

ОБЩЕСТВЕННЫЕ НАУКИ  
ЗА РУБЕЖОМ

ФИЛОСОФИЯ И СОЦИОЛОГИЯ

Д. КЭМПБЕЛЛ  
МОДЕЛИ  
ЭКСПЕРИМЕНТОВ  
В  
СОЦИАЛЬНОЙ  
ПСИХОЛОГИИ  
И  
ПРИКЛАДНЫХ  
ИССЛЕДОВАНИЯХ

**ОБЩЕСТВЕННЫЕ НАУКИ  
ЗА РУБЕЖОМ**

---

**ФИЛОСОФИЯ  
И ПСИХОЛОГИЯ**



DONALD T. CAMPBELL  
MODELS OF EXPERIMENTS  
IN SOCIAL PSYCHOLOGY  
AND  
APPLIED RESEARCHES

OCR А.И. Степов

**Д. КЭМПБЕЛЛ**

**МОДЕЛИ**  
**ЭКСПЕРИМЕНТОВ**  
**В**  
**СОЦИАЛЬНОЙ**  
**ПСИХОЛОГИИ**

**И**

**ПРИКЛАДНЫХ ИССЛЕДОВАНИЯХ**

Перевод с английского

Составление и общая редакция

М. И. БОБНЕВОЙ

Вступительная статья профессора

Г. М. АНДРЕЕВОЙ

МОСКВА

„ПРОГРЕСС“

Редактор Э. М. ПЧЕЛКИНА

*Редакция литературы по философии  
и педагогике*

© «Прогресс», 1980

10508-343

**Методологические проблемы  
экспериментального метода  
в социальных науках**

(Вступительная статья)

Среди многообразных проблем методологии научного познания, пожалуй, нет более актуальной и дискуссионной, чем проблема использования экспериментального метода в социальных науках. Основным вопросом, который поднимается в ходе дискуссий, является вопрос о том, могут ли социальные — а также в ряде случаев и гуманитарные — науки достичь того статуса «истинной» науки, основным методом которой является эксперимент, и, соответственно, могут ли эти науки получать знание, добытое экспериментальным путем.

Какие бы более частные проблемы при этом ни возникали, суть дела сводится к сравнению, сопоставлению двух типов научного знания — экспериментального и неэкспериментального. Чаще всего сопоставление это завершается бесспорным признанием приоритета первого и — в результате — своеобразным «обличением» наук, которые либо по определению не располагают такого рода знанием, либо находятся на первых этапах освоения экспериментального метода. Успехи естественнонаучных дисциплин прибавляют существенные аргументы в пользу такого приговора. Эти установки проявляются и в противопоставлении «сциентистского» и «гуманистического» подходов в науке.

Сложившаяся в XX в. самостоятельная область науки — логика и методология научного исследования — хотя по замыслу и не должна была выступать в качестве мачехи к какой-либо одной из ветвей научного знания, в значительной мере способствовала тому, что социальным наукам отводилась роль Золушки.

Философская традиция неопозитивизма, в недрах которой первоначально сложились соответствующие нормативы, сделала эталоном науки физику и математику, и дальнейшая дискуссия свелась к доказательству несовершенства всякого рода знания, не отвечающего предложенным канонам.

Вместе с тем общая ситуация, сложившаяся в науке XX в., в значительной мере доказывала преимущества тех знаний, которые получены экспериментальным путем. Специалисты в области социальных наук, осознавшие потребность в экспериментальном методе, с самого начала столкнулись с необходимостью не только утвердить свое право на этот метод, но и доказать, что образцы его применения в социальных исследованиях не уступают таковым в естественных науках. Такой постановкой вопроса представители социальных наук сами наносили себе удар: вместо признания права на существование своеобразия социального познания они стали оправдывать его несходство со знанием, получаемым «каноническими» дисциплинами. Потребовались многие годы исследований и дискуссий, прежде чем среди представителей социальных наук начали раздаваться трезвые

голоса относительно применения экспериментального метода с учетом глубокой специфики самого предмета исследования. В разработке этих проблем видное место занимают исследования известного американского ученого Дональда Т. Кэмпбелла.

Кэмпбелла по праву можно причислить к лучшим представителям прогрессивной традиции не только в области психологической науки, но также социологии и антропологии. Как психолог он сформировался в 40-х годах под влиянием как американской, так и европейской психологической мысли. В университете Беркли, который он закончил, он был учеником таких крупнейших психологов, как Э. Толмен и К. Левин. Около двадцати лет Кэмпбелл был профессором одного из старейших американских университетов — Северо-Западного (близ Чикаго) — и одновременно читал курсы лекций в таких ведущих университетах, как Гарвард, Йель, Оксфорд и т. д. Он проводил экспериментальные исследования по восприятию и обучению. По его инициативе и при его личном участии осуществлялся цикл сравнительных психологических и культурноантропологических исследований как в США, так и в ряде других стран, в частности в некоторых странах Африки. Кэмпбелл выступил инициатором и соавтором капитального научного труда по проблемам этноцентризма, в котором с прогрессивных позиций проанализированы многочисленные теории этноцентризма, этнических особенностей и межгрупповых конфликтов. Имя Кэмпбелла известно также среди специалистов по методологии науки и эпистемологии. В последние годы он проводил исследования по проблемам альтруизма и генетической психологии, участвовал в создании нового междисциплинарного направления — оценивания программ и т. д.

## 6

Широта взглядов, глубокая образованность и большой круг проблем, при разработке которых Кэмпбелл выступает как профессионал-исследователь, не только показывают его компетентность в ряде областей социальной науки, но и придают особый вес его методологическим исследованиям. Работы Д. Т. Кэмпбелла широко известны не только в Соединенных Штатах, но и во многих других странах. Советский читатель знаком с некоторыми его статьями, опубликованными в переводе<sup>1</sup>, а специалисты имели возможность слышать ряд его устных выступлений во время пребывания Кэмпбелла в СССР в 1976 г., а также на заседаниях Европейской ассоциации экспериментальной социальной психологии.

Д. Т. Кэмпбелл был удостоен ряда научных премий, многие годы был руководителем секции социальной психологии и психологии личности, избирался президентом Американской психологической ассоциации. В настоящее время он является почетным профессором по курсу Альберта Швейцера в Сиракузском университете штата Нью-Йорк.

В последнее десятилетие не выходило фактически ни одной зарубежной работы в области методологии, методов и методик социальных исследований, в которых не упоминались бы его работы, не использовались бы

разработанные им категории, приемы, принципы исследования.

В области методологических проблем научного познания работы Кэмпбелла тем

значительнее, что сам он является крупным специалистом в области применения

экспериментального метода (и вообще количественных методов) в социальных

---

<sup>1</sup> К э м п б е л л Д. Т. Слепые вариации и селективный отбор как главная стратегия процессов познания. — В кн.: Самоорганизующиеся системы. М., «Мир», 1964; К э м п б е л л Д. Т. Социальные Диспозиции индивида и их групповая функциональность: эволюционный аспект. — В кн.: Психологические механизмы регуляции социального поведения М., «Наука», 1979.

## 7

разработанные им категории, приемы, принципы исследования.

В области методологических проблем научного познания работы Кэмпбелла тем значительнее, что сам он является крупным специалистом в области применения экспериментального метода (и вообще количественных методов) в социальных

исследованиях и в то же время человеком, склонным к постоянному размышлению по поводу природы, особенностей и проблем этих методов. Его позиция не может быть уподоблена позиции тех исследователей, которые «измеряют, не размышляя», или тех, которые «размышляют, не измеряя». Д. Кэмпбелл выступает и как высококвалифицированный профессионал, и как человек с чрезвычайно широким, подлинно философским взглядом на природу науки, ее действительные возможности, затруднения и просчеты.

В настоящее время проведение большого объема как фундаментальных, так и особенно прикладных психологических и социально-психологических исследований с особой остротой ставит вопросы методологии и методики исследований, проблемы планирования экспериментов, разработки моделей экспериментальных и прикладных исследований, валидности полученных данных, достоверности и надежности научных выводов и рекомендаций. Эти специальные вопросы методологии научного исследования и разрабатываются Д. Т. Кэмпбеллом.

Предлагаемая советскому читателю книга включает ряд важнейших работ Д. Т. Кэмпбелла в области методологии и методики научного исследования. Для этого тома по рекомендации самого автора были отобраны работы, цель которых состоит, прежде всего, в том, чтобы ввести читателя в круг проблем и понятий, сообщить тот минимум знаний о предмете, который необходим для дальнейшего, более глубокого ознакомления с ним и для самостоятельного размышления и исследования. (Более специальные работы автора, относящиеся к этой же проблематике, отмечены в примечаниях редактора-составителя.) Хотя книга является сборником, при отборе работ и автор, и составитель тома стремились к тому, чтобы эти работы в совокупности составили целостный труд, отражающий в то же

время различные проблемы, рассматриваемые автором на протяжении многих лет и находящиеся в центре его научных интересов. Этот характер всей работы задан вводной статьей Кэмпбелла, подготовленной специально для данного издания.

Содержание настоящего тома можно разделить на три части: 1) общие методологические проблемы соотношения качественного и количественного знания в социальных науках; 2) описание специфики основных моделей и планов экспериментальных и квазиэкспериментальных исследований; 3) некоторые прикладные проблемы методологии социальных наук, прежде всего проблемы оценки эффективности предлагаемых практике программ. Эта последовательность отражает общую логику рассуждений автора, настойчиво стремящегося к достижению надежности результатов, получаемых в социальных исследованиях.

\* \* \*

Первый блок проблем, обсуждаемых Кэмпбеллом, связан с основным содержанием полемики о специфической природе социального познания, которая продолжается вот уже почти целое столетие. Дискуссии середины XX в. внесли много нового в эту полемику. Самым существенным, пожалуй, является то, что интерес к проблемам качества научного знания перестал быть традиционным интересом лишь философии науки, но породил особую «внутринаучную методологическую рефлексию» самих исследователей в конкретных областях знания. Этим объясняется и то многообразие наименований, которые даются названной области исследования: это уже давно не только «философия науки», но и «социология науки», «социальная психология науки» и, наконец (самое нейтральное наименование), «науковедение», к которому можно отнести весь спектр проблем, ранее возникавших в различных областях методологической рефлексии.

Со времени появления книги Т. Куна «Структура научных революций» статус той области научного знания, предметом которой является само научное знание, стал особенно прочным. Нет сомнений в том, что объективной основой повышения интереса к этим проблемам

явилось возрастание роли и изменение социальных функций науки в обществе, особенно в связи с научно-технической революцией. С другой стороны, лавинообразная масса научных исследований, ежегодно обрушивающаяся на человечество, требует на каких-то этапах остановиться, оглянуться, отдать себе отчет в том, что же может и чего не может наука сегодня.

Поэтому не только неудовлетворенность предписаниями традиционной «философии науки» побуждает исследователей самих взяться за анализ своих собственных возможностей, но и практическая потребность: все чаще и чаще общество предъявляет своеобразный счет к научным исследованиям, особенно когда они в массовом масштабе начинают проводиться на прикладном уровне. С особой остротой проблемы *качества* знания встают перед общественными науками. Не только социальные, но и гуманитарные науки обращаются сейчас к старому вопросу — что же обеспечивает валидность научного вывода: строгое применение количественных методов или что-то иное, не охватываемое традиционными понятиями надежности измерительных процедур.

Д. Кэмпбелл, рассматривая эту проблему, называет свой подход «дескриптивной эпистемологией», желая подчеркнуть, что предлагаемый им анализ не есть анализ с позиций «философии науки», но есть философское осмысление реальной исследовательской практики, предпринятое самим исследователем.

Точка зрения Кэмпбелла особенно интересна потому, что сам он относит себя к сторонникам точных методов в социальных науках. Он признает, что распространенная формулировка альтернативы возможных стратегий исследования предполагает преимущество количественной ориентации. Это проявляется в определенной оценочной нагрузке, которую несут на себе условно обозначаемые стороны альтернативы: именно количественный метод отождествляется с «научным», «сциентистским», «естественнонаучным», в то время как метод качественного анализа определяется как «гуманистический», «гуманитарный», «философский», «клинический», «оценивающий» или даже «обыденное знание».

Те, кто пытается защитить второй вид стратегии научного исследования «изнутри», то есть работая в рамках качественного подхода, обычно проигрывают, поскольку

его представители — вольно или невольно — занимают оборонительную позицию. Иное дело, когда эта альтернатива рассматривается человеком, занимающим прочное положение в области применения количественных методов и являющимся здесь бесспорным авторитетом.

Кэмпбелл выступает именно с такой позиции. Его защита качественного подхода в социальных науках строится по двум линиям. Во-первых, путем признания здравого смысла как важного компонента качественных методов. Даже такая каноническая наука, как физика, по мнению Кэмпбелла, не гнушается перепроверкой многих добытых ею фактов в свете соображений здравого смысла. Поэтому нет оснований третировать его и в социальных науках: хотя никакое научное знание не есть просто знание, основанное на здравом смысле, оно вместе с тем не может развиваться и вне зависимости от здравого смысла. Несправедливо, когда апелляция к здравому смыслу в естественных науках допускается, в то время как в социальных науках немедленно вызывает критику. Вместе с тем здесь перепроверка данных научного исследования с позиций здравого смысла приобретает особое значение, прежде всего в связи со спецификой предмета исследования. Если рассматривать социальную науку не только как лабораторную науку (а такой ее статус весьма сомнителен, по словам Кэмпбелла), то станет очевидным, что она, по

определению, имеет дело с реальными человеческими проблемами, с теми процессами, в которые включены реальные люди в реальных обстоятельствах. Именно это умножает не только возможность, но и необходимость качественного знания в этих науках, в котором велик «удельный вес» здравого смысла.

Кэмпбелл описывает различные формы качественного знания в социальных науках: это и примитивное качественное знание, источником которого являются сами участники различного рода исследований; и мнения экспертов, к которым сплошь и рядом приходится прибегать; даже системный подход, — в принципе неотрицаемый и с точки зрения строгого «сциентизма», — который выступает своеобразной альтернативой количественному методу. Все представители общественных наук — и социологи, и психологи, и историки, и педагоги, и политологи — так или иначе сталкиваются

## 11

с теми «вкраплениями» здравого смысла, которые неизбежно присутствуют в этих различных формах качественного знания. Хотя истина, излагаемая в этой первой линии защиты, может показаться достаточно банальной, в действительности нужно большое мужество, чтобы во весь голос в век компьютеров и изоощренных математических выкладок, завоевавших прочное место в науке, сделать заявление о допустимости здравого смысла как важнейшего компонента научного метода.

Вторая линия защиты качественного метода в социальных науках развертывается автором как выявление ограниченности количественного метода, как необходимость его дополнения. Своеобразным резюме внутренней недостаточности количественного знания служит рассуждение автора о том, что когда измерительные приборы в физических лабораториях грешат неточностями, то неисправность обнаруживается лишь благодаря качественному знанию и предположениям по поводу того, что должен был бы показывать этот прибор! Конечно, дело не ограничивается лишь приведением этого афоризма: автор остается верен своему стилю изложения, где соседствует образность и самый изоощренный педантизм. На примере количественного подхода, в данном случае отождествляемого с экспериментальным методом, выявляются и внутренние проблемы такой стратегии исследования. Самые главные из них связаны с неизбежным возникновением артефактов в экспериментальных исследованиях, Кэмпбелл исследует внимательнейшим образом все основные типы артефактов, неоднократно обсуждаемые, в частности, в социально-психологической литературе (эффект Розенталя, фоновые взаимодействия, «внедрение» фактора времени в проблему репрезентативности выборки и т. д.). Разумеется, современная наука много сделала для разработки средств контролирования артефактов, однако в силе остается чрезвычайно глубокий вывод, к которому приходит Кэмпбелл: в любой проблеме *измерения* в социальных науках так или иначе сохраняется проблема *смещения*, каким бы парадоксальным ни казалось такое положение.

Аргументация в пользу такого утверждения сводится к тому, что эта проблема пристрастности могла бы отсутствовать в измерении при одном единственном

## 12

условии, а именно при абсолютно строгом проведении принципа операционализма. Однако точное знание ситуации как в философии неопозитивизма вообще, так и в логике и методологии научного исследования позволяет Кэмпбеллу поставить совершенно точный диагноз: строгий операционализм давно преодолен, а раз так, то проблема смещения измерения сохраняется. И в этом случае, несмотря на кажущуюся тривиальность вывода, он в действительности очень глубок и по плечу лишь подлинному

ученому. Как часто, к сожалению, можно встретить в специальной литературе, особенно в социологической или психологической, надменные и легковесные утверждения о том, что только на пути абсолютной операционализации может быть достигнут стандарт подлинно научного исследования. Профессионализм таких утверждений лишь кажущийся, на деле за его фасадом — достаточно приблизительное понимание того, сколь сложным и превратным было осмысление принципа операционализма на различных исторических рубежах развития философии науки. Как часто складывается карикатурная ситуация для тех, кто знает лишь понаслышке о сложнейших проблемах философии научного познания: в самой философии давно отказались от чрезмерно жестких акцентов в разработке того или другого принципа, а в конкретных науках — по инерции — все еще продолжают настаивать на том, что является вчерашним днем философской мысли... Тем веселее и авторитетнее слова серьезного ученого, в равной мере искушенного и в специальных, и в метанаучных проблемах. Кэмпбелл не боится показаться банальным и заявляет совершенно четко: в социальных науках проблема артефактов в эксперименте настолько сложна, что, увы, приходится считаться с суровой критикой по адресу этого метода, признать его существенные ограничения в этой сфере знания.

