mind of man: Implications for decision making in the nuclear age. In H.J. Otway (Ed.), *Risk vs. benefit: Solution or dream?* (Report LA 4860-MS). Los Alamos, N.M.: Los Alamos Scientific Laboratory, 1972. (Available from the National Technical Information Service). (28) (b)

Slovic, P. Psychological study of human judgment: Implications for investment decision making. *Journal of Finance*, 1972, 27, 779-799. (31, 33) (c)

Slovic, P. Hypothesis testing in the learning of positive and negative linear functions. *Organizational Behavior and Human Performance*, 1974, 11, 368-376. (23)

Slovic, P., & Fischhoff, B. On the psychology of experimental surprises. *Journal of Experimental Psychology: Human Perception and Performance*, 1977, 3, 544-551. (23,31)

Slovic, P., Fischhoff, B., & Lichtenstein, S. Cognitive processes and societal risk taking. In J.S. Carroll & J.W. Payne (Eds.), *Cognition and social behavior*. Hillsdale, N.J.: Erlbaum, 1976. (6, 19, 23)

Slovic, P., Fischhoff, B., & Lichtenstein, S. Behavioral decision theory. *Annual Review of Psychology*, 1977, 28, 1-39. (8, 19, 34, 35)

Slovic, P., Fischhoff, B., & Lichtenstein, S. Accident probabilities and seat

belt usage: A psychological perspective. *Accident Analysis and Prevention*, 1978, *W*, 281-285. (33)

Slovic, P., Fischhoff, B., & Lichtenstein, S. Rating the risks. *Environment*, 1979, *21*(3), 14-20, 36-39. (33)

Slovic, P., Fischhoff, B. & Lichtenstein, S. Facts vs. fears: Understanding perceived risk. In R. Schwing & W. A. Albers, Jr. (Eds.), *Societal risk assessment: How safe is safe enough?* New York: Plenum, 1980. (33)

Slovic, P., Fischhoff, B., & Lichtenstein, S. Characterizing perceived risk. In R.W. Kates & C. Hohenemser (Eds.), *Technological hazard management*. Cambridge, Mass.: Oelgeschlager, Gunn and Hain, in press. (33) (a)

Slovic, P., Fischhoff, B., & Lichtenstein, S. Response mode, framing, and information processing effects in risk assessment. In R. M. Hogarth (Ed.), *New directions for methodology of social and behavioral science: The framing of questions and the consistency of response*. San Francisco: Jossey-Bass, in press. (33) (b)

Slovic, P., Fischhoff, B., Lichtenstein, S., Corrigan, B., & Combs, B. Preference for insuring against probable small losses: Insurance implications. *Journal of Risk and Insurance*, 1977, *44*, 237-258. (33)

Slovic, P., Kunreuther, H., & White, G. F. Decision processes, rationality, and adjustment to natural hazards. In G. F. White (Ed.), *Natural hazards: Local, national, and global*. New York: Oxford University Press, 1974. (33)

Slovic, P., & Lichtenstein, S. Comparison of Bayesian and regression approaches to the study of information processing in judgment. *Organizational Behavior and Human Performance*, 1971, *6*, 649-744. (1, 3, 4, 23, 28)

Slovic, P., Lichtenstein, S., & Fischhoff, B. Images of disaster: Perception and acceptance of risks from nuclear power. In G. T. Goodman & W. D. Rowe (Eds.), *Energy risk management*. London: Academic Press, 1979. (33)

Slovic, P., & Tversky, A. Who accepts Savage's axiom? *Behavioral Science*, 1974, *19*, 368-373. (34)

Smedslund, J. The concept of correlation in adults. *Scandinavian Journal of Psychology*, 1963, *4*, 165-173. (15, 19)

Smedslund, J. Note on learning, contingency, and clinical experience. *Scandinavian Journal of Psychology*, 1966, *7*, 265-266. (11, 19)

Smith, E.E., Shoben, E.J., & Rips, L.J. Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review*, 1974, *81*, 214-241. (6)

Snapper, K.J., & Fryback, D.G. Inference based on unreliable reports. *Journal of Experimental Psychology*, 1974, *87*, 401-404. (26)

Snyder, M., & Swann, W.B., Jr. When actions reflect attitudes: The politics of impression management. *Journal of Personality and Social Psychology*, 1976, *34*, 1034-1042. (15)

Snyder, M., & Swann, W.B., Jr. Behavioral confirmation in social interaction: From social perception to social reality. *Journal of Experimental Psychology*, 1978, *14*, 148-162. (9) (a)

Snyder, M., & Swann, W.B., Jr. Hypothesis-testing processes in social

- interaction. *Journal of Personality and Social Psychology*, 1978, 36, 1202-1212. (9) (b)
- Snyder, M., Tanke, E. D., & Berscheid, E. Social perception and interpersonal behavior: On the self-fulfilling nature of social stereotypes. *Journal of Personality and Social Psychology*, 1977, 35, 656-666. (9, 15)
- Snyder, R.E. Mammography: Contributions and limitations in the management of cancer of the breast. *Clinical Obstetrics and Gynecology*, 1966, 9, 207-220. (18)
- Sokolov, E.N. The modeling properties of the nervous system. In I. Maltzman and K. Cole (Eds.), *Handbook of Contemporary Soviet Psychology*, New York: Basic Books, 1969. (35)
- Soskin, W. F. Bias in postdiction from projective tests. *Journal of Abnormal and Social Psychology*, 1954, 49, 69-74. (20)
- Sowby, F.D. Radiation and other risks. *Health Physics*, 1965, 11, 879-887. (33)
- Srinivisan, V. A theoretical comparison of the predictive power of the multiple regression and equal weighting procedures (Research Paper No. 347), Stanford, Calif.: Stanford University, Graduate School of Business, 1977. (28)
- Staddon, J.E.R., & Simmelhag, V.L. The "superstition" experiment: A reexamination of its implications for the principles of adaptive behavior. *Psychological Review*, 1971, 78, 3-43. (19)
- Stael von Holstein, C.-A.S. An experiment in probabilistic weather forecasting. *Journal of Applied Meteorology*, 1971, 10, 635-645. (22, 31) (a)
- Stael von Holstein, C.-A.S. Two techniques for assessment of subjective probability distributions-An experimental study. *Acta Psychologica*, 1971, 35, 478-494. (1) (b)
- Stael von Holstein, C.-A. S. Probabilistic forecasting: An experiment related to the stock market. *Organizational Behavior and Human Performance*, 1972, 8, 139-158. (31)
- Starr, B.J., & Katkin, E.S. The clinician as an aberrant actuary: Illusory correlation and the incomplete sentence blank. *Journal of Abnormal Psychology*, 1969, 74, 670-675. (15)
- Starr, C, & Whipple, C. Risks of risk decisions. *Science*, 1980, 208, 1114-1119. (33)
- Steinbrugge, K.V., McClure, F.E., & Snow, A.J. Studies in seismicity and earthquake damage statistics (Report [Appendix A] COM-71-00053). Washington, D.C.: U.S. Department of Commerce, 1969. (33)
- Stevens, S.S., & Greenbaum, H.B. Regression effect in psychophysical judgment. *Perception and Psychophysics*, 1966, 1, 439-446. (4)
- Storms, M.D. Videotape and the attribution process: Reversing actors' and observers' points of view. *Journal of Personality and Social Psychology*, 1973, 27, 165-175, (9, 12)
- Storms, M.D., & Nisbett, R.E. Insomnia and the attribution process. *Journal of Personality and Social Psychology*, 1970, 16, 319-328. (7)
- Strachey, L. *Eminent Victorians*. New York: Putnam, 1918. (23)

Strickland, L.H. Surveillance and trust, *Journal of Personality*, 1958, 26, 200-215. (9)

Strickland, L.H., Lewicki, R.J., & Katz, A.M. Temporal orientation and perceived control as determinants of risk-taking. *Journal of Experimental Social Psychology*, 1966, 2, 143-151. (16)

Svenson, O. Are we all less risky and more skillful than our fellow drivers? *Acta Psychologica*, 1981, 47, 143-148. (33)

Swets, J.A., Tanner, W.P., Jr., & Birdsall, T. Decision processes in perception. *Psychological Review*, 1961, 68, 301-340. (22)

Taft, R. The ability to judge people. *Psychological Bulletin*, 1955, 52, 1-23. (20)

Tawney, R.H. *The agrarian problems in the sixteenth century*. New York: Franklin, 1961. (23)

Taylor, H.C, & Russell, J.T. The relationship of validity coefficients to the practical effectiveness of tests in selection: Discussion and tables. *Journal of Applied Psychology*, 1939, 23, 565-578, (19)

Taylor, S.E., & Crocker, J. The processing of context information in person perception. Unpublished manuscript, Harvard University, 1979. (34) (a)

Taylor, S.E., & Crocker, J. Schematic bases of social information processing. In E.T. Higgins, P. Herman, & M.P. Zanna (Eds.), *The Ontario Symposium on Personality and Social Psychology* (Vol. 1). Hillsdale, N.J.: Erlbaum, 1979. (13) (b)