Возникает вопрос, какой же выход можно предложить из проанализированной таким образом ситуации? Кэмпбелла не убеждают доводы тех критиков, которые приходят к доказательству приоритета «гуманистического», качественного знания в социальных науках и на этой основе вообще готовы изгнать количественные методы. Для него выход заключается в другом: дело не в том, какой метод — количествен-

ный или качественный — лучше, нужно органическое сочетание их обоих. Специфика социальных наук требует того, чтобы исследователи овладели сложным искусством сочетания математической точности вывода и адекватного представления о существенных характеристиках изучаемого объекта. Кэмпбелл не апеллирует при такой постановке вопроса к принципам марксистской методологии научного познания, но, по существу, его позиция близка исследователю, сознательно принимающему эти принципы. Именно поэтому для советского читателя этот раздел работы Кэмпбелла представляет несомненный интерес.

Нельзя сказать, что знание общего методологического принципа, утверждающего необходимость сочетания качественного и количественного подходов, автоматически означает решение всех конкретных исследовательских проблем. Дискуссии о способах применения количественных методов в социальных науках до сих пор продолжаются и среди советских ученых. Можно, пожалуй, констатировать, что наиболее примитивная форма этих дискуссий, когда в открытую сталкивались наиболее радикальные позиции как «сциентистов», так и «гуманистов», осталась позади. По-видимому, сегодня наиболее продуктивным является поиск конкретных путей построения специальной «метрологии» для гуманитарных и социальных наук, которая смогла бы предложить особую форму измерительной процедуры, с самого начала учитывающую специфику характеристик объектов исследования. Создание такой новой «метрологии» — дело будущего. Сейчас же оправданным представляется тот подход, который предлагается Кэмпбеллом: тщательный анализ существующих методик, выявление в них таких факторов, которые осложняют их применение в системе социального или гуманитарного знания. При этом утверждается вновь простая на первый взгляд мысль: если нельзя провести совершенный эксперимент, то нужно хотя бы минимизировать его несовершенства. Кажущаяся простота такой рекомендации не должна скрыть те многочисленные трудности, которые в действительности кроются за ней. Именно здесь необходим высокий профессионализм и огромный опыт в осуществлении экспериментов, а вместе с тем

широта научного кругозора. Эти качества в полной мере проявляются в настоящей книге автора.

Второй круг проблем составляет содержание самой объемной работы, включенной в настоящий том, — «Эксперименты и квазиэксперименты». Пожалуй, это наиболее «трудная» часть книги Кэмпбелла, и вряд ли она рассчитана на легкое чтение. Если при обсуждении общих проблем методологии социального познания автор придерживается стиля, свойственного широкой философской, «гуманитарной» традиции в науковедении, здесь он скорее «сциентист», но «сциентист», постоянно размышляющий о том, что можно, нужно и при каких обстоятельствах применить из этой традиции в социальных науках. Главная идея автора — выявить те специфические факторы, которые угрожают валидности вывода по данным эксперимента, проводимого в экономике, социологии, педагогике или психологии. Учет этих факторов поможет, по мнению Кэмпбелла, избежать как неоправданных иллюзий, так и чрезмерных разочарований, которые сулит экспериментальный метод в этих науках.

Приступая к решению поставленной задачи, Кэмпбелл, прежде всего, уточняет понятия «внутренней» и «внешней» валидности эксперимента. Нельзя сказать, что это какой-то совершенно новый подход к оценке возможностей экспериментального метода: современная социологическая и социально-психологическая литература уже давно содержит поиски способов, повышающих валидность вывода в экспериментальном исследовании. Существует целый ряд попыток, предпринятых на этом пути со стороны таких исследователей, как М. Орн, Р. Розенталь, Дж. Сильвермен<sup>1</sup>.

<sup>1</sup> Подробный анализ этих попыток предпринят в советской социально-психологической литературе. См.: Андреева Г. М. Методологические проблемы развития американской социальной психологии. — «Вопросы психологии», 1977, № 2; Жуков Ю. М., Гржегоржевская И. А. Эксперимент в социальной психологии. Проблемы и перспективы. — В кн.: Методология и методы социальной психологии. М., 1977; Панферов В. Н., Трусов В. П. Эксперимент. — В кн.: Методы социальной психологии, изд. ЛГУ, 1977.

Полезно напомнить, по каким направлениям ведется критика экспериментального метода, чтобы отчетливее был виден вклад автора «Экспериментов и квазиэкспериментов». Одна группа критиков концентрирует внимание, прежде всего на ограниченных возможностях *лабораторного* эксперимента в социальных науках. В этом смысле лабораторный эксперимент рассматривается как определенная исследовательская стратегия отличная от другой стратегии, связанной преимущественно с методом наблюдения. По мнению таких следователей, как В. Мак-Гайр, М. Аргайл, А. Тэшфел главная слабость лабораторного эксперимента в том, что стремление выделить минимальное количество переменных с целью их всестороннего контроля приводит в ряде случаев к игнорированию такой характеристики, которая является наиболее существенной для понимания природы того или иного социального явления. Утрата специфики именно *социальных* связей — наиболее часто встречающаяся «погрешность» лабораторных экспериментальных ситуаций. Мак-Гайр полагает, что стремление к ограничению числа переменных в эксперименте настолько сильно у исследователей, что даже переход к другим формам эксперимента, а именно к полевым экспериментам, не помогает делу. Полевой эксперимент, по крайней мере, в многочисленных социально-психологических исследованиях, лишь номинально является «естественным экспериментом», поскольку в действительности и в этом случае сохраняется та же самая исследовательская стратегия, общая схема действия, которая отработана для лабораторного эксперимента: исследователь просто «присутствует» в

реальной ситуации, саму же эту ситуацию он заботливо препарировывает под проверяемую гипотезу, то есть опять-таки ограничивает число переменных, оставляя лишь нужный ему «по гипотезе» минимум<sup>1</sup>.

К этой же группе аргументов следует отнести и те, которые связаны с постановкой проблемы репрезентативности экспериментальных исследований. Речь идет в данном случае не о том, как при проведении экспериментальных исследований осуществляются расчеты

---

<sup>1</sup> См. подробно: Андреева Г. М., Богомолова Н. Н., Петровская Л. А. Современная социальная психология на Западе. М., 1978, гл. VI.

норм выборки и т. д., но о весьма специфическом смещении выборки, которое возникает из-за способа отбора испытуемых. По мнению Дж. Мак-Дэвида и Г. Хэртри, люди приходят на лабораторный эксперимент в качестве испытуемых добровольно, и уже этот факт порождает смещение выборки. Оно становится тем более значимым, если учесть практические трудности приглашения испытуемых на эксперименты в лабораторию: как правило, сюда приходят те люди, которые оказались «под рукой», и естественно, что это, прежде всего, студенты, причем студенты психологических колледжей. Недаром, некоторые авторы именуют американскую социальную психологию наукой о «второкурсниках», имея в виду основной контингент испытуемых.

Другой ряд возражений касается возможностей эксперимента в плане выявления действительной позиции испытуемых, а также возможностей распространения полученного вывода на более широкую сферу; если речь идет о лабораторном эксперименте, то это вопрос о возможности распространить его выводы на «реальную жизнь». Именно в этой области особенно тщательно исследованы некоторые факторы, которые могут помешать корректному выводу. По мысли М. Орна эксперимент в социальном исследовании не удовлетворяет требованиям так называемой «экологической валидности», то есть возможности переноса выводов за пределы экспериментальной ситуации. Причина этого заключается в том, что между испытуемым и экспериментатором устанавливается особого рода взаимодействие, которое предполагает включение в эксперимент особого класса переменных (характеристики личности экспериментатора, лабораторной обстановки, специфические черты коммуникативного процесса, который осуществляется между испытуемым и экспериментатором, и т. д.). М. Розенберг вводит особый термин «предвосхищающая оценка», который обозначает позицию испытуемого, его стремление «подыграть» экспериментатору, угадать то поведение, которого ждет экспериментатор, и продемонстрировать именно его, а не то, которое было бы в данной ситуации «подлинным» поведением испытуемого. Дж. Сильвермен полагает, что по этим причинам эксперименты — по крайней мере, в социальной психологии — не выявляют мотивацию людей

в реальных жизненных ситуациях, а потому сам метод оказывается нерелевантным реальным социальным проблемам.

Приведенные иллюстрации показывают, что в настоящее время самые разнообразные позиции критики экспериментов в социальных науках стали довольно распространенным явлением. Это определяет актуальность работы Кэмпбелла по анализу и систематизации всех тех действительных «опасностей», которые подстерегают исследователя, коль скоро он вступает на трудный путь экспериментирования в социальных науках. Его задача заключается в том, чтобы описать все встречающиеся в реальной практике

экспериментирования факторы, которые угрожают валидности вывода. Для этого необходимо, прежде всего, учесть те различные понятия валидности, которые употребляют критики экспериментального метода в социальных науках. Кэмпбелл указывает четыре возможных вида валидности: внутренняя и внешняя валидность, валидность статистического вывода и конструктивная валидность. Однако в данной работе содержится материал подробного исследования лишь двух видов валидности, о которых речь идет наиболее часто. Это — внутренняя валидность, то есть такое качество эксперимента, которое означает, что именно экспериментальное воздействие явилось причиной изменений, зарегистрированных в экспериментальной ситуации, и внешняя валидность, то есть такое качество эксперимента, которое гарантирует возможность обобщения вывода, возможность его распространения на совершенно определенные внеэкспериментальные ситуации. Для каждого исследователя крайне важно научиться выбирать такие оптимальные модели экспериментального исследования, где были бы обеспечены оба типа валидности.

Для обучения исследователя решению такой задачи Кэмпбелл выявляет и описывает двенадцать факторов, «угрожающих» внутренней (восемь факторов) и внешней (четыре фактора) валидности эксперимента, а затем раскрывает действие каждого из этих факторов в различных планах экспериментального исследования. Нет необходимости описывать здесь полученные комбинации: их изложение составляет основное содержание работы автора. Необходимо сказать только, что слож-

ный материал преподнесен здесь в максимально дидактичной форме, чему способствуют составленные Кэмпбеллом таблицы, помогающие следить за ходом изложения и постоянно сверяться с тем, как описанные двенадцать факторов, «угрожающих» валидности, накладываются на шестнадцать различных планов эксперимента. Эта часть книги Кэмпбелла носит характер своего рода справочника, необходимого для каждого, кто приступает к экспериментальному исследованию в социальных науках. Нельзя, конечно, упрощенно представлять себе решение проблемы: стоит «выучить» предложенные автором шестнадцать схем и проявление в них двенадцати факторов, как заветная оптимальная модель эксперимента будет найдена. Весь ход рассуждений Кэмпбелла опровергает возможность такого легкого вывода. Напротив, вся сложность и тонкость работы экспериментатора в социальных науках состоит в том, что только глубокое понимание специфики материала должно руководить исследователем при выборе наиболее адекватной модели исследования. Поэтому нельзя научиться проводить эксперименты, прочитав книгу Кэмпбелла, но, умея проводить эксперименты, можно сделать их более совершенными, «сверив» свои планы с планами, проанализированными автором.

Среди шестнадцати возможных планов эксперимента Кэмпбелл выделяет три группы: доэкспериментальные (планы 1—3), экспериментальные (планы 4—6) и квазиэкспериментальные (планы 7—16).

Что касается первых двух групп, то планы экспериментов классифицированы здесь на основании двух показателей, обычно фигурирующих при изложении теории экспериментального метода: наличие или отсутствие контрольной и экспериментальной группы, наличие или отсутствие тестирования до и после экспериментального воздействия. Каждый план отличается от другого тем, насколько полно представлены в нем комбинации предварительных и последующих тестирований в контрольной и экспериментальной группах. В принципе такая классификация уже довольно давно предложена в социологии (например, известная схема Стауффера). По сравнению с традиционно выделяемыми на основании сочетания предложенных показателей комбинациями Кэмпбелл лишь вводит различие доэкс-

периментальных и «истинно» экспериментальных планов. Но и это, казалось бы, незначительное дополнение представляет большой интерес с точки зрения общего подхода к экспериментированию в социальных науках. В частности, особое значение имеет трактовка в рамках логики экспериментального метода такого далекого от традиционных форм эксперимента типа исследования, как «исследование единичного случая» (case-study). В многочисленных руководствах, например по социологии, хотя и признается ценность такого рода исследований, их редко отваживаются считать экспериментальными. Кэмпбелл и в данном случае смело раздвигает рамки допустимых в социальных науках экспериментальных моделей.

Однако наибольший интерес представляет описание *квазиэкспериментальных* планов как совершенно специфического вида экспериментирования в обществоведении. Сам автор считает эту часть своей работы наиболее оригинальной. Здесь отчетливо выявляется основная *линия* его рассуждений о специфике познавательных средств в социальных и гуманитарных науках. По мнению Кэмпбелла, трудности, которые осознал исследователь, применяющий более традиционные планы эксперимента, не должны привести его к выводу, что лучше вообще отказаться от столь сложного метода и перейти к еще более неформальным методам исследования. Список источников невалидности в этих планах не должен отпугнуть исследователя. Напротив, следует искать возможность еще менее строгих планов эксперимента, осознать их несовершенства и *тем не менее* пойти на их применение, «если применение лучшего плана невозможно». Автор решает здесь две важные задачи: познавательную — обосновывает *допустимость* нетрадиционных планов эксперимента в социальных науках — и просветительскую — обучает тому, как *минимизировать* несовершенство этих планов при постоянном его осознании.

Квазиэкспериментом автор называет такой исследовательский прием, когда осуществляется нечто вроде планирования эксперимента при сборе данных (то есть в отношении того, *когда* и *на ком* производить тестирование), но нет полного контроля порядка экспериментальных воздействий (то есть *когда* и *кому* предъявля-

ются воздействия, рандомизация предъявлений). Нельзя бегло прочитать этот раздел работы, как нельзя бегло пересказать его. Все достоинство анализа именно в скрупулезнейшем выяснении всех сторон сложного процесса осуществления квазиэкспериментального исследования. Основанием, по которому квазиэкспериментальные схемы излагаются в определенной последовательности, является мера, с которой экспериментатор может управлять воздействием экспериментальной переменной, то есть вторгаться в реальный ход событий (именно с этой точки зрения оцениваются возможности таких типов исследований, как корреляционное, панельное, эксперимент *ex-post-facto*). Красной нитью при анализе всех десяти моделей квазиэкспериментальных исследований проходит мысль о том, что в случае невозможности применения лучших планов исследователь может использовать такой план, где отсутствует оптимальный контроль, но в этом случае он обязан всякий раз осуществлять проверку всех двенадцати перечисленных факторов валидности.

Организация материала в данной части работы позволяет рассматривать ее как справочное руководство, и потому возникает необходимость предостеречь от тех соблазнов, которые возможны при использовании любого справочника. Самым главным из них является желание рассматривать руководство как некую догму, найти в каждой строке предложенных таблиц точное указание на все возможные реальные трудности, встречающиеся в реальной исследовательской ситуации. Кэмпбелл решительно предупреждает

против такого использования его работы: «Для любой конкретной реализации каждого плана картина плюсов и минусов, скорее всего, будет отличаться от соответствующей строчки в таблице» (с. 184). Главный вывод, который делает сам автор, заключается в том, что всякое экспериментальное исследование в социальных науках представляет собой творческий процесс и при проектировании способов проверки гипотез в таком исследовании необходима, прежде всего, апелляция к анализу конкретной исследовательской ситуации. Такой вывод вновь возвращает нас к идеям ранее рассмотренной части книги, а именно к необходимости включения в систему социального познания наряду с количественными развитыми форм

качественного знания. Дальнейшая конкретизация этой идеи содержится в третьей части предлагаемой работы.

Мы обозначили третий логический блок проблем, рассматриваемых Кэмпбеллом, как анализ некоторых прикладных проблем социальных наук под углом зрения эффективности разрабатываемых для внедрения в практику программ. Ряд статей, содержащихся в книге, с разных сторон исследует вопрос о качестве планирования социальных экспериментов, которые осуществляются исследователями непосредственно для решения конкретных проблем, выдвигаемых обществом. Сам факт рождения многочисленных практических программ, предлагаемых социологами, педагогами, психологами и политологами, есть свидетельство возросшей роли социального знания в различных звеньях общественного организма. Иными словами, это означает, что общество многого ждет от общественных наук. Ответственность за качество рекомендаций заставляет разработать специальные критерии оценки составляемых программ.

Естественно, что в этом разделе Кэмпбелл остается наиболее «американским» автором, поскольку он анализирует те способы приложения социальных наук, которые характерны для американского общества. В этой связи для советского читателя данных работ Кэмпбелла отчетливо выявляются две перспективы. Во-первых, знакомство с содержанием ряда конкретных программ, которые в течение последних лет были выдвинуты представителями различных общественных наук в США: педагогами, социологами, психологами и др. И, во-вторых, своеобразное «вычленение» общей проблематики прикладного исследования в социальных науках: его принципиальных возможностей, места в системе социальной практики общества. Отношение к этим двум перспективам не может быть одинаковым.

Что касается первой перспективы, то она представляет, естественно, определенный познавательный интерес, поскольку содержит информацию о сравнительно мало известной стороне деятельности американских исследователей. В последние годы практика создания

ими разнообразных программ для нужд общества приобрела довольно широкое распространение. Так, например, много усилий было сконцентрировано на разработке программы «Head Start», предполагающей облегчение условий для начальных этапов образования неимущих слоев населения. И хотя содержание этой и других программ далеко от действительности нашего общества, знакомство с ними полезно для познания другого мира и его проблем. Однако в рамках этой первой перспективы для советского читателя открывается не только возможность приобретения новой информации, но и возможность наиболее острой полемики с Кэмпбеллом. Эта полемика неизбежна, прежде всего, по вопросам общего, скорее идеологического, чем чисто профессионального плана. Она касается принципиального понимания роли и значения отдельных программ, разрабатываемых представителями общественных наук в рамках буржуазного общества. В

работе Кэмпбелла не рассматривается вопрос о том, каковы *в принципе* возможности социального реформаторства в условиях данного типа общественных отношений. Он решает *другой* вопрос: насколько хороши предлагаемые программы с точки зрения их профессиональной обоснованности, надежности, обусловленной квалифицированным сочетанием методов их разработки? Конечно, эта вторая сторона анализа программ весьма важна, но она остается именно «второй стороной», — более острой является постановка вопроса о том, могут ли совершенные с профессиональной точки зрения программы дать действительный социальный эффект, если предлагаемые в них меры «сталкиваются» с более общими нерешенными проблемами общества. Та же программа «Head Start» порождает много размышлений не только о системе образования в США, но о противоречиях социальной структуры, о наличии расовой дискриминации и т. д. Можно ли ожидать решения и этих вопросов в случае успешного внедрения программы «Head Start» — вот проблема, которая, на наш взгляд, является главной и которая, к сожалению, обойдена в работе Кэмпбелла. Развести два аспекта оценивания качества практических программ, конечно, можно, но оба аспекта должны быть обозначены. Кэмпбелл же лишь весьма глухо говорит о значении социальных условий, которые должны

23

быть приняты в расчет. Поэтому полезная информация о конструировании представителями социальных наук практических «программ действия» порождает вместе с тем чувство неудовлетворенности, оттого что она не вписывается автором в более широкий контекст обсуждения данной проблемы.

Можно, конечно, согласиться, что обсуждаемый здесь угол зрения есть угол зрения, профессионально присущий лишь политику, но ведь Кэмпбелл постоянно подчеркивает, что усилия для составления «программ действия» предпринимаются всеми специалистами в области общественных наук. В этом случае вряд ли можно согласиться с игнорированием политических аспектов дискуссии. Но и с точки зрения методологии социального познания обсуждение вопроса только выиграло бы, если бы автор предложил, например, такую проблему, как соотношение объективных законов развития общества и сознательного воздействия на ход этого развития. Именно такой поворот дискуссии представляется естественным в рамках марксистской методологии. Коль скоро такой «макросоциальный» аспект дискуссии отсутствует, неизбежно остается чувство, что не рассмотрены весьма важные вопросы.

Более сильной представляется позиция Кэмпбелла, когда мы рассматриваем вторую перспективу, раскрывающуюся при знакомстве с третьей логической частью книги. Автор выступает здесь с трактовкой социальных реформ как своего рода социальных экспериментов, обходя «трудный» вопрос о возможности или невозможности социального реформаторства в условиях буржуазного общества и сосредоточиваясь именно на профессиональной стороне дела. Задача заключается в том, чтобы научиться измерять воздействие «запланированного социального изменения» на ход общественного развития. Под «запланированным социальным изменением» понимаются, прежде всего, те локальные программы, которые разработаны социальными науками. Кэмпбелл в данном случае очень точно отдает себе отчет в том, что эту чисто профессиональную задачу нельзя решить по отношению ко всему спектру социальных «нововведений», их значимость необходимо ранжировать: «Обсуждаемая мной технология непригодна для измерения социальных воздействий какой-

24

либо революции. Даже в рамках стабильной политической ситуации ее приложения следует, возможно, ограничить сравнительно второстепенными нововведениями» (с. 335—

336). Только по отношению к этим второстепенным нововведениям и обсуждается далее вопрос об эффективности практических программ. Нужно отдать должное автору, что в этой части своих рассуждений он ставит весьма точный и тонкий диагноз ситуации, сложившейся именно в США при оценивании эффективности таких программ. Во-первых, он выявляет различие в позиции тех, кто осуществляет, и тех, кто оценивает программы. Это различие сплошь и рядом переходит в конфликт, поскольку всякое оценивание качества программ неизбежно превращается в оценивание определенных социальных *институтов* и, следовательно, предполагает столкновение различных социальных сил. Во-вторых, при оценивании эффективности программ выявляется такой фактор, как *содержание* самих программ: одно дело — когда программа касается совершенно конкретного и в принципе решаемого вопроса, другое дело — когда оцениваются программы, посвященные решению «хронически нерешаемых вопросов». Наконец, в-третьих, возникает трудность, связанная с намерениями составителей программ: иногда это тенденция «предпринимать действие лишь на словах либо осуществлять косметические операции, рассчитанные скорее на то, чтобы убедить общественность в том, что меры принимаются, чем на решение проблемы» (с. 357).

Можно смело утверждать, что Кэмпбелл здесь демонстрирует высокую гражданскую позицию, позицию ученого, испытывающего чувство ответственности не только за формальные атрибуты качества социальных исследований, но и за их действительную социальную направленность. Конкретным анализом всевозможных обстоятельств, затрудняющих чисто профессиональную оценку программ, он отчасти восполняет те пробелы в общей постановке вопроса о возможностях социального реформирования, которые мы отмечали выше.

Содержание этой части работы Кэмпбелла как бы приводит к «общему знаменателю» всю совокупность его рассуждений в трех частях книги. Здесь получает

не только свое оправдание, но и свое доказательство тезис о том, что именно в социальных науках необходимо органическое единство количественных и качественных методов, что соображения здравого смысла имеют здесь право на существование и что, несмотря на затруднения, порожденные применением на практике результатов социальных исследований, ученый обязан сделать максимум возможного для обеспечения надежности его выводов. Здесь, на наш взгляд, возможна другая рекомендация, связанная с подготовкой будущих социальных исследователей: профессионализм в этой области не может ограничиваться лишь искусным владением измерительными процедурами; он неизбежно должен включать в себя и формирование у исследователя определенного «социального видения»... Переводя содержание этой рекомендации на тот язык, которым мы начали данный очерк, можно сказать, что диалог между *сциентизмом* и *гуманизмом* Кэмпбелл понимает как профессиональное («сциентистское»!) применение исследовательских процедур при глубоком («гуманистическом») понимании сущности изучаемых социальных проблем.

Именно такой подход Дональда Кэмпбелла выгодно отличает его от тех односторонних подходов, которые столь запальчиво защищаются порой сторонниками «крайних» взглядов. Это делает книгу, составленную из произведений Кэмпбелла, полезной и нужной также и советскому читателю.

**НЕСКОЛЬКО ЗАМЕЧАНИЙ  
ОТНОСИТЕЛЬНО ТЕОРИИ НАУКИ  
И МЕТОДОЛОГИИ ЭКСПЕРИМЕНТА  
В СОЦИАЛЬНОЙ ПСИХОЛОГИИ<sup>1</sup>**

Работа, которую я веду в последние годы, посвящена тому, чтобы придать социальным наукам, и в частности социальной психологии, по возможности более научный характер, выявить трудности, возникающие на пути научного вывода в социальных исследованиях и оценить научный статус тех типов выводов, к которым обычно приходят представители социальных дисциплин. Эти цели предполагают как разработку способов повышения валидности данных там, где это возможно, так и рекомендации быть осмотрительными и скромными там, где это необходимо.

В этой чрезвычайно широкой области я сосредоточиваю свои усилия на методах экспериментирования в полевых условиях, в первую очередь на квазиэкспериментальных моделях исследования, а также на теории и практике измерения, уделяя особое внимание вопросам валидности, погрешности, надежности и т. п. Это те аспекты методологии социальных наук, в которые я, как мне кажется, более всего готов внести свой вклад.

Эти цели соответствуют общественным интересам — разработке приемов оценки воздействия и эффективности социальных программ и выработке отличных от экономических показателей социальных индикаторов качества и образа жизни. На уровне статистики проблемы научного вывода в значительной мере те же,

---

<sup>1</sup> Campbell D. T. Some comments on theory of science and methodology of experimentation in social psychology.

что и в теоретической социальной науке. Однако применение методов социальных наук для оценки практических программ заставляет обратить внимание на некоторые собственно социальные проблемы, при решении которых помощь методологов также весьма полезна. Многие из этих проблем актуальны и для сугубо академического социального исследования.

Попытки достичь большей научности в такой проблематичной маргинальной области, как социальная наука, приводят к плодотворному взаимодействию с теорией науки, включая философию, социологию и историю науки, а также потенциально и психологию науки. Не только мы, методологи, нуждаемся в понимании природы научного метода, которое возникает на основе данных этих областей знания, но и наши проблемы также могут быть полезны им, так как связаны с постановкой фундаментальных вопросов, которые, возможно, никогда и не возникли бы, если бы эти отрасли знания ориентировались только на явно успешные научные дисциплины.

Указанные выше цели могут быть отнесены к трем основным областям: 1) теория науки, 2) социальное экспериментирование, выбор социальных индикаторов и оценивание социальных программ и 3) проблемы статистического анализа, планирования эксперимента и разработки моделей и планов эксперимента<sup>1</sup>.

Объединение этих трех направлений исследования можно, вероятно, оправдать и с рациональных позиций. Наука представляет собой коллективный продукт множества частичных знаний, полученных отдельными учеными. Основную роль в создании этого продукта играют фундаментальные дисциплины, которые охватывают

<sup>1</sup> В большинстве вышедших на русском языке работ по методологии экспериментальных исследований и статистическим методам в качестве эквивалента английского термина «design» обычно используется термин «план». Во избежание недоразумений при переводе работ, вошедших в данную книгу, мы следовали этой традиции. Однако мы считаем необходимым отметить, что наиболее точным было бы использование термина «дизайн», получившего права гражданства в таких областях, как эргономика, системное проектирование, техническая эстетика и др. Суть работ в этой области методологии науки — разработка проектов экспериментов — экспериментальный дизайн, проектирование и построение моделей эксперимента. — *Прим. ред.*

смежные специальности. Однако определенная роль принадлежит и таким формам научной компетенции, которые предполагают большую широту кругозора, на известный риск поверхностности и ошибочности. Подобный подход создает возможность интеграции данных ряда наук. Эта возможность была бы упущена, если бы имело место объединение лишь узкоспециализированных дисциплин, не выходящих за Пределы смежных областей<sup>1</sup>. Сказанное не означает, что мы объявляем поверхностной нашу прошлую и будущую работу в этих областях (в действительности мы обладаем в разработке некоторых специальных проблем в ряде случаев почти что монополией), просто мы допускаем, что, если бы все наши усилия были направлены на разработку какого-то одного аспекта, могла бы быть достигнута большая глубина, которая и необходима для коллективного продукта науки, подвергающегося перекрестной валидации в работах специалистов более узкого профиля, но при этом нам, по крайней мере, иногда, приходилось бы ловить себя на ошибках.

Три указанные выше области, составляющие предмет наших исследований, действительно релевантны друг другу, как это будет показано в более специальных работах. В целом вопросы общесоциального порядка теснейшим образом связаны как с возможностью применения экспериментальных моделей различного типа, так и с валидностью этих моделей и планов эксперимента. В современной философии, истории, социологии науки большое внимание уделяется взаимодействию социальных процессов с научными представлениями, при этом очерчивается круг вопросов, которые могут рассматриваться в самой непосредственной связи с более практическими проблемами оценивания программ и планирования эксперимента.

Традиционные термины «эпистемология» и «философия науки» не слишком подходят для большинства

---

<sup>1</sup> Методологические проблемы узкодисциплинарных и междисциплинарных исследований рассматривались специально в работе) C a m p b e l l D. T. Ethnocentrism of Disciplines and the Fish-Scale Model of Omniscience. — In: *Interdisciplinary Relationships in the Social Sciences*. Ed. by M. Sherif and Caroline W. Sherif. Chicago, Aldine, 1969.

впечатляющих исследований последних лет, все еще фигурирующих под этими названиями. В то время как классическая эпистемология пыталась решить проблему знания, не рассматривая всякое знание в процессе его приобретения, современные исследователи намеренно ставят этот вопрос и предпринимают эпистемологические исследования, считая, что сегодняшние физика и астрономия дают нам аппроксимативно валидное знание о познаваемом мире, а эволюционная биология, психология и социология дают нам знания о познающем его человеке. Куайн называет это «натурализованной эпистемологией», Шимони — «коперниканской эпистемологией». Исходя из традиционного различения аналитического и синтетического, ее можно назвать синтетической или, точнее, «дескриптивной эпистемологией». Такая эпистемология вполне совместима с традиционной эпистемологией в случае принятия

негативных результатов скептической традиции, например положения о необоснованности индуктивного вывода в научном знании. При аналитическом подходе ответ на традиционный эпистемологический вопрос «Возможно ли знание?» гласит: «Нет». Мы не можем быть уверены в том, что знаем, или знать, когда мы знаем, а когда нет. Поскольку дело касается знания о мире, наши самые обоснованные представления будут содержать допущения, которые мы не можем доказать, и любые дополнения к «знанию» (в практическом смысле) будут получены ценой принятия дополнительных предположений. Deskриптивная эпистемология, которая идет дальше этого, в свете традиционного различия будет более похожа на науку, чем на философию. Она будет гипотетическим, условным, презумптивным описанием субъекта познания и мира как объекта познания. В этих пределах она может все-таки иметь дело с проблемами истины: в какого рода мире, какого рода процедуры приведут познающее сообщество к повышению валидности его моделей мира? Но в то время как традиционная эпистемология пыталась строить эпистемологию без онтологии, любая deskриптивная эпистемология будет включать и гипотетическую онтологию. Следует оставить несбыточные мечты о достоверности, о доказательности представлений или даже о полной экспликации референтов представления.

### 30

В настоящее время наука не только стала главной ареной, на которой ведутся споры вокруг вопросов, поднимаемых deskриптивной эпистемологией, но во многом изменился и характер самой «философии науки». Она больше не довольствуется одним лишь логическим анализом статуса научной истины. Не только проявляется глубокий интерес к выявлению нормативных правил выбора, которыми *должна* пользоваться наука, осуществляя выбор между теориями, ведутся также и споры относительно тех правил выбора, которыми уже пользовалась наука (явно или неявно), когда она осуществляла *предположительно валидный* выбор в прошлом, и это может рассматриваться как гипотетический, условный поиск нормативных правил. Для обозначения подобных исследований термин «теория науки» технически более приемлем, чем «философия науки». К такой теории науки история науки имеет самое прямое отношение, что исключается при традиционном пуристском определении задачи философов и логиков науки. Книга Куна «Структура научных революций»<sup>1</sup> явно имеет эпистемологическое значение, очевидно и то, что она связана с вопросом о научной истине. Философы сразу восприняли ее как таковую. Однако в условиях принятых традиционных обозначений Кун вынужден был отрицать, что он занимается философией науки, и выдавать свою работу всего лишь за «историю науки», «социологию науки» или «социальную психологию науки». Суть его работы становится ясной в свете вопросов, оказавшихся на повестке дня deskриптивной эпистемологии, или теории науки. Объединенные под общей рубрикой теории науки, такие специальные дисциплины, как история науки, социология науки, психология науки, антропология науки и т. д., могут стать релевантными при решении вопроса о валидности науки, несмотря даже на то, что, как это ныне имеет место, они сплошь и рядом обходят этот вопрос.

За последние двадцать лет я написал, наверно, с дюжину статей, которые в настоящее время склонен рассматривать как работы в области deskриптивной

---

<sup>1</sup> Кун Т. Структура научных революций. М., «Прогресс», 1977.

### 31

эпистемологии<sup>1</sup>. В известной мере эти проблемы нашли отражение и в полдюжине моих статей о методах исследования, а также в ряде заметок, обзоров и лекций. Некоторые из этих работ вошли и в настоящий сборник.

В работах этого цикла затрагиваются не менее существенные вопросы, связанные с определенным пониманием социологии науки, обращающейся к выявлению социальных источников валидности науки и социальных условий, благоприятствующих развитию науки. Этот круг вопросов исследуется нами в контексте методологии социального экспериментирования и прикладных исследований. В одной из последних работ<sup>2</sup> мы рассмотрели ряд примеров макроэкспериментов, осуществленных на практике. Работа в этой прикладной области обладает своей спецификой. Мы столкнулись здесь с важнейшей социальной проблемой — природой бюрократической организации, оказывающей влияние и на планирование экспериментов, и на реализацию различных моделей экспериментов, и на валидность оценок разрабатываемых и осуществляемых социальных программ. Эти вопросы подлежат дальнейшему детальному изучению, но отчасти они затронуты и в нашей обобщающей работе об оценке социальных программ, вошедшей в данный том.