Taylor, S.E., & Fiske, S.T. Point of view and perceptions of causality. *Journal of Personality and Social Psychology*, 1975, 32, 439-445. (9,12, 13)

Taylor, S.E., & Fiske, S.T. Salience, attention, and attribution: Top of the head phenomena. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 11). New York: Academic Press, 1978. (9)

Taylor, S.E., Fiske, S.T., Close, M., Anderson, D., & Ruderman, A. Solo status as a psychological variable: The power of being distinctive. Unpublished manuscript, Harvard University, 1976. (9, 13)

Taylor, S.E., Fiske, S.T., Etcoff, N., & Ruderman, A. The categorical and contextual bases of person memory and stereotyping. *Journal of Personality and Social Psychology*, 1978, 36, 778-793. (13)

Taynor, J., & Deaux, K. When women are more deserving than men: Equity, attribution, and perceived sex differences. *Journal of Personality and Social Psychology*, 1973, 28, 360-367. (15)

Testa, T.J. Causal relationships and the acquisition of avoidance responses. *Psychological Review*, 1974, 81, 491-505. (15)

Thibaut, J.W., & Riecken, H.W. Some determinants and consequences of the perception of social causality. *Journal of Personality*, 1955, 24, 113-133. (9)

Thornton, B. Linear prediction of marital happiness: A replication. *Personality and Social Psychology Bulletin*, 1977, 3, 674-676. (28) (a)

Thornton, B. Personal communication, 1977. (28) (b)

Tufte, E. R., & Sun, R.A. Are there bellwether electoral districts? *Public Opinion Quarterly*, 1975, 39, 1-18. (23)

Tukey, J. W. Some thoughts on clinical trials, especially problems of multiplicity. *Science*, 1977, 198, 679-690. (23)

Tulving, E., & Pearlstone, Z. Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior*, 1966, 5, 381-391. (11)

Tune, G.S. Response preferences: A review of some relevant literature. *Psychological Bulletin*, 1964, 61, 286-302. (2, 3)

Turoff, M. An alternative approach to cross-impact analysis. *Technological Forecasting and Social Change*, 1972, 3, 309-339. (8)

Tversky, A. Intransitivity of preferences. *Psychological Review*, 1969, 76, 31-48. (19)

Tversky, A. Features of similarity. *Psychological Review*, 1977, 84, 327-352. (6, 8)

Tversky, A., & Kahneman, D. The belief in the "law of small numbers." *Psychological Bulletin*, 1971, 76, 105-110. (1, 3, 4, 6, 7, 23, 34)

Tversky, A., & Kahneman, D. Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 1973, 5, 207-232. (1, 4, 9, 12, 13, 14, 23, 30, 33)

Tversky, A., & Kahneman, D. Judgment under uncertainty: Heuristics and biases. *Science*, 1974, 185, 1124-1131. (6, 8, 13, 19, 22, 27, 28, 30, 31, 33, 34, 35)

Tversky, A., & Kahneman, D. Causal schemas in judgments under uncertainty. In M. Fishbein (Ed.), *Progress in social psychology*. Hillsdale, N.J.: Erlbaum, 1980. (6, 10, 19, 30)

Tversky, A., & Kahneman, D. The framing of decisions and the rationality of choice. *Science*, 1981, 211, 453-458. (33, 34)

U.S. House of Representatives. Browns Ferry nuclear plant fire: Hearing. 94th Congress, 1st Session. Washington, D.C.: U.S. Government Printing Office, September 16, 1975. (33)

U.S. Nuclear Regulatory Commission. Reactor safety study: An assessment of accident risks in U.S. commercial nuclear power plants (WASH 1400 [NUREG-75/014]). Washington, D.C.: NRC, 1975. (22, 23, 33)

U.S. Nuclear Regulatory Commission. Risk assessment review group report to the U.S. Nuclear Regulatory Commission (NUREG/CR-0400). Washington, D.C.: NRC, 1978. (33)

U.S. Weather Bureau. Report on weather bureau forecast performance 1967-8 and comparison with previous years (Tech. Memorandum WBTM FCST, 11), Silver Spring, Md.: Office of Meteorological Operations, Weather Analysis and Prediction Division, March 1969. (22)

Valins, S. Persistent effects of information about internal reactions: Ineffectiveness of debriefing. In H. London & R.E. Nisbett (Eds.), *Thought and feeling: Cognitive modification of feeling states*. Chicago: Aldine, 1974. (9)

Vitz, P.C., & Todd, D.C. A coded element model of the perceptual processing of sequential stimuli. *Psychological Review*, 1969, 76, 433-449. (3)