Общие вопросы теории науки, валидности научного вывода, прикладные проблемы валидной оценки социальных программ и эффективности социальных нововведений непосредственно связаны с более специальными моими исследованиями в сфере социально-психологического эксперимента. Важнейшими представляются здесь вопросы планирования эксперимента, создания моделей экспериментов в соответствии с требованиями

---

<sup>1</sup> См., например: Campbell D. T. Evolutionary Epistemology. — In: The philosophy of Karl Popper. Ed. by P. A. Schilpp. The library of living philosophers. La Salle, Ill. The Open Court Publishing Co., 1974, v. 14, p. 413—463; Campbell D. T. Variation and selective retention in socio-cultural evolution — In: Social change in developing areas: a reinterpretation of evolutionary theory. Cambridge, Mass., Schenkman, 1965, p. 19—49.—Прим. ред.

<sup>2</sup> См.: Riecken H. W., Boruch R. F., Campbell D. T., Caplan N., Glennan T. K., Pratt J., Rees A., Williams W. Social Experimentation: A Method for Planning and Evaluating of Social Intervention. New York, Academic Press, 1974.

валидности научного знаний, разработки средств и способов контроля переменных в эксперименте, анализа артефактов и конкурентных гипотез, выявления приемлемых и адекватных статистических процедур и количественных мер и т. д. Исследования в этой области также проводятся мною уже многие годы. В известной мере эти вопросы получили отражение в более ранней работе, написанной совместно с Дж. Стэнли, — «Экспериментальные и квазиэкспериментальные планы для исследования». Эта книга многократно переиздавалась в США и широко используется рядом исследователей. Хотя эта работа была первоначально ориентирована главным образом на педагогические исследования, она имеет более общее методологическое значение. Указанные проблемы рассматривались также в работе, написанной совместно с Т. Д. Куком, в которой собраны и проанализированы различные типы и планы экспериментов в прикладных областях и в полевых исследованиях, в частности в сфере организационной и индустриальной психологии<sup>1</sup>. Особенно плодотворной, по оценке многих авторов, следует признать разрабатываемую нами концепцию квазиэкспериментального исследования<sup>2</sup>. Проблемам квазиэкспериментов посвящена подготовленная нами совместно с Т. Д. Куком специальная монография<sup>3</sup>.

В заключение мы хотели бы выразить надежду, что вопросы, рассматриваемые в работах, включенных в данный сборник, представят интерес и для широкого круга советских специалистов — психологов, социологов, методологов науки, практических работников.

---

<sup>1</sup> Cook T. D., Campbell D. T. The Design and Conduct of Quasi-Experiments and True Experiments in Field Settings. — In: Handbook of Industrial and Organizational Psychology. Ed. by Dunnette M. Chicago, Rand McNally, 1976.

<sup>2</sup> См.: Г л а с с Дж., С т э н л и Дж. Статистические методы в педагогике и психологии. М., 1976, с. 452—453.

<sup>3</sup> C o o k T. D., C a m p b e l l D. T. Quasi-Experimentation. Chicago, Rand McNally, 1979.

## ЭКСПЕРИМЕНТЫ И КВАЗИЭКСПЕРИМЕНТЫ <sup>1</sup>

Экспериментом мы называем ту часть исследования, которая заключается в том, что исследователь осуществляет манипулирование переменными, и наблюдает эффекты, производимые этим воздействием на другие переменные. В этом очерке мы рассмотрим валидность 16 планов эксперимента в связи с 12 факторами, которые могут угрожать валидности вывода по данным эксперимента. Наша работа не является руководством по планированию эксперимента в традиции Фишера [37, 38], когда экспериментатор, полностью владея ситуацией, может организовать воздействие и осуществить измерения для достижения оптимальной статистической эффективности и когда сложность плана обусловлена лишь стремлением к достижению эффективности. Сложность планов, обсуждаемых в данной работе, является следствием специфики материала и отсутствия у экспериментатора возможности полного контроля. Хотя в настоящей работе и есть некоторые точки соприкосновения с традицией Фишера, для знакомства с последней лучше обратиться к следующим источникам:

---

<sup>1</sup> C a m p b e l l D.T., S t a n l e y J. C. Experimental and Quasi-Experimental Designs for Research. Chicago, Rand McNally, 12th ed., 1976. © 1963 by the American Educational Research Association.

Работа публикуется с некоторыми сокращениями, а также уточнениями и добавлениями, внесенными Д. Т. Кэмпбеллом специально для данного издания.

Редактор-составитель благодарит проф. Дж. Верча (Северо-Западный университет США) за консультацию по поводу некоторых терминологических вопросов, возникших при переводе работ Д. Т. Кэмпбелла. — *Прим. ред.*

Brownlee [8], Cox [28], Edwards [33], Ferguson [36], Johnson [53], Johnson, Jackson [54], Lindquist [68], McNemar [75], Winter [136], а также Stanley [109].

## ПОСТАНОВКА ПРОБЛЕМЫ

В 1923 г. У. Э. Мак-Кол опубликовал книгу под названием «Как проводить эксперименты в педагогике» [74]. В предисловии к ней он писал: «Существуют прекрасные учебники и руководства по статистической обработке экспериментальных данных, но мало что можно найти по методам получения адекватных относящихся к делу данных, к которым применимы статистические процедуры». Слова эти остаются справедливыми и по сей день и могут служить лейтмотивом данной работы. Хотя под влиянием работ, выполненных в традиции Фишера, ситуация в некоторых существенных отношениях изменилась, все же наиболее заметным следствием этих работ является, видимо, совершенствование статистического анализа, а не средств получения «адекватных относящихся к делу данных».

Вероятно, ввиду ее ориентации на практику и здравый смысл и отсутствия претензии на фундаментальность классическая книга Мак-Кола не была оценена по достоинству. В то время, когда она появилась, то есть за два года до первого издания книги Фишера «Статистические методы для исследователей» (1925), не было ни одной работы, которая могла бы идти хоть в какое-то сравнение с книгой Мак-Кола ни в области агробиологии, ни в психологии. Она предвосхитила ортодоксальную методологию этих областей по ряду фундаментальных вопросов.

Возможно, наиболее значительным вкладом Фишера является идея уравнивания групп перед экспериментом путем рандомизации. Принять эту идею и вместе с тем отказаться от «попарного уравнивания» (способа интуитивно привлекательного, но фактически ошибочного) оказалось трудным для исследователей в области педагогики. Мак-Кол же еще в 1923 г. продемонстрировал глубокое понимание существа дела. В качестве своего первого метода формирования сопоставимых групп он предложил «группы, уравненные случайно». «Точно так же, как репрезентативность может обеспечиваться

35

с помощью метода случайного выбора... эквивалентность может быть достигнута случайно, при условии использования достаточно большого числа испытуемых» [74, с. 41]. Была предвосхищена и другая идея Фишера. Под названием «ротационного эксперимента» был предложен план латинского квадрата, который применялся Торндайком, Мак-Колом и Чепмэном [124] как в форме 5x5, так и 2x2 еще в 1916 г., то есть почти за 10 лет до того, как Фишер включил его в свою систему планирования эксперимента с применением рандомизации<sup>1</sup>.

Способ использования Мак-Колом «ротационного эксперимента» служит прекрасным примером для выявления специфики как его книги, так и данной работы. Ротационный эксперимент был предложен не с целью достижения статистической эффективности, а, скорее, для обеспечения некоторой степени контроля в тех случаях, когда невозможно осуществить рандомизацию при формировании эквивалентных групп. Сходным образом в настоящей работе будут рассмотрены недостатки многочисленных планов эксперимента, и тем не менее будет отстаиваться их использование в тех случаях, когда невозможно применение лучших планов. В этом смысле большинство обсуждаемых здесь планов, в том числе и нерандомизированный «ротационный эксперимент», обозначаются как квазиэкспериментальные планы.

### **Разочарование в эксперименте**

В данной работе эксперимент рассматривается как единственный способ разрешения споров о практике обучения, как единственный метод верификации усовершенствований в обучении, как единственный путь создания традиции накопления знаний, при которой каждое нововведение может быть осуществлено без опасения, что старая мудрость приносится в жертву менее совершенному, но более новому знанию. При всей нашей горячей защите экспериментирования не следует думать, что наша позиция нова. Были периоды как пессимистического отношения к возможностям эксперимента, так

---

<sup>1</sup> Кендалл и Бакленд [61] пишут, что латинский квадрат был изобретен Эйлером в 1782 г. Торндайк, Мак-Кол и Чепмэн не пользуются этим термином.

36

и энтузиазма (см., например: Good, Scates [43], Monroe [79]). Можно отметить, что отход от экспериментирования к описательным работам, нередко сопровождавшийся поворотом от

торндайковского бихевиоризма к гештальтпсихологии или психоанализу, совершался людьми, получившими хорошую экспериментальную подготовку.

Чтобы предупредить новые разочарования, мы должны понять некоторые причины предшествующей реакции и попытаться избежать неверных ожиданий, которые вызвали такую реакцию. Можно отметить несколько моментов. Во-первых, утверждалось, что темп и степень прогресса, которые должен был обеспечить эксперимент, оценивались слишком оптимистически, и это сопровождалось неоправданным пренебрежением к неэкспериментальному знанию. Сторонники экспериментирования считали, что прогресс в технологии обучения был медленным *именно потому*, что не применялся научный метод. Они рассматривали традиционную практику как несостоятельную лишь потому, что она не была основана на экспериментировании. Когда же выяснилось, что эксперименты часто оказывались скучными, двусмысленными и подчас невоспроизводимыми и подтверждали донаучные знания, чрезмерный оптимизм уступил место разочарованию, пренебрежению ими и отказу от них.

Это разочарование испытали как наблюдатели, так и участники экспериментов. Что касается экспериментаторов, то можно отметить, что у них выработалась некая реакция избегания экспериментирования. Для исследователя, обычно глубоко заинтересованного в своей работе, неподтверждение выношенной гипотезы является весьма болезненным. Экспериментатор тоже подчиняется биологическим и психологическим законам научения и связывает свои огорчения со смежными во времени событиями и стимулами. Таким стимулом станет, скорее всего, сам экспериментальный процесс, а не «истинный» источник фрустрации — неадекватная теория. Это может привести, возможно, неосознанно, к избеганию или отказу от экспериментирования. Если экология нашей науки такова, что в ней неправильные ответы встречаются чаще, чем правильные (а, по-видимому, так оно и есть), то мы можем ожидать, что боль-

шинство экспериментов окажется неутешительным, и мы обязаны каким-то образом подготовить начинающих экспериментаторов к этому и вообще смотреть на экспериментирование более пессимистически — не как на панацею, а как на единственно возможный путь постоянного накопления знаний. Мы должны внушить нашим ученикам, что они могут столкнуться с рутинной и разочарованиями, и воспитывать в них упорство, которое отличает в настоящее время исследователей в биологических и физических науках. Наши ученики должны примириться не только со скудостью денежных средств, но и со скудостью результатов экспериментов.

Точнее, нам необходимо более значительное время и осознание того, что длительное и многократное экспериментирование более характерно для науки, чем единичные окончательные эксперименты. Если эксперименты, которые мы проводим сегодня, являются успешными, то они нуждаются в повторении и перекрестной валидации в другое время и в других условиях, прежде чем они смогут стать признанными данными науки и получить обоснованную теоретическую интерпретацию. Далее, хотя мы считаем экспериментирование основным средством доказательства, единственным судьей в споре конкурирующих теорий, мы, по-видимому, не должны ожидать однозначных результатов «критических» экспериментов, в которых сталкиваются соперничающие теории. Когда мы обнаружим, например, что компетентные наблюдатели защищают прямо противоположные точки зрения, то можно утверждать, почти что априорно, что оба они наблюдали нечто действительно относящееся к реальной ситуации и оба представляют часть истины. И чем сильнее противоречие, тем больше вероятность того, что так оно и есть. Поэтому в таких случаях мы можем ожидать неоднозначных результатов эксперимента или результатов, в которых истинные данные неуловимо меняются от эксперимента к эксперименту. Более зрелой позицией — а к такой в значительной мере пришла экспериментальная психология

(см., например, Underwood [129]) — является стремление избегать критических экспериментов и заменять их исследованиями соотношений и взаимодействий разнообразных многоуровневых экспериментальных переменных.

Не следует также недооценивать важность значительно усовершенствованных статистических методов, которые постепенно проникли в психологию и педагогику. В период наибольшего увлечения экспериментом в педагогике применялись весьма несовершенные методические средства. Мак-Кол [74] и его современники проводили исследования с одной переменной. Ввиду огромной сложности ситуации обучения человека этот подход оказался слишком ограниченным. Теперь мы знаем, насколько важным может быть совместное действие двух или более экспериментальных переменных. Стэнли [108, 110, 112, 113, 114], Стэнли и Уайли [116] и другие указали на важность оценки таких взаимодействий.

Эксперимент может быть многомерным в следующих двух отношениях. План эксперимента может содержать более чем одну «независимую» переменную (пол, год обучения, метод обучения арифметике, тип и размер шрифта в учебнике и т. д.) и/или более чем одну «зависимую» переменную (число ошибок, скорость, количество правильных ответов, данные различных тестов и т. д.). Процедуры, введенные Фишером, многомерны в первом смысле и одномерны во втором. Специалисты по математической статистике, например, Рой и Гнанадесикан [95], разрабатывают планы и методы анализа, объединяющие оба типа многомерных планов. Возможно, что, обратившись к ним, исследователи смогут уменьшить обычный огромный разрыв между изложением статистических методов в специальной литературе и их практическим применением в исследованиях.

Несомненно, более основательное знакомство исследователей с *современной* экспериментальной статистикой поможет повысить качество экспериментальных исследований.

## ВИДЫ ВАЛИДНОСТИ<sup>1</sup>

### Проверка каузальных отношений

Эксперименты являются средством проверки каузальных гипотез. По традиции предполагается, что су-

---

<sup>1</sup> Проблемы валидности – центральная проблема в этом цикле исследований Кэмпбелла. К анализу валидности автор обращается

ществуют три необходимых условия для утверждения с определенной долей уверенности о том, что отношение между двумя переменными является каузальным и что направление причинности от *A* к *B*. Первое условие относится к временной последовательности и формулируется следующим образом: причина должна предшествовать по времени эффекту. Обычно это условие соблюдения нетрудно, если исследователь знает, когда испытуемые подвергались экспериментальному воздействию. В квазиэкспериментах исследователь может сочетать такого рода влияние со своим знанием результатов испытуемых до и после экспериментального воздействия и может связать воздействие с некоторыми показателями изменения зависимой переменной. Исследователь, проводящий истинный эксперимент, знает, что соответствующая рандомизация обеспечивает, в

вероятностном плане, предварительную (до введения экспериментального воздействия) эквивалентность различных экспериментальных групп. Поэтому если после экспериментального воздействия обнаружатся различия, связанные с воздействием, то он полагает, что такие различия, по всей вероятности, вызваны экспериментальным воздействием.

Второе необходимое условие для обоснованного вывода о причинной зависимости  $B$  от  $A$  состоит в том, что воздействие или воздействия должны быть статистически связаны с эффектом, так как, если возможная причина и эффект не связаны друг с другом, одно не может быть причиной другого. Для проверки существования такой ковариации используются статистические показатели; разработаны критерии для принятия решения о существовании «действительной» ковариации в результатах (например,  $p < 0,05$ ). Таким образом, статистические показатели действуют как фильтры. К, сожалению, они подвержены ошибкам даже в тех случаях, когда используются должным образом; с их помощью не всегда удастся выявить как действительно

---

в большом числе своих работ. Данный раздел был впервые опубликован в работе: Cook T. D., Campbell D. T. The Design and Conduct of Quasi-Experiments and True Experiments in Field Settings. — In: Handbook of Industrial and Organizational Psychology. Ed by M. Dunnette. Chicago, Rand McNally, 1976, p. 223—326. © 1976, Rand McNally Inc. — *Прим. ред.*

40

существующую, так и ложную композицию ковариаций. Поскольку использование статистических показателей ведет к весьма важным решениям, разумно было бы показать те причины, которые порождают неверные выводы о ковариации. Обозначим их как угрозы для *валидности статистического вывода*.

Третье необходимое условие каузального вывода состоит в том, что не должно быть правдоподобного альтернативного, объяснения появления  $B$  помимо  $A$ . Это условие труднее всего соблюсти в связи с другими обстоятельствами. Первое — относящееся в основном к квазиэкспериментам — связано с валидностью альтернативных интерпретаций, которые предполагают, что кажущаяся каузальная связь  $A$  и  $B$  фактически может быть обусловлена действием третьей переменной, которая и вызывает изменения  $B$ . К примеру, можно установить на фабрике новую машину и отмечать, связано ли это с ростом производительности труда. Если это так, то можно приписать это увеличение нововведению. Однако всегда существует вероятность того, что это увеличение никак не связано с новой машиной и может быть обусловлено сезонным ростом производительности, который происходит в такое время каждый год. И это только один из примеров такой третьей переменной, и в дальнейшем мы представим систематизированный перечень таких дополнительных переменных, обозначая их как угрозы для *внутренней валидности*.

Существование угроз для внутренней валидности наводит на мысль, что предполагаемая каузальная связь может быть лишь видимой. Это отличается от второго значения «альтернативной интерпретации», когда  $A$ , которым манипулируют, и  $B$ , которое измеряется, и в самом деле каузально связаны, но существуют сомнения, действительно ли эмпирические операции осуществляются с объектами, представленными конструктами  $A$  и  $B$  которые исследователь пытался применить для обозначения  $A$  и  $B$ . Большинство теоретических контroversий в психологии именно такого рода — например, контroversии о том, можно ли объяснить связь между увеличением оплаты и более высокой производительностью тем, что высокая зарплата формирует чувство неравенства или разрушает ожидания или «Я- концепцию» и т. д. В таких исследованиях речь

41

идет *не* о внутренней валидности, то есть не о том, являются ли изменения в оплате причиной изменений в производительности. Скорее, это вопрос о том, как оплата должна быть выражена в терминах соответствующей теории, доступных обобщению. Другой пример: для некоторых исследователей проблема интерпретации известных хоуторнских экспериментов Ротлисбергера и Диксона<sup>1</sup> заключается в обозначении того, что заставило женщин увеличить производительность труда, а не в определении того, действительно ли экспериментальное вмешательство привело к этому увеличению. Был ли каузальной переменной сам факт изменения независимо от его характера, или обратная связь, несущая информацию о поведении обследуемого и содержащаяся в новых изменениях, или же это следствие возрастания групповой сплоченности, или следствие нового восприятия интересов руководства, или еще что-нибудь? Мы рассматриваем опасности для *конструктивной валидности*, которые следует понимать как угрозы для правильного обозначения причины и эффекта с помощью абстрактных терминов, взятых из обыденного языка или из формальной теории. В действительности проблемы конструктивной валидности несколько шире и, очевидно, имеет отношение к попыткам обозначения любых аспектов эксперимента, включая характер обстановки, в которой проводится эксперимент, особенности участвующих лиц и т. д.

Стоит отметить, что термин «внутренняя валидность» в прошлом употреблялся не вполне правильно, потому что с его помощью выражали как сомнения в правильности вывода о причинных связях между *A* и *B*, так и сомнения в том, как следует обозначать причину и следствия. Такое смешение понятий может возникнуть вследствие того, что альтернативные интерпретации должны быть отброшены в процессе установления как внутренней, так и конструктивной валидности.

Однако установление внутренней валидности предполагает отбрасывание альтернативных интерпретаций вероятных каузальных связей между *A*, которым манипулируют, и *B*, которое измеряется, а установление конструктивной валидности предполагает отбрасывание аль-

---

<sup>1</sup>Roethlisberger F. J., Dickson W. J. Management and the Worker: Cambridge, Harvard Univ. Press, 1939.

тернативных интерпретаций того, как *A* и *B* соотносятся с гипотетическими понятиями. Поскольку в экспериментах крайне важно выяснить, является ли связь между двумя переменными каузальной, *необходимым* условием для выведения заключения о каузальности является отбрасывание альтернативных интерпретаций связи между переменными (то есть устранение угроз для внутренней валидности), а не отбрасывание альтернативных интерпретаций операций с причиной и следствием (то есть угроз для конструктивной валидности).

Хороший эксперимент (а) делает ясной временную последовательность; (б) достаточно чувствителен и действен для того, чтобы показать, что вероятные причина и эффект взаимосвязаны (ковариантны); (в) исключает возможность влияния третьих переменных, которыми можно было бы объяснить связь между причиной и эффектом; а также (г) исключает альтернативные гипотезы о конструктах, включенных в эту связь. Полезно сделать еще один шаг. Заключение о каузальной связи в какой-то промежуток времени, в одной обстановке и при одной выборке испытуемых дает мало уверенности в том, что обнаруживаемая каузальная связь является устойчивой. Вопрос о возможности обобщения результатов для различных периодов времени, различных условий и групп мы рассматриваем как вопрос о *внешней валидности*, и вкратце перечислим угрозы и для этого вида валидности.

Сказанное не должно создавать впечатления, что эксперимент является единственным средством установления каузальности. Такая наука, как астрономия, развивалась без

экспериментирования отчасти потому, что ей посчастливилось иметь надежные методы наблюдения и количественные теории, позволяющие предсказывать точное местоположение, точные орбиты и временные интервалы движения в пространстве. Точность численных предсказаний означает, во-первых, что предсказания могут быть проверены самым тщательным образом и, во-вторых, что различные теории, на основе которых делаются различные численные предсказания, могут быть противопоставлены друг другу. Это не значит, что в астрономии решены все проблемы валидности или что исследователь может отказаться от детального анализа

43

всех возможных альтернативных гипотез и тщательного сопоставления их с данными для того, чтобы выяснить, могут ли они быть отброшены. Мы хотим сказать только, что было бы меньше угроз валидности, если бы измерения были такими же надежными, как в астрономии, а теории столь же точными.

К сожалению, социальным наукам не посчастливилось иметь столь же точные теории, столь же надежные измерения или повторяющийся циклический порядок в результатах наблюдений. Представьте себе, что обнаружено различие в производительности труда работника до и после того, как он прошел специальное обучение. Как мы можем определить, что в этом различии следует отнести за счет собственно обучения, что за счет естественного развития работника, или улучшений в навыке выполнения задания при обследовании, или в результате различных событий в жизни человека, которые оказывают влияние на зависимую переменную между предварительным тестированием и тестированием после экспериментального воздействия<sup>1</sup>, и что объясняется любым сочетанием этих факторов? Более того, даже если бы мы могли определить конкретные численные величины, связанные с каждым из перечисленных выше объяснений, можно ли быть уверенными в том, что нам удастся измерить соответствующие результаты с такой надежностью, которая позволила бы осуществить выбор между теориями? Ответ на эти вопросы зависит, конечно, от величины предсказываемых различий в приросте производительности труда и от конкретного вида проверки. Мы полагаем, однако, что в социальных науках найдется не много неэкспериментальных условий, в которых точные предсказания могли бы быть с успехом использованы для проверки конкурентных гипотез.

Хотя предшествующее изложение наводит на мысль, что эксперименты более пригодны для проверки предположений о причинности, чем неэкспериментальные исследования, из этого не следует делать вывод, что эксперименты являются надежным средством получения

---

<sup>1</sup> Термин «тестирование» в данной работе используется в широком смысле – как обследование, проверка, измерение – и лишь в редких случаях для обозначения процедуры использования психометрических и диагностических тестов. – *Прим. ред.*

44

ответов на все вопросы, связанные с проверкой гипотез о каузальных связях. Перечень угроз внутренней и внешней валидности, валидности статистического вывода и конструктивной валидности<sup>1</sup>, а также выявление взаимосвязей между этими видами валидности делают очевидным несовершенство экспериментирования. Экспериментальные планы нуждаются в улучшении, могут и должны быть улучшены, чтобы способствовать установлению каузальной зависимости. Но мы заблуждались бы, если бы считали, что один эксперимент или даже целая исследовательская программа, рассчитанная на несколько лет, могли бы дать окончательный ответ на основные вопросы, связанные с надежным установлением каузальной связи, обозначением ее составляющих и определением возможности обобщения вывода.

## Факторы, угрожающие внутренней и внешней валидности<sup>2</sup>

В следующих разделах приводится 12 факторов, угрожающих валидности различных экспериментальных планов. Каждый фактор будет подробнее описан при обсуждении тех планов, в которых его действие проявляется наиболее отчетливо, и этот список будет завершен после обсуждения 10 из 16 планов. Основным при составлении этого списка было различие *внутренней* и *внешней валидности*.

*Внутренняя валидность* — это тот минимум, без которого не может быть интерпретирован ни один эксперимент: действительно ли именно это экспериментальное воздействие привело к изменениям в данном эксперименте? *Внешняя валидность* относится к возможности *обобщения* вывода: на какие популяции, ситуации, другие независимые переменные, параметры воздействия и переменные измерения могут быть распространены результаты эксперимента. Несомненно, что оба типа критериев являются важными, хотя часто они находятся в противоречии в том смысле, что увеличение валидности одного типа может угрожать валидности другого типа. Хотя *внутренняя валидность* есть условие

---

<sup>1</sup> Два последних вида валидности в данной работе специально не рассматриваются. — *Прим. ред.*

<sup>2</sup> Большая часть этого параграфа основана на работе Кэмпбелла [10].

*sine qua non*<sup>1</sup> и хотя проблема *внешней валидности*, как и проблема индуктивного вывода, никогда не может быть полностью решена, очевидно, что нашим идеалом является выбор таких экспериментальных планов, которые обеспечивают оба типа валидности. Это особенно важно для педагогических исследований, где весьма желательным является распространение данных на практические ситуации, определенного характера. Различия и взаимоотношения между этими двумя видами валидности станут яснее после рассмотрения примеров при обсуждении конкретных типов эксперимента.

Далее будут представлены восемь различных классов внешних переменных, относящихся к *внутренней валидности*. Если эти переменные не контролируются в экспериментальном плане, то они могут дать эффекты, которые смешиваются с эффектом экспериментального воздействия. Эти переменные представляют влияние:

1) *фона* (history) — конкретных событий, которые происходят между первым и вторым измерением наряду с экспериментальным воздействием;

2) *естественного развития* (maturation) — изменений испытуемых, являющихся следствием течения времени *per se* (не связанных с конкретными событиями), например взросление, усиление голода, усталости и т. п.;

3) *эффекта тестирования* (testing) — влияния выполнения заданий, применяемых для измерения, на результаты повторного испытания;

4) *инструментальной погрешности, нестабильности измерительного инструмента* (instrumentation), при которой изменения в калибровке инструмента или изменения, характеризующие наблюдателя или оценочные показатели, могут вызвать изменения в результатах измерения;

5) *статистической регрессии* (statistical regression), имеющей место тогда, когда группы отбираются на основе крайних показателей и оценок;

6) *отбора испытуемых* (selection) — неэквивалентности групп по составу, вызывающей появление систематической ошибки в результатах;

7) *отсева в ходе эксперимента* (experimental mortality) — неравномерности выбывания испытуемых из сравниваемых групп;

8) *взаимодействий фактора отбора с естественным развитием* и др., которые в ряде квазиэкспериментальных планов с несколькими группами (таких, как план 10) ошибочно принимаются за эффект экспериментальной переменной<sup>1</sup>.

К факторам, ставящим под угрозу *внешнюю валидность*, или репрезентативность эксперимента, относятся:

9) реактивный эффект, или *эффект взаимодействия тестирования*, — возможное уменьшение или увеличение сензитивности, или восприимчивости, испытуемых к экспериментальному воздействию под влиянием предварительного тестирования. Результаты лиц, прошедших предварительное тестирование, будут нерепрезентативны по отношению к тем, кто не подвергался предварительному тестированию, то есть тем, из кого состоит генеральная совокупность, из которых были отобраны испытуемые;

10) *эффекты взаимодействия фактора отбора и экспериментального воздействия*;

11) *условия организации эксперимента, вызывающие реакцию испытуемых на эксперимент*, которая не позволяет распространить полученные данные о влиянии экспериментальной переменной на лиц, подвергающихся такому же воздействию в неэкспериментальных условиях;

12) *взаимная интерференция экспериментальных воздействий*, нередко возникающая, когда одни и те же испытуемые подвергаются нескольким воздействиям, поскольку влияние более ранних воздействий, как правило, не исчезает. Это относится особенно к планам экспериментов с одной группой (планы типа 8 и 9).

Для представления различных характеристик экспериментальных планов в дальнейшем будет применяться постоянная система графических и символических обо-

---

<sup>1</sup> В других работах Д. Т. Кэмпбелла этот список факторов, ставящих под угрозу внутреннюю валидность, уточняется и дополняется. Так, приводится еще шесть факторов: 1) сомнение по поводу направления причинной связи; 2) диффузия или имитация воздействия; 3) компенсаторное уравнивание воздействия; 4) компенсаторное соперничество; 5) деморализация испытуемых, вызванная обидой в связи с обделенностью желаемым воздействием; 6) особенности фона в данной группе, взаимодействие факторов отбора и фона (см.: Соок Т. Д., Сампелл Д. Т. Op. cit., p. 228 — 229). — Прим. ред.

значений. Символ  $X$  будет обозначать экспериментальную переменную или событие, влияние которого подлежит измерению; символ  $O$  — некоторый процесс наблюдения или измерения<sup>1</sup>;  $X$  и  $O$ , стоящие в одной строке, относятся к одним и тем же конкретным лицам. Направление слева направо обозначает временной порядок, а расположение  $X$  и  $O$  одно под другим — одновременность. Для обозначения важных различий, например между планами 2 и 6 или между планами 4 и 10, используется символ  $R$ , указывающий на случайное распределение испытуемых по разным режимам эксперимента (то есть по различным группам). Предполагается, что такая рандомизация осуществляется в определенное время и служит универсальным методом уравнивания групп перед введением воздействия в известных статистических границах. Наряду с этим вводится дополнительная графическая символика: параллельные строки, не разделенные пунктирной линией, представляют группы, уравненные посредством рандомизации, тогда как сопоставляемые группы, не уравненные рандомизацией, отделяются друг от друга пунктирной линией. Не предусматривается никакого символа для обозначения групп, сформированных путем попарного уравнивания, поскольку ценность этой процедуры слишком переоценивается, и она чаще ведет к ошибочному выводу, чем служит средством достижения валидного

вывода. (См. далее обсуждение плана 10 и последний раздел по корреляционным планам.) Символ  $M$  для обозначения материалов будет использован в плане 9.

## ТРИ ДОЭКСПЕРИМЕНТАЛЬНЫХ ПЛАНА

### 1. Исследование единичного случая

В настоящее время многие педагогические исследования строятся по плану, предусматривающему проведение однократного обследования на единственной группе, подвергшейся воздействию, которое, как предполагается, и вызывает изменения. Такие исследования

---

<sup>1</sup> Автор этим символом обозначает как саму процедуру измерения, так и ее результаты. — *Прим. перев.*

могут быть схематически представлены следующим образом:

$X$   $O$ .

Как уже отмечалось (см., например, Boring [5], Stouffer [120]), в таких исследованиях полностью отсутствует контроль, так что они почти не представляют научной ценности. Этот план приводится здесь, чтобы служить в качестве нижней точки отсчета. Но поскольку в такого рода исследования по-прежнему вкладываются средства и на основании таких исследований делаются выводы о причинных связях, требуются некоторые пояснения. Основным для научных результатов (а также для всех когнитивно-диагностических процессов, включая те, что происходят в сетчатке глаза) является процесс сравнения, регистрации различий или контраста. Анализ показал, что существование абсолютного знания или знания сущности отдельных изолированных объектов является иллюзией. Получение научных результатов связано с проведением, по меньшей мере, одного сравнения. Для того чтобы из такого сравнения можно было извлечь пользу, данные об обеих его сторонах должны быть получены с одинаковой точностью и тщательностью.

В случае плана 1 тщательно исследованное единичное событие имплицитно сравнивается с другими случайно наблюдаемыми и запоминающимися событиями. Вывод основывается на предположениях о том, какими могли бы быть данные, если бы не было  $X$  и т. п. Часто результаты таких исследований представляют собой унылое нагромождение конкретных деталей, скрупулезных наблюдений, измерений и т. п., причем точности в таких случаях добиваются не там, где надо. Насколько более ценным могло бы быть исследование, если бы этот ряд наблюдений был сокращен наполовину, а сэкономленные усилия направлены на столь же тщательное изучение соответствующего случая для сравнения. По-видимому, в настоящее время считается неэтичным представление такого типа исследований единичного случая, то есть с одной единственной раз наблюдаемой группой, в качестве диссертаций по педагогике. Не лучше обстоит дело и с применением «стандартизированных» тестов, поскольку всегда обнаружи-

вается такое множество конкурентных источников различий, не связанных с  $X$ , что «стандартная» (нормативная) группа становится в качестве «контрольной» почти бесполезной. По тем же самым причинам нет надежды, что с такими исследованиями могут быть сопоставлены возможные будущие исследования, ведь существует огромное число

неконтролируемых источников различий между изучаемым случаем и теми, которые будут исследоваться в будущем. В целом предпочтительней распределить исследовательские усилия на обе стороны интересующего нас сравнения.

План 1, если предполагать использовать его данные для сравнения с обыденным знанием, обладает большинством недостатков каждого из последующих планов. Поэтому обсуждение таких недостатков будет проведено ниже при анализе этих конкретных случаев.