Von Winterfeldt, D., & Edwards, W. Flat maxima in linear optimization

models (Tech. Rep. 011313-4-T). Ann Arbor: University of Michigan, Engineering Psychology Laboratory, 1973. (22, 28)

Wagenaar, W.A. Subjective randomness and the capacity to generate information. In A.F. Sanders (Ed.), *Attention and Performance, III*, Acta Psychologica, 1970, 33, 233-242. (3)

Wainer, H. Estimating coefficients in linear models: It don't make no nevermind. *Psychological Bulletin*, 1976, 83, 312-317. (28)

Wainer, H., & Thissen, D. Three steps toward robust regression. *Psychometrika*, 1976, 41, 9-34. (28)

Wallace, H.A. What is the corn judge's mind? *Journal of the American Society of Agronomy*, 1923, 15, 300-304. (28)

Wallach, M.A., & Wing, C.W. Is risk a value? *Journal of Personality and Social Psychology*, 1968, 9, 101-106. (16)

Wallis, W.A., & Roberts, H.V. *Statistics: A new approach*. New York: Free Press, 1956. (4)

Wallsten, T. Processes and models to describe choice and inference. In T. Wallsten (Ed.), *Cognitive processes in choice and behavior*. Hillsdale, N.J.: Erlbaum, 1980. (19)

Watlsten, T.S., & Budescu, D.F. Encoding subjective probabilities: A psychological and psychometric review. Research Triangle Park, N.C.: U.S. Environmental Protection Agency, Strategies and Air Standards Division, 1980. (31) Walster, E., Berscheid, E., Abrahams, D., & Aronson, E. Effectiveness of debriefing following deception experiments. *Journal of Personality and Social Psychology*, 1967, 6, 371-380. (9)

Ward, W.C., & Jenkins, H.M. The display of information and the judgment of contingency. *Canadian Journal of Psychology*, 1965, 39, 231-241. (11, 15, 19)

Warren, R.M. Perceptual restoration of missing speech sounds. *Science*, 1970, 167, 393-395. (35)

Wason, P.C. On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, 1960, 12, 129-140. (19)

Wason, P.C. Reasoning. In B. Foss (Ed.), *New horizons in psychology*. Middlesex, England: Penguin, 1966. (34)

Wason, P.C. Regression in reasoning? *British Journal of Psychology*, 1969, 60, 471-480. (34)

Wason, P.C., & Evans, J.St.B.T. Dual processes in reasoning? *Cognition*, 1975, 3, 141-154. (34)

Wason, P.C., & Johnson-Laird, P.N. A conflict between selecting and evaluating information in an inferential task. *British Journal of Psychology*, 1970, 61, 509-515. (34)

Wason, P.C., & Johnson-Laird, P.N. *Psychology of reasoning: Structure and Content*. London: Batsford, 1972. (6, 9)

Wason P.C., & Shapiro, D. Natural and contrived experience in a reasoning problem. *Quarterly Journal of Experimental Psychology*, 1971, 23, 63-71. (34)

Weatherwax, R.K. Virtues and limitations of risk analysis. *Bulletin of the*

Atomic Scientists, 1975, 33, 29-32. (22)

Weiner, B. Achievement motivation and attribution theory. Morristown, N.J.: General Learning Press, 1974. (9, 15)

Weiner, B., Frieze, I., Kukla, A., Reed, L., Rest, S., & Rosenbaum, R. Perceiving the causes of success and failure. Morristown, N.J.: General Learning Press, 1971. (7)

Weinstein, N.D. Unrealistic optimism about future life events. *Journal of Personality and Social Psychology*, 1980, 39, 806-820. (33)

Wells, G.L., & Harvey, J.H. Do people use consensus information in making causal attributions? *Journal of Personality and Social Psychology*, 1977, 35, 279-293. (10)

Wells, G.L., & Harvey, J.H. Naive attributors' attributions and predictions: What is informative and when is an effect an effect? *Journal of Personality and Social Psychology*, 1978, 36, 483-490. (10)

Wessman, A.E., & Ricks, D.F. Mood and personality. New York: Holt, Rinehart & Winston, 1966. (7)

Wheeler, G., & Beach, L.R. Subjective sampling distributions and conservatism. *Organizational Behavior and Human Performance*, 1968, 3, 36-46. (3)

Wheeler, W.M. An analysis of Rorschach indices of male homosexuality. *Rorschach Research Exchange*, 1949, 13, 97-126. (17)

White, R.W. Lives in progress: A study of the natural growth of personality. New York: Dryden Press, 1952. (20)

White, R.W. Motivation reconsidered: The concept of competence. *Psychological Review*, 1959, 66, 297-333. (12, 16)