## 2. План с предварительным и итоговым тестированием на одной группе

Хотя этот план все еще находит широкое применение в педагогических исследованиях и считается лучше, чем план 1, в тех случаях, когда нет ничего лучшего (см. далее обсуждение квазиэкспериментальных планов), здесь он приводится в качестве «дурного примера» для иллюстрации некоторых примешивающихся к  $X$  посторонних факторов, которые могут поставить под угрозу *внутреннюю валидность*. Эти факторы дают основание для правдоподобных гипотез, объясняющих различие между  $O_1$  и  $O_2$  и конкурирующих с гипотезой о том, что причиной этого различия является  $X$ :

$$O_1 \quad X \quad O_2.$$

Первой из этих неконтролируемых конкурентных гипотез является *фон*. Помимо экспериментального воздействия  $X$ , между  $O_1$  и  $O_2$  может происходить множество других событий, вызывающих изменения. Если предварительное ( $O_1$ ) и итоговое ( $O_2$ ) тестирования проводятся в разные дни, то причиной различия между ними могут оказаться некоторые события, происшедшие в промежутке между тестированиями. Чтобы стать *правдоподобной* конкурентной гипотезой, такое событие

50

должно произойти с большинством испытуемых экспериментальной группы, например, в часы совместных занятий или вследствие широко распространившихся известий. В исследовании Колльера [23] (проведенном в 1940 г., а опубликованном в 1944 г.) в то время, когда студенты изучали материалы по нацистской пропаганде, пала Франция, и наблюдавшееся изменение в установках явилось, по-видимому, следствием скорее этого события, чем знакомства с самой по себе пропагандой<sup>1</sup>. Чем больше промежуток времени между  $O_1$  и  $O_2$ , тем вероятнее, что *фон* является правдоподобным конкурентным объяснением зарегистрированного изменения. Когда же эксперимент длится всего один или два часа, можно считать, что его роль незначительна, хотя и в этом случае надо следить за возможными посторонними воздействиями, например возникновением смеха, отвлекающих внимание событий, и т.д. К этой переменной (*фону*) имеет отношение такое явление, как *экспериментальная изоляция*, которая достигается во многих физических лабораториях в такой мере, что позволяет считать план 2 приемлемым для значительной части исследований в физике. Столь эффективная экспериментальная изоляция почти никогда не возможна при исследовании, например, методов обучения. Поэтому в табл. 1, в колонке «фон» для плана 2 стоит знак «минус». Мы относим к *фону* возможные влияния времени года или событий, возникающих на институциональном уровне, хотя такие влияния можно относить и к *естественному развитию*. Так, оптимизм может зависеть от времени года, а тревожность — быть связана с приближением экзаменационной сессии (см., например, Crook [29]; Windle [135]). Влияние таких событий на изменение результатов от  $O_1$  к  $O_2$  может быть ошибочно принято за действие  $X$ .

Второй конкурентный фактор или класс факторов обозначен как *естественное развитие*. Здесь этот термин используется для обозначения всех тех биологических или

психологических процессов, которые независимо от конкретных внешних событий систематически изменяются с течением времени. Так, например, в промежутке

---

<sup>1</sup> В действительности Колльер использовал более адекватный план, обозначенный в данной системе как план 10.

## Источники невалидности для экспериментальных планов 1—6

	Источники невалидности																	
	внутренней								внешней									
	фон	естественное развитие	эффект тестирования	инструментальная погрешность	регрессия	состав групп	выбывание	взаимодействие состава групп с естественным развитием и др. факторами	взаимодействие тестирования и X	взаимодействие состава групп и X	реакция испытуемых на эксперимент	взаимодействие между разными X						
<i>Доэкспериментальные планы</i>																		
1. Исследование единичного случая	X	O	—	—	—	—	—	—	—	—	—	—						
2. План с предварительным и итоговым тестированием на одной группе	O	X	O	—	—	—	—	?	+	+	—	—	—	?				
3. Сравнение статических групп	X	O	—	—	—	—	—	—	+	?	+	+	+	—	—	—		
	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—		
<i>Планы истинных экспериментов</i>																		
4. План с предварительным и итоговым тестированием и контрольной группой	R	O	X	O	—	—	—	—	+	+	+	+	+	+	+	—	?	?
	R	O	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—

				Источники невалидности												
				внутренней						внешней						
				фон	естественное развитие	эффект тестирования	инструментальная погрешность	регрессия	состав групп	выбывание	взаимодействие состава групп с естественным развитием и др. факторами	взаимодействие тестирования и X	взаимодействие состава групп и X	реакция испытуемых на эксперимент	взаимодействие между разными X	
5.	План Соломона для 4-х групп			+	+	+	+	+	+	+	+	+	+	+	?	?
	R	O	X	O												
	R	O		O												
	R		X	O												
	R			O												
6.	План с контрольной группой и тестированием только после воздействия			+	+	+	+	+	+	+	+	+	+	+	?	?
	R	X	O													
	R		O													

Примечание: В таблицах «минус» означает безусловную слабость плана, «плюс» указывает на то, что фактор находится под контролем, вопросительный знак — на возможность некоторых осложнений, а пробел — на иррелевантность фактора.

Мы приводим эти таблицы весьма неохотно, поскольку они могут оказаться «слишком полезными» и применяться вместо более сложного и квалифицированного анализа. Ни один «плюс» или «минус» не должен браться на вооружение, пока читатель не поймет, почему поставлен данный знак. Мы не хотели бы, чтобы эти таблицы вызвали у читателя безотчетный страх перед теми или иными планами или, наоборот, доверие к ним.

между  $O_1$  и  $O_2$  испытуемые могли стать старше, проголодаться, устать, их может одолеть скука и т. д., и полученное различие может оказаться следствием этих процессов, а не действия X. При компенсаторном обучении, которому подвергаются исключительно лица

с явными нарушениями развития, за эффект корректирующего X может быть ошибочно принят процесс «спонтанной ремиссии», аналогичный затягиванию раны. (Не приходится и говорить, что такую ремиссию нельзя считать «спонтанной» в каком-либо каузальном

смысле, точнее, эта ремиссия представляет собой кумулятивный эффект процессов обучения и влияния среды, который проявился бы и без применения  $X$ ).

Третьим возможным объяснением, могущим быть источником погрешности, является эффект *тестирования*, то есть влияние предшествующего обследования на результаты последующих. При тестировании интеллекта или уровня знаний испытуемые обычно показывают лучшие результаты при повторном выполнении того же теста или его альтернативной формы по сравнению с лицами, впервые подвергающимся испытанию (см., например, Anastasi [2, с. 190—191]; Cane, Heim, [17]). Для нетренированных испытуемых такое улучшение в среднем на 3—5 единиц по шкале умственного развития ( $IQ$ ) происходит без каких-либо указаний или инструкций по поводу тех пунктов, в которых первый раз была допущена ошибка. Сходное явление отмечено и в исследованиях с применением личностных тестов — повторное использование теста обычно выявляет лучшую приспособленность, хотя иногда наблюдается и весьма значительный обратный эффект (Windle [135]). Применение повторного теста при изучении отношений к национальным меньшинствам может обнаружить большую степень предубежденности, хотя этот эффект выражен незначительно (Rankin, Campbell [91]). Очевидно, на характер результатов оказывают влияние условия анонимности обследования, возрастающая осведомленность о том, какой ответ более социально приемлем, и т. д., При исследованиях предрассудков в условиях анонимности уровень адаптации, создаваемый предъявляемыми суждениями, выражающими враждебность, может изменить представления испытуемых о приемлемых установках в сторону большей враждебности. При применении неанонимных вопросников для изучения личностных характеристик и приспособленности первое тестирование приобретает характер проблемной ситуации, в которой испытуемый пытается выявить скрытую роль теста. Когда ему это удается (или после обсужде-

ния со своими друзьями их ответов на некоторые «странные» вопросы), он лучше понимает, как показать себя во второй раз в более выгодном свете.

С проблемой эффекта тестирования связана проблема различных способов измерения *реактивности* испытуемого. Эта тема весьма важна для данной работы, и поэтому мы всегда будем подчеркивать желательность применения нереактивных в этом смысле измерений<sup>1</sup>. В социальных науках давно уже известно, что процесс измерения может сам по себе изменять то, что измеряется. Увеличение показателей от предварительного тестирования к итоговому является важным аспектом таких изменений. (Другой аспект — взаимодействие между измерением и  $X$  — будет обсуждаться ниже в связи с планом 4. Кроме того, важно избегать возможной реактивности на начальное тестирование даже в тех случаях, когда процедура измерения оказывает различное воздействие на разных испытуемых.) Реактивности можно ожидать во всех тех случаях, когда процесс тестирования сам по себе является скорее стимулом к изменению, чем пассивной регистрацией поведения. Так, например, начальное взвешивание в экспериментах с регулированием веса само по себе, даже без терапевтического воздействия, может служить стимулом для уменьшения веса. Аналогично этому появление в классе наблюдателя для предварительной оценки умений преподавателя устанавливать и поддерживать межличностные отношения может само по себе изменить стиль поведения преподавателя. Вид подключенного микрофона может изменить характер группового взаимодействия и т. д. В целом можно ожидать, что, чем более мотивирующим и привлекающим внимание будет процесс измерения, тем значительнее могут быть побочные реакции.

Термин *инструментальная погрешность*, или «дефект инструмента» (Campbell [10]), используется для обозначения четвертой неконтролируемой конкурентной гипотезы. Он относится к автономным изменениям в

---

<sup>1</sup> Проблеме нереактивных измерений в социальных исследованиях посвящена работа: Webb E. J., Campbell D. T., Schwartz R. D., Sechrest L. Unobtrusive measures: Nonreactive research in the social sciences. Chicago, Rand McNally, 1966. — *Прим. ред.*

измерительных инструментах, которые могут привести к различиям между  $O_1$  и  $O_2$ . Эти изменения аналогичны растяжению или потере упругости в пружинных весах, конденсации в камере Вильсона и т. д. При использовании для получения значений  $O_1$  и  $O_2$  наблюдателей их утомление, научение и т. д. приведут к различиям между  $O_1$  и  $O_2$ . При выставлении оценок за сочинения могут измениться эталоны оценивания при переходе от  $O_1$  и  $O_2$ . (Можно преодолеть этот недостаток, если оценивать сочинения из  $O_1$  и  $O_2$  в один прием так, чтобы не было известно, какие написаны раньше.) При обследовании школьных занятий наблюдатели могут стать более опытными или более пресыщенными к моменту повторного тестирования. Если проводится интервьюирование родителей, то причиной отличий данных первого интервью от данных второго может быть лучшее знакомство интервьюера с порядком проведения интервью и с опрашиваемыми. Изменения самого наблюдателя, происшедшие между  $O_1$  и  $O_2$ , могут привести к различиям в результатах наблюдений.

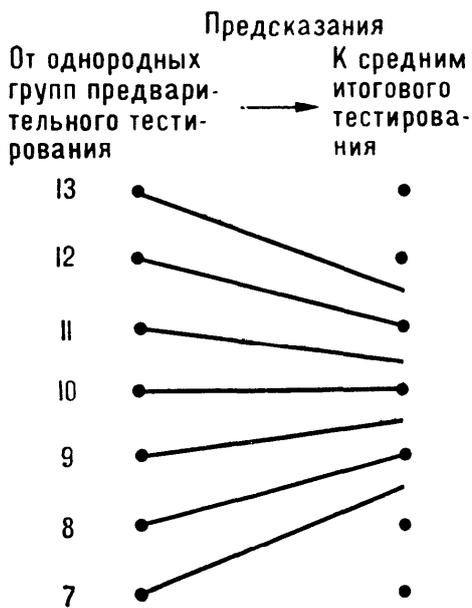
Пятым фактором, примешивающимся к  $X$  в некоторых случаях использования плана 2, является эффект *статистической регрессии*. Так, если для экспериментов с корректирующим обучением учащиеся выбираются потому, что они показали наихудшие результаты в тестах достижений (и эти результаты будут для них  $O_1$ ), то последующее тестирование с использованием того же теста или аналогичной его формы покажет, что  $O_2$  для этой группы испытуемых почти наверняка будет в среднем выше, чем  $O_1$ . Этот несомненный результат не обусловлен каким-либо подлинным эффектом  $X$ , обучением в ходе тестирования и т. д. Скорее всего, это проявление отсутствия 100%-ной корреляции между  $O_1$  и  $O_2$ . Поскольку ошибки, возникающие вследствие недооценки эффекта регрессии, столь опасны в педагогических исследованиях, поскольку понимания сущности этого эффекта нет даже у студентов, прошедших более сложный курс современной статистики, и в дальнейшем обсуждении (например, плана 10 и анализа *ex post facto*<sup>1</sup>) предполагается знание этого вопроса, мы изложим здесь эту проблему, хотя и в несколько упрощенной и традиционной форме. На рис. 1 представлены некоторые гипотезы

---

<sup>1</sup> После того, как факт произошел (*лат.*).



а



б

Рис. 1б



в

Рис. 1в

Р и с. 1. Регрессия для предсказания результатов итогового тестирования по данным предварительного тестирования, и наоборот. 1а. Разброс результатов итогового тестирования для каждого класса результатов предварительного тестирования, и наоборот.

тетические данные, где коэффициент корреляции между результатами предварительного и итогового тестирования для всей популяции равен 0,50 при одном и том же среднем и одинаковой дисперсии по группе. (Данные подобраны таким образом, чтобы расположение значений по столбцам и строкам было легко обзримым. Значение коэффициента 0,50 также выбрано по соображениям удобства.) В этом гипотетическом примере нет подлинного изменения, но, как обычно, ретестная корреляция показателей теста значительно меньше единицы. Если, как в приведенном примере, мы будем сравнивать с данными итогового тестирования только результаты тех учащихся, которые получили при первом тестировании самые низкие оценки — к примеру 7,— то обнаружим, что, хотя результаты этих учеников имеют определенный разброс, в целом они выше и в среднем «приблизились» к среднегрупповым данным на половину отрезка, отделявшего их от среднего по группе (поскольку коэффициент регрессии или корреляции составляет 0,50), и равны теперь 8,5. Однако это вовсе не свидетельствует о реальном прогрессе, а является лишь тавтологическим, хотя и специфическим, подтверждением факта отсутствия полной корреляции.

Поскольку между первым и вторым тестированием прошло определенное время, и произошли некоторые события, возникает искушение связать это изменение с естественным течением времени. Но ведь в данном случае возможно и рассуждение в обратном направлении: выделим тех, кто при втором тестировании получил 7 очков, и посмотрим на их результаты при начальном тестировании. Мы обнаружим, что показатели вначале были более высокими. К наибольшим ошибкам при попытках определения причинных связей ведет представление данных, показанное на рис. 1б (это относится как к верхней, так и к нижней части рис. 1б). Здесь все выглядит так, будто умные стали глупее, а глупые — умнее, что могло бы быть результатом действия институциональных факторов, приводящих к нивелированию и гомогенизации. Если бы такое представление соответствовало фактическому положению, то разброс данных последующего тестирования был бы ниже, чем предварительного, но в действительности он в обоих случаях одинаков. Более того, если бы мы нача-

ли свой анализ с подгруппы, однородной в отношении результатов последующего тестирования, например с линии регрессии  $c$  и рис. 1в, то пришли бы к противоположному выводу. Как отмечал Мак-Нимар [75], для предупреждения такого рода ошибок интерпретации полезно провести реверсивный контрольный анализ и обратиться непосредственно к данным об изменениях вариабельности в популяции.

Мы можем рассматривать регрессию в направлении средних значений и несколько по-другому. Чем сильнее отклоняется оценка от среднего значения, тем больше вероятность того, что она связана с ошибкой измерения. Так, как правило, получающему самые высокие оценки в некотором смысле необычайно «везет» (большая позитивная ошибка), а получающему самые низкие оценки «не везет» (большая негативная ошибка). Но везение — вещь капризная, и мы вправе ожидать, что при последующем тестировании показатели тех, кто получил высокие оценки, несколько сдвинутся к среднему по группе, а получившие низкие оценки улучшат свои позиции. (Та же логика действительна и в том случае, если мы начнем с анализа данных итогового тестирования и обратимся затем к данным предварительного тестирования.)

Регрессия в направлении среднего — широко распространенное явление, и она связана не только с исследованиями, в которых тест применяется повторно (или используются его эквивалентные формы). Директор, который обнаружит, что учащиеся с самыми высокими показателями  $IQ$  получают не самые высокие (хотя и достаточно высокие) оценки в тестах достижений, тогда как учащиеся с наиболее низким коэффициентом умственного развития обычно обнаруживают не самые низкие (хотя и довольно низкие) результаты в тестах до-

стижений, и на этом основании решит, что в его школе недооценивают наиболее способных учеников и переоценивают неспособных, был бы введен в заблуждение именно этим фактом регрессии. Столь же нелогичным был бы и противоположный вывод, основанный на сравнении показателей групп учащихся, подобранных на основе наивысших и самых низких оценок в тестах достижения, с их показателями в тесте умственного развития.

59

Хотя проблемы регрессии обсуждались здесь в связи с ошибками измерения, в более широком плане эффект регрессии является функцией степени корреляции: чем ниже корреляция, тем в большей степени выражена регрессия к среднему. Отсутствие полной корреляции данных может быть вызвано как «ошибками», так и систематическими источниками дисперсии, характерными для того или иного количественного показателя.

Поэтому для групп, подобранных на основе крайних показателей, неполная корреляция результатов предварительного и итогового тестирования неизбежно сопровождается эффектами регрессии. Однако эти эффекты возникают не всегда, когда мы встречаемся с экстремальными показателями. Если средний показатель группы, сформированной по независимым основаниям, окажется экстремальным, то в этом случае меньше априорных оснований ожидать, что среднее группы будет регрессировать при повторном измерении, так как случайные или посторонние источники дисперсии могут вызывать смещение исходных показателей в обоих направлениях. Другое дело, если группа подобрана на основании крайних показателей по переменной, подверженной ошибкам. Ее экстремальность является артефактом, и она регрессирует в сторону среднего той популяции, из которой была выбрана данная группа.

Более косвенные эффекты регрессии могут быть вызваны отбором по другим экстремальным показателям, а не по данным предварительного тестирования. Рассмотрим случай, когда учащиеся, которые «провалились» на контрольной, отбираются в группу экспериментального обучения. Начальное тестирование проводилось с помощью «Формы А» стандартного теста достижений, а при повторном тестировании применяется «Форма В». Вполне вероятно, что результаты контрольной имеют более высокую корреляцию с результатами почти одновременно используемой «Формы А», чем с результатами, полученными при предъявлении после трехмесячного перерыва «Формы В» (если тест предлагается каждый раз всему классу). Чем выше степень корреляции, тем меньше регрессия к среднему. Таким образом, при проведении предварительного тестирования оценки неудачников смещаются к среднему в меньшей степени, чем при проведении итогового тестирования,

60

что может создать ошибочное впечатление успешного влияния корректирующего обучения. (Для более детального изучения этой проблемы см.: Lord [70, 71]; McNemar [77]; R. L. Thorndike [125]).

На этом мы закончим перечень недостатков плана 2, которые уместно рассмотреть на данном этапе. Если обратиться к табл. 1, то можно обнаружить еще один минус, относящийся к внутренней валидности и связанный с фактором, который мы будем рассматривать при обсуждении плана 10 в разделе о квазиэкспериментальных планах, и два недостатка, относящиеся к внешней валидности, которые будут рассмотрены при обсуждении плана 4.

### 3. Сравнение статических групп

Третий доэкспериментальный план, который мы используем при обсуждении источников невалидности,— это сравнение статических групп. Это план, где для установления эффекта  $X$  группа, подвергающаяся воздействию  $X$ , сравнивается с группой, не испытавшей такого воздействия:

$$\frac{X \quad O_1}{O_2}$$

Примерами такого рода исследований могут служить сравнение школ, в которых от преподавателя требуется наличие степени бакалавра ( $X$ ), с теми школами, где не предъявляется такое требование; сравнение учащихся, прошедших курс обучения скоростному чтению, с теми, кто не проходил такого курса; сравнение тех, кто смотрел определенную телевизионную программу, с теми, кто не смотрел ее, и т. д. В отличие от «истинного» эксперимента, представленного планом 6, в этих примерах использования плана 3 отсутствуют какие-либо формальные средства, удостоверяющие, что сравниваемые группы были бы эквивалентными без  $X$ . Отсутствие эквивалентности обозначается на схеме пунктирной линией, разделяющей две группы, и говорит о том, что необходим контроль за фактором отбора, или состава группы. Если  $O_1$  и  $O_2$  различны, это различие может быть обусловлено различием способов отбора

61

испытуемых в группы: группы могли бы, так или иначе, отличаться и при отсутствии  $X$ . Как будет показано ниже (при анализе плана *ex post facto*), попарное уравнивание на основе характеристик, отличных от  $O$ , не дает желательных результатов, оно обычно неэффективно и вводит в заблуждение, особенно в тех случаях, когда испытуемые «экспериментальной группы» стремятся подвергнуться воздействию  $X$ .

Последняя соперничающая переменная может быть обозначена как *выбывание*. Она связана с возникновением различий  $O_1 — O_2$  в группах в результате различного выбывания испытуемых из сравниваемых групп. Так, даже если бы в плане 3 две группы и были когда-то эквивалентны, к моменту проведения сравнения они могут отличаться, но не по причине каких-либо изменений, происшедших с членами групп, а по причине селективного выбывания испытуемых из одной группы. В педагогических исследованиях с этой проблемой чаще всего сталкиваются в работах, цель которых — установить влияние обучения в колледже путем сравнения характеристик новичков, не подвергавшихся воздействию  $X$ , и выпускников, подвергавшихся такому воздействию. Если в подобных исследованиях устанавливается, что, к примеру, вновь поступившие девушки более красивы, чем выпускницы, то мы воздержимся от вывода, что наш суровый способ обучения пагубно влияет на красоту, а скорее осмелимся утверждать, что красивые девушки чаще выбывают из колледжа до завершения обучения в связи с замужеством. Это явление обозначается здесь как экспериментальное *выбывание*. (Конечно, если бы мы сравнивали *тех же самых* девушек, когда они поступили и когда они окончили колледж, такая проблема не возникла бы, и мы имели бы дело с планом 2.)

### ТРИ ИСТИННЫХ ЭКСПЕРИМЕНТАЛЬНЫХ ПЛАНА

Три основных плана, которые будут рассматриваться в данном разделе, обычно и рекомендуются в методологической литературе. Мы также рекомендуем использовать в первую очередь именно эти планы, хотя это связано со многими конкретными ограничениями, налагае-

62

мыми обычной практикой, и хотя эти планы не всегда обеспечивают *внешнюю валидность* (см. табл. 1). Чаще всего используется план 4, поэтому ему и уделено наибольшее внимание. Надо отметить, что все три плана представлены так, как будто отдельное воздействие  $X$  уравнивается с *отсутствием*  $X$ . Планы, которые в соответствии с фишеровской традицией многофакторного планирования эксперимента предполагают несколько независимых переменных, не имеют прямого отношения к основной теме данного раздела и будут обсуждаться в конце раздела после рассмотрения плана 6. Но в связи с этим необходимо отметить здесь, что сравнение  $X$  с без  $X$  является очень сильным упрощением. Фактически всегда проводится сравнение экспериментальной группы, которая подвергается воздействию  $X$ , с контрольной группой, которая в это время осуществляет какую-то конкретную деятельность. А это значит, что можно говорить скорее о сравнении  $X_1$  и  $X_c$ , или  $X_1$  и  $X_0$ , или  $X_1$  и  $X_2$ . Такая деятельность контрольной группы часто является не установленным точно дополнением и создает неопределенность при интерпретации воздействия  $X$ . Имея это в виду, мы, тем не менее, и в дальнейшем изложении будем пользоваться обозначением без  $X$  для контрольной группы.

#### 4. План с предварительным и итоговым тестированиями и контрольной группой

##### Средства контроля «внутренней валидности»

По тем или иным из приведенных выше соображений между 1900 и 1920 гг. исследователи в области психологии и педагогики начали добавлять к плану 2 контрольную группу, создав тем самым современный ортодоксальный план с контрольной группой. В работах Мак-Кола [74], Соломона [104] и Боринга [5] излагается история этого плана, а просмотр «Teacher College Record» за этот период показывает, что уже к 1912 г. контрольные группы использовались как нечто само собой разумеющееся (см., например, Pearson [88]). Эти планы вводятся в данной работе под двумя рубриками: в плане 4 уравнивание групп достигается

63

путем рандомизации, в квазиэкспериментальном плане 10 используются реально существующие группы без уверенности в том, что они эквивалентны. План 4 имеет следующий вид:

$$\begin{array}{cccccc} R & O_1 & X & R & O_2 & \\ & R & O_3 & & O_4 & \end{array}$$

Поскольку применение этого плана дает возможность осуществить почти полный контроль за *всеми* семью рассмотренными до сих пор конкурентными гипотезами, необходимость контроля при его использовании обычно не эксплицируется. В традиционных исследованиях научения к признанию необходимости контрольных групп привело выявление эффектов *тестирования*. В педагогических исследованиях главной проблемой было *естественное развитие*, как в детской психологии — проблема «врожденное — приобретенное». В исследованиях изменения установок, например в ранних исследованиях влияния кино, наиболее важным является необходимость контролировать *фон*. Во всяком случае, полезно, видимо, кратко обсудить здесь, как или при каких условиях осуществляется контроль за этими факторами.

*Фон* («история») контролируется в той мере, в какой общие «исторические» события, которые могут привести к различию  $O_1$  —  $O_2$ , способны привести также к различию  $O_3$  —  $O_4$ . Однако отметим, что во многих возможных применениях плана 4 (а также 5 или 6) *не*

контролируются специфические события, имевшие место в группе в ходе обследования («внутригрупповая история»). Если все испытуемые экспериментальной группы, сформированной на основе рандомизации, подвергаются однократному экспериментальному воздействию и контрольная группа также исследуется однократно, то посторонние частные события, происходящие в каждой из этих групп (шутка, вызвавшая оживление, пожар на улице, предварительные замечания экспериментатора и т. д.), могут стать конкурентными объяснениями отличия  $O_1 - O_2$  от  $O_3 - O_4$ . Эксперименты такого рода *не* являются истинными экспериментами, хотя они иногда и приводятся в качестве примера, как в случае с экспериментом Соломона [104] по обучению произношению. (Справедливости ради необходи-

64

мо сказать, что этот пример выбран с целью иллюстрации других моментов.) Если рассматривать проблему, имея в виду наши «лучшие образцы» экспериментального исследования, то может показаться, что эксперименты по такому плану — простительный грех, но наши «лучшие образцы» слишком часто невозпроизводимы, а поэтому вполне возможно, что именно внутригрупповые события и есть тот самый источник «значимых», но посторонних различий, который является дефектом нашей исследовательской практики. Кроме того, в типичном эксперименте из публикуемых в «Journal of Experimental Psychology» работ контроль внутригрупповых событий обеспечивается тем, что тестирование испытуемых проводится в индивидуальном порядке и путем случайного распределения опытов и испытуемых по экспериментальным или контрольным режимам. Тем не менее, надо отметить, что даже при индивидуальной работе *фон* может оказаться неконтролируемым, если вся экспериментальная группа проходит тестирование до контрольной и т. д. План 4 рассчитан на одновременное проведение исследования в экспериментальной и контрольной группах. Если же действительно проводить опыты одновременно, то необходимо использовать различных экспериментаторов, при этом различия между экспериментаторами в свою очередь становятся одним из видов внутригрупповой истории, смешиваемой с  $X$ .

Оптимальным решением является рандомизация отдельных сеансов эксперимента с учетом тех ограничений, которые связаны с необходимостью уравнивания таких важных источников смещений, как экспериментатор, время дня, день недели, начало или конец семестра, близость экзаменов и т. п. Общепринятый способ объединения испытуемых в группы вместо индивидуальной работы с каждым неприемлем, если это группирование игнорируется при статистическом анализе. (См. ниже раздел об экспериментировании с реальными группами.) Все испытуемые, участвующие в одном сеансе, подвергаются воздействию одних и тех же посторонних событий, что может оказаться отличной от  $X$  причиной сходства их реакций. Если такие сеансы будут проводиться в случайном порядке, то корректными с точки зрения статистики будут такие процедуры предъявления воздействия, которые рассматриваются

65

ниже при обсуждении того, как проводить экспериментальные исследования с целыми классами. (Для некоторых исследований, связанных с групповым тестированием, можно распределить экспериментальные воздействия и предъявлять их отдельным лицам внутри одной группы в случайном порядке, подобно тому, как это делается при использовании множественных форм теста, когда изучается влияние порядка трудности отдельных заданий по тесту. В таких случаях особенности «внутригрупповой истории» будут общими для обоих видов процедур и при объяснении полученных различий они не смогут стать правдоподобной конкурентной гипотезой, смешиваемой с  $X$ .)

*Естественное развитие и эффект тестирования* контролируются постольку, поскольку они одинаковым образом проявляются в экспериментальной и контрольной группах. В тех случаях, когда созданы условия для контроля «внутригрупповой истории», легко контролировать и *инструментальную погрешность*, особенно если для *O* используются ответы испытуемых, полученные с помощью фиксированной методики, такой, как напечатанный тест. Однако при использовании наблюдателей или интервьюеров проблема становится более сложной. В случае, когда число наблюдателей не позволяет осуществлять их рандомизированное распределение по отдельным сеансам, необходимо использовать каждого наблюдателя и в экспериментальных, и в контрольных сеансах; кроме того, для предупреждения искажения результатов наблюдений необходимо, чтобы наблюдатели не знали, каким видам воздействий подвергается каждый из испытуемых. То, что такие тенденции к смещению результатов являются частыми источниками различий, подтверждается необходимостью введения «двойного слепого эксперимента»<sup>1</sup> в медицинских исследованиях, как было показано в недавно проведенных работах (Rosenthal [94]) и в более ранних исследованиях (например, Kennedy, Uphoff [62], Stanton, Baker [117]). Регистрация группового взаимодействия облегчит контроль инструментального фактора в исследованиях поведения и взаимодействия в группах учащихся, поскольку отдельные фрагменты записи, по-

---

<sup>1</sup> Эксперимент, при котором ни испытуемый, ни экспериментатор не знают, имеет место воздействие или нет. — *Прим. ред.*

лученные при предварительном и итоговом тестировании в экспериментальной и контрольной группах, могут предъявляться «судьям» в случайном порядке.

Эффект *регрессии* контролируется, когда речь идет о различиях между средними, независимо от того, насколько экстремальными были показатели группы при предварительном тестировании, если экспериментальная и контрольная группы были составлены из некоторой экстремальной совокупности на основе рандомизации. В этом случае контрольная группа имеет такую же регрессию, что и экспериментальная. Однако даже при использовании плана 4 часто возникают ошибки интерпретации из-за регрессионных артефактов. Экспериментатор может использовать контрольную группу для подтверждения группового среднего эффекта *X*, а затем при определении того, какие предварительные показатели подгрупп испытуемых из экспериментальной группы особенно подверглись влиянию *X*, оставить результаты контрольной группы без внимания. Если у экспериментальной группы в целом обнаружится улучшение результатов, то экспериментатор столкнется с поразительным явлением, состоящим в том, что у тех, кто вначале имел самые низкие показатели, это улучшение было наибольшим, в то время как у тех, кто вначале имел наивысшие показатели, не наблюдается никакого прогресса. В этих условиях такие результаты достигаются за счет того, что регрессионный артефакт способствует увеличению показателей у тех, кто имел наихудшие предварительные показатели, и нейтрализует их у тех, кто имел наилучшие предварительные показатели. (Если же в целом не наблюдается улучшения результатов, то экспериментатор может ошибочно «обнаружить», что это произошло вследствие двух взаимно компенсирующих процессов: прироста у «низких» и потери у «высоких».)

Одно из средств предупреждения такой ошибки интерпретации — проведение параллельного анализа экстремальных показателей предварительного тестирования в контрольной группе и использование этих данных при интерпретации увеличения показателей путем сравнения результатов итогового тестирования соответствующих — выделенных по данным предварительного тестирования — экспериментальных и контрольных под-

групп. (Отметим, что асимметрия распределений, получаемых при отборе подгрупп, ставит под сомнение пригодность процедур, основанных на нормальной кривой.)

Действие фактора *отбора, или состава групп*, в качестве объяснения различий исключается в той степени, в какой рандомизация обеспечивает эквивалентность групп в момент  $R$ . Эта степень определяется принятой нами выборочной статистикой. Поэтому уверенность в равенстве выше для большого числа случайных распределений, чем для малого. В пределах, определяемых величиной ошибки, для гипотезы об отсутствии различий это утверждение иногда оказывается неверным. В случае плана 4 это означает, что иногда могут обнаруживаться явно «значимые» различия между результатами предварительного тестирования. Поэтому, хотя простая, или стратифицированная, рандомизация и обеспечивает несмещенное распределение испытуемых по группам, она не дает абсолютной уверенности в исходной эквивалентности таких групп. Тем не менее, рандомизация — единственный существенный способ решения. Это утверждение столь категорично потому, что в течение последних 30 лет в педагогических исследованиях, наблюдается весьма распространенное ошибочное предпочтение способа попарного уравнивания групп. Мак-Кол [74] и другие способствовали этому недоразумению. Как будет более детально показано ниже, при обсуждении плана 10 и анализа *ex post facto*, попарное уравнивание фактически не способствует преодолению исходных групповых различий. Это не исключает применения попарного подбора в качестве дополнения к рандомизации, когда для достижения статистической точности подбирают учащихся по парам, а затем в случайном порядке одного члена каждой пары включают в экспериментальную группу, а другого — в контрольную. В статистической литературе этот прием известен как «образование блоков» (см. более подробное обсуждение этого вопроса в: Cox [27], Feldt [35], Lindquist [68]). Но применение попарного уравнивания вместо рандомизации недопустимо даже в квазиэкспериментальных планах, использующих лишь две реально существующие группы: одну — экспериментальную, другую — контрольную. Даже в этих «слабых

экспериментах» существуют лучшие, чем попарное уравнивание, способы уменьшить исходные средние значения различия в этих двух выборках.

Данные, которые можно получить с помощью плана 4, позволяют установить, может ли *выбывание* объяснить прирост  $O_1$  —  $O_2$ . Потери, выбывания и случаи, когда доступны лишь частичные данные, представляют немалые трудности, и обычно о них не упоминают. Как правило, эксперименты с методами обучения длятся в течение нескольких дней, недель или месяцев. Если предварительное и итоговое тестирования проводятся в классах, из которых составляются экспериментальная и контрольная группы, и условия эксперимента требуют присутствия испытуемых на занятиях, в то время как испытуемые контрольной группы эти занятия не посещают, то различная посещаемость на трех этапах — предварительное тестирование, экспериментальное воздействие, итоговое тестирование — ведет к выбыванию. Этот фактор может вызвать неуловимые смещения выборки. Если исключить из экспериментальной группы тех, кто первоначально был намечен для участия в ней, но не смог посещать экспериментальные занятия, это вызовет такое селективное сокращение группы, которое сделает ее несопоставимой с контрольной группой (в частности, экспериментальная группа оказывается «более добросовестной»). По-видимому, более предпочтительным было бы — хотя это редко наблюдается — использование результатов всех испытуемых контрольной и экспериментальной групп, которые подвергались предварительному и итоговому тестированию, включая в экспериментальной группе и тех,

кто не подвергался воздействию  $X$ . Очевидно, при таком способе анализа эффект  $X$  выглядит ослабленным, но зато это позволяет избежать смещения выборки.

Такая процедура основывается на допущении, что нет более простых смещений вследствие выбывания; такое допущение может быть частично проверено путем анализа числа и предварительных показателей тех лиц, кто проходил предварительное тестирование, но отсутствовал при итоговом тестировании. Возможно, что некоторые  $X$  влияют скорее на степень выбывания, а не на изменение индивидуальных показателей. Конечно, даже если степень выбывания одна и та же, остается

возможность сложных взаимодействий, которые делают различным характер выбывания из экспериментальной и контрольной групп.

Проблема выбывания выступает в гораздо более острой форме в тех исследованиях *компенсаторного* обучения, в которых используется принцип *добровольного* участия. Рассмотрим пример, в котором одна выборка плохо читающих школьников приглашается для участия в добровольном порядке в исследовании по методу компенсаторного обучения, в то время как эквивалентная группа таких приглашений не получает. Возможно, что 30% из числа приглашенных примут участие в эксперименте. Показатели предварительного и итогового тестирования получают с помощью стандартных тестов успешности чтения; тестированию подвергаются все классы. Неправильно было бы сравнивать 30% добровольцев со всей контрольной группой, так как эти добровольцы представляют собой тех, кто больше всего обеспокоен своими результатами предварительного тестирования, кто, вероятно, наиболее энергично работает над собой и т. д. Но в этом случае невозможно выделить соответствующих им лиц в контрольной группе. Сравнение результатов всех приглашенных с результатами неприглашенных — хотя это и кажется несправедливым с точки зрения гипотезы о терапевтической эффективности — на самом деле является приемлемым, хотя и осторожным решением проблемы. Отметим, однако, что, возможно, само по себе приглашение, а не терапия вызывает данный эффект. В общем, неприглашенная контрольная группа должна быть столь же осведомлена о результатах предварительного тестирования, как и приглашенная. Другое решение состоит в том, чтобы приглашать всех, кто нуждается в компенсаторном обучении, и составлять из желающих две группы — настоящую экспериментальную и группу-плацебо<sup>1</sup>; но на данном уровне развития любая плацебо-терапия, которая выглядит достаточно правдоподобной, чтобы ее приняли за реальную помощь, вероятно, будет столь же хорошей терапией, что

---

<sup>1</sup> Группа-плацебо (*placebo* — *лат.*) — группа, которая якобы подвергалась воздействию, но воздействие в действительности не осуществляется. (См. подробнее статью «Научный вывод, артефакты и контроль» в данной книге). — *Прим. ред.*

и исследуемый метод. Отметим, однако, обоснованность заключения о том, что гораздо легче провести экспериментальные проверки сравнительной эффективности двух терапевтических процедур, чем оценить абсолютную эффективность каждой из них. Единственным практическим решением является создание экспериментальной и контрольной групп из желающих пройти курс компенсаторного обучения путем манипулирования временем ожидания (см., к примеру, Rogers, Dymond [93]). Иногда это создает другие трудности, например чрезмерное выбывание из контрольной группы, для которой терапия отложена. Об успешном и, по-видимому, не вызывающем особых реакций испытуемых использовании жеребьевки для решения о немедленном или отсроченном проведении курса смотри в работе Рида [92].

## Факторы, ставящие под угрозу внешнюю валидность

Описанные выше факторы внутренней невалидности были факторами, непосредственно влияющими на данные  $O$ . И эти факторы сами по себе могли вызвать изменения, которые можно ошибочно принять за результат действия  $X$ . Если в план эксперимента включается еще и контрольная группа, эти факторы вызывают эффекты, проявляющиеся и в контрольной группе: в экспериментальной же группе их действие накладывается на действие  $X$ . В терминах дисперсионного анализа *фон («история»)*, *естественное развитие*, *эффект тестирования* и т. д. могут быть представлены как главные эффекты и как таковые они могут контролироваться в эксперименте по типу плана 4, обеспечивая ему *внутреннюю валидность*. С другой стороны, угрозы *внешней валидности* могут быть отнесены к эффектам взаимодействия  $X$  с некоторыми другими переменными. Тем самым они представляют собой потенциальные ограничения эффектов  $X$  некоторой нежелательно узкой совокупностью условий.

Забегая вперед, отметим, что в плане 4, как известно, наблюдаемые эффекты действия  $X$  могут проявляться только в группах, «подготовленных» предварительным тестированием. Логически мы не имеем права распространять эти результаты на ту большую, пред-

71

варительно не обследованную популяцию, о которой мы хотели бы говорить.

В этом разделе мы обсудим несколько таких угроз обобщаемости и способы их уменьшения. Так, поскольку существуют валидные планы, в которых не используется предварительное тестирование, и поскольку во многих ситуациях (не обязательно в исследованиях обучения) хотят распространить результаты на группы, не подвергавшиеся предварительному тестированию, такие планы более предпочтительны, так как обеспечивают большую *внешнюю валидность* или возможность обобщения. В области обучения, часто высказываемые сомнения по поводу возможности применения результатов крайне искусственных экспериментов к реальной практике являются суждениями о *внешней валидности*. Поэтому учет таких соображений при обсуждении оптимальных экспериментальных планов встретит одобрение тех практических работников, которые правильно понимают, что такие соображения бесосновательно игнорируются в обычных формальных курсах по экспериментальной методологии. В дальнейшем мы покажем несколько путей повышения внешней валидности экспериментов, создания более подходящих условий для обобщения их результатов и применения их в практике обучения без утраты *внутренней валидности*.

Прежде чем перейти к обсуждению этих вопросов, необходимо сделать несколько предостережений. Они касаются некоторых острых в учении об индукции проблем. Они являются таковыми вследствие периодически повторяющегося вынужденного признания трюизма Юма о том, что *индукция или обобщение никогда не могут быть полностью логически обоснованы*. В то время как проблемы *внутренней валидности* разрешимы в рамках логики вероятностной статистики, проблемы *внешней валидности* логически не разрешимы никаким ясным, убедительным способом. Оказывается, что обобщение всегда связано с экстраполяцией в области, не представленные в некоторой выборке. Такая экстраполяция осуществляется посредством *предположения* о знании релевантных закономерностей. Так, если имеется внутренне валидный план 4, то эффект продемонстрирован только для тех конкретных условий, которые оказались общими для экспериментальной и контроль-

72

ной групп, то есть только для предварительно обследованных групп, включающих лиц определенного возраста, интеллекта, социально-экономического статуса, географического региона, взятых в определенный исторический момент, при данной ориентации светил, определенном атмосферном давлении, уровне гамма-излучения и т. д.

*Логически* мы не имеем права выходить за эти пределы, то есть, не имеем права вообще обобщать. Но в действительности мы пытаемся делать обобщения, устанавливая закономерности и отбрасывая некоторые из этих обобщений в других столь же конкретных, но отличных условиях. Изучая историю науки, мы узнаем о «подтверждении» обобщения посредством накопления опыта обобщения, но это не логическое обобщение, выводимое из данных исходного эксперимента. Столкнувшись с этим, мы при обобщении высказываем предположения о пока еще не доказанных закономерностях, включая и такие, которые совсем не исследовались. Так, при изучении преподавания мы полностью принимаем предположение о том, что ориентация в магнитном поле не оказывает влияния на результаты. Но на основе отдельных исследований мы знаем, что предварительное тестирование часто влияет на результаты исследования, и поэтому его как ограничение для обобщения следует устранить. Если же мы изучаем железные бруски, то по опыту знаем, что процедура предварительного взвешивания не оказывает никакого влияния на результаты последующих замеров, но ориентация в магнитном поле, если она систематически не контролируется, может в значительной степени ограничить обобщение результатов наших исследований. Таким образом, источники внешней невалидности являются предположениями об общих законах науки о науке,— предположениями о том, какие факторы закономерно взаимодействуют с нашими независимыми переменными, и предположениями о том, чем можно пренебречь.

Помимо этого, существует общая эмпирическая закономерность, которую мы признаем наряду со всеми учеными. Речь идет о современной трактовке постулата Милля о законообразности природы. В своей современной, более мягкой трактовке он выступает как утверж-

дение о «склеенности» («stickiness») природы: мы предполагаем, что, чем ближе два события в пространстве и во времени, а также по измеренным значениям по некоторым или по всем параметрам, тем в большей степени они будут следовать одним и тем же законам. Чем больше экспериментальные условия отличаются от тех, на которые хотелось бы распространить результаты, тем в большей степени следует ожидать сложных взаимодействий и криволинейных зависимостей, затрудняющих обобщение. Поэтому наш призыв к достижению большей внешней валидности есть призыв к достижению максимального сходства экспериментальных условий с условиями применения при сохранении внутренней валидности.

Подчеркивая это, мы должны иметь в виду, что такие «успешные» науки, как физика и химия, развиваются, не обращая никакого внимания на репрезентативность (зато придавая большое значение воспроизводимости результатов независимыми исследователями). Искусственная лабораторная наука, создаваемая в башнях из слоновой кости, имеет ценность, даже если она и нерепрезентативна, а искусственность часто бывает необходимой при аналитическом вычленении переменных, что в свою очередь является очень важным для достижений во многих науках. Но, несомненно, и то, что весьма важным, особенно для таких прикладных дисциплин, как педагогика, является достижение внешней валидности, когда это не препятствует внутренней валидности или аналитической работе.

*Взаимодействие тестирования и X.* Если ограничиваться только работами по планированию экспериментов как таковых, то приоритет в выделении угрозы внешней валидности со стороны предварительного тестирования принадлежит Соломону [104], хотя

еще раньше аналогичные соображения привели отдельных экспериментаторов к использованию плана б, в котором нет предварительного тестирования. Особенно отчетливо эта проблема предстает в исследованиях по изменению установок, где тесты для измерения последних содержат много весьма необычного материала (например, редко можно встретить в печатном тексте столько враждебных заявлений, сколько их содержится в типичном тесте на национальные предрассудки), и вполне веро-

74

ятно, что предварительное тестирование оказывает влияние на установки личности и ее подверженность убеждению. Психолог будет серьезно сомневаться в возможности сравнивать киноаудиторию, видевшую фильм «Джентльменское соглашение», направленный против предрассудков, сразу после ответов на 100 пунктов теста на антисемитизм, с другой аудиторией, видевшей фильм без предварительного тестирования. Эти сомнения касаются не только основного влияния предварительного тестирования, но также и реакции на убеждающее воздействие. Предположим, что замысел фильма реализован настолько приглушенно, что некоторые зрители могут с интересом посмотреть его, уделив внимание какой-то одной линии, не осознав тех социальных проблем, которых он касается. Такие люди, возможно, не войдут в группу, подвергавшуюся предварительному тестированию. Если сделать аудиторию сензитивной посредством предварительного тестирования, то может возрасти обучающее влияние  $X$ , так как внимание аудитории будет определенным образом сфокусировано. Понятно, что такое воздействие  $X$  может быть эффективно лишь для предварительно тестируемой группы.

Хотя сензитивные влияния такого рода часто приводятся в качестве иллюстрации эффекта, данные нескольких публикаций указывают либо на отсутствие такого эффекта (см., например, Anderson [3], Duncan et al. [31], Glock [42], Lana [64, 65], Lana, King [66], Piers [89], Sobol [103], Zeisel [139]), либо на озадачивающий эффект взаимодействия. Так, Соломон [104] установил, что проведение предварительного тестирования снижало эффективность экспериментального обучения чтению, а Ховленд, Ламздейн и Шеффилд [51] предположили, что предварительное тестирование уменьшает убеждающее воздействие фильмов. Таких эффектов взаимодействия следует избегать, даже если они не столь обманчивы, как сенсбилизация (так как благодаря избытку опубликованных открытий ложные подтверждения являются более серьезной проблемой в нашей литературе, чем ложные опровержения. См. Campbell [11, с. 168—170]).

Снижение внешней валидности, обусловленное влиянием предварительного тестирования на  $X$ , конечно, за-

75

висит от того, в какой степени повторяющиеся обследования характерны для той популяции, на которую хотят распространить выводы. Исследовательские интервью и процедуры измерения установок совсем нетипичны для сферы массовой коммуникации. Но когда мы хотим обобщить данные исследований в области педагогики, мы сталкиваемся с условиями, в которых тестирование — обычное явление, особенно если экспериментатор в качестве тестирования использует обычные опросы в классе. Но, возможно, нежелательное взаимодействие тестирования с  $X$  не возникает и тогда, когда экспериментальное тестирование  $O$  будет сходно с тем, которое всегда применяется в этой области. В тех же случаях, когда тестирование осуществляется с помощью необычных процедур или связано с маскировкой цели исследования, перестройкой восприятия и мышления, удивлением, стрессом и т. д., желательно, а возможно, и необходимо применять планы экспериментов с предварительно не тестируемыми группами.

*Взаимодействие отбора (состава группы) и X.* Так как в плане 4 контролируется влияние отбора только при интерпретации различий между экспериментальной и контрольной группами, остается опасность того, что полученные результаты валидны лишь для той популяции, из которой были составлены экспериментальная и контрольная группы. Такая возможность тем вероятней, чем больше трудностей мы встречаем при поиске испытуемых для эксперимента. Рассмотрим ситуацию, когда от экспериментирования с преподаванием отказались девять школ и только десятая согласилась участвовать в эксперименте. Почти наверняка эта десятая школа отличается по многим показателям от девяти других и от основной массы школ, на которую мы хотим распространить полученные данные. Эта школа, таким образом, нерепрезентативна. Скорее всего, ее штат обладает более высокими моральными качествами, в ней меньше опасаются проверок, больше стремятся к совершенствованию, чем в обычных «средних» школах. И хотя полученные результаты и будут обладать внутренней валидностью, они справедливы только для школ такого рода. Чтобы выносить суждения по этим вопросам, в отчет об исследовании следует включать данные о том, сколько и каких школ и классов были

76

приглашены сотрудничать, и отказались от проведения эксперимента; в таком случае читатель сможет оценить степень возможных смещений выборки. Вообще говоря, чем больше было попыток наладить сотрудничество, чем в большей мере нарушается сложившаяся практика и чем выше процент отказов от экспериментирования, тем больше вероятность наличия эффекта специфичности отбора.

Рассмотрим более подробно, что означает «*взаимодействие фактора отбора (состава группы) и X*». Если мы намерены проводить исследование в рамках одной школы, которая выразила готовность сотрудничать, то при случайном распределении испытуемых в экспериментальную и контрольную группы мы можем не заботиться об «основном эффекте» школы как таковой. Если средние результаты экспериментальной и контрольной групп лишь в равной мере улучшены посредством такого исследования, то этому не стоит придавать значения. Если же окажется, что особенности этой школы способствуют большей (по сравнению с той совокупностью, на которую собираются распространить данные) эффективности экспериментального курса, то это обстоятельство может быть серьезным. Надо убедиться в незначительности взаимодействия характеристик школы (связанных, вероятно, с готовностью сотрудничать) с экспериментальным воздействием. Некоторые экспериментальные переменные могут быть довольно чувствительны к характеристикам школы (взаимодействовать с ними), другие — нечувствительны. Такие взаимодействия *могут* возникать при использовании в исследовании школ со сходными средними *IQ* и могут отсутствовать, если значения *IQ* сильно отличаются друг от друга. Однако мы должны ожидать, что взаимодействия будут более вероятны тогда, когда школы сильно отличаются друг от друга по различным признакам, а не тогда, когда они сходны.

Часто значительные смещения выборки возникают из-за инерции экспериментаторов, которые не предусмотрели того, что более репрезентативная выборка школ может отказаться от участия в эксперименте. Вследствие этого большинство педагогических исследований осуществляется в школах, где обучается непропорционально большой процент детей университетских препо-

77

давателей. Хотя достижение выборочной репрезентативности в полной мере невозможно и хотя такой подход почти полностью отрицается во многих науках (например, в большинстве публикаций на страницах «*Journal of Experimental Psychology*»),

репрезентативность может и должна считаться желательной в педагогических исследованиях. Одним из путей повышения репрезентативности является уменьшение числа испытуемых или классов из одной школы или класса и увеличение числа школ и классов, в которых проводятся эксперименты. Конечно, мы никогда не достигнем того, чтобы эксперименты проводились на выборке, репрезентативной для всех школьных классов в США или всего мира. Мы можем лишь шаг за шагом, путем проб и ошибок узнавать, сколь далеко мы можем экстраполировать, как широко обобщать внутренние валидные результаты. Но такая работа будет более успешной, если уже в исходном эксперименте мы будем наблюдать явление в широком диапазоне условий.

Если обратиться к «плюсам» и «минусам» табл. 1, то, очевидно, нельзя сказать