Widdund, R.A. Objective self-awareness. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 9). New York: Academic Press, 1975. (9)

Wiggins, N., & Kohen, E.S. Man vs. model of man revisited: The forecasting of graduate school success. *Journal of Personality and Social Psychology*, 1971, 19, 100-106. (28)

Wilks, S.S. Weighting systems for linear functions of correlated variables when there is no dependent variable. *Psychometrika*, 1938, 8, 23-40. (23, 28)

Williams, P. The use of confidence factors in forecasting. *Bulletin of the American Meteorological Society*, 1951, 32, 279-281. (22)

Wilson, R. Analyzing the daily risks of life. *Technology Review*, 1979, 81, 40-46. (33)

Winch, R.F., & More, D.M. Does TAT add information to interviews? Statistical analysis of the increment. *Journal of Clinical Psychology*, 1956, 12, 316-321. (20)

Winkler, R.L. The assessment of prior distributions in Bayesian analysis. *Journal of the American Statistical Association*, 1967, 62, 776-800. (31)

Winkler, R.L., & Murphy, A.H. Evaluation of subjective precipitation probability forecasts. In *Proceedings of the First National Conference on*

Statistical Meteorology, Hartford, Conn., May 27-29. Boston: American Meteorological Society, 1968. (22) (a)

Winkler, R.L., & Murphy, A.H. "Good" probability assessors. *Journal of Applied Meteorology*, 1968, 7, 751-758. (22) (b)

Wohlstetter, R. Pearl Harbor: Warning and decision. Stanford, Calif.: Stanford University Press, 1962. (23, 31)

Wolfe, J.N. Mammography: Report on its use in women with breasts abnormal and normal on physical examination. *Radiology*, 1964, 83, 244-254. (18)

Wolfe, J.N. Mammography: Errors in diagnosis. *Radiology*, 1966, 87, 214-219.

Wolfe, J.N. Mammography (2nd ed.). Springfield, 111.: Charles C Thomas, 1967. (18)

Wolosin, R.J., Sherman, S.J., & Till, A. Effects of cooperation and competition on responsibility attribution after success and failure, *Journal of Experimental Social Psychology*, 1973, 9, 220-235. (9)

Wood, G. The knew-it-all-along effect. *Journal of Experimental Psychology: Human Perception and Performance*, 1978, 4, 345-353. (31)

Woodworth, R.S. *Experimental psychology*. New York: Holt, 1938. (4)

Wright, G.N., & Phillips, L.D. Personality and probabilistic thinking: An experimental study (Tech. Rep. 76-3). Uxbridge, England: Brunel Institute of Organisational and Social Studies, 1976. (22)

Wright, G.N., Phillips, L.D., Whalley, P.C., Choo, G.T.G., Ng, K.O., Tan, I., & Wisudha, A. Cultural differences in probabilistic thinking. *Journal of Cross-cultural Psychology*, 1978, 9, 285-299. (22)

Wright, G.N., & Wisudha, A. Differences in calibration for past and future events. Paper presented at the 7th Research Conference on Subjective Probability, Utility and Decision-Making, Goteborg, Sweden, August 1979. (22)

Yarrow, M., Campbell, J.D., & Burton, R.V. Recollections of childhood: A study of the retrospective method. *Monographs of the Society for Research in Child Development*, 1970, 35(5). (23)

Yntema, D.B., & Torgerson, W.S. Man-computer cooperation in decisions requiring common sense. *IRE Transactions of the Professional Group on Human Factors in Electronics*, 1961, 2(1), 20-26. (28)

Zadeh, L.A. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Set and Systems I*, 1978, 3-28. (6)

*Даниель Канеман
Пауль Словик
Амос Тверски*

**Принятие решений
в неопределенности
Правила и предубеждения**

**Judgment under Uncertainty:
Heuristics and biases**

Научный редактор: *Г. В. Суходольский*
Редактор: *С. Г. Тарасов*
Литературный редактор: *В. В. Артеменко*
Компьютерная верстка: *Д. В. Козлов*

Свидетельство ДК №2009 от 10.11.2004 г.

Издательство Институт прикладной
психологии “Гуманитарный Центр”.
61166, Украина, Харьков,
отдел продаж (057) 719-52-40, e-mail: huce@kharkov.ua
Подписано в печать 23. 12. 2004.
Формат 70x100/16. Бумага оф.
Усл. печ. л. 39,5 Тираж 1500 экз. Заказ №40
Отпечатано в типографии ГП ХМЗ ФЭД. Украина,
Харьков, ул. Сумская, 132 тел. (057) 719-67-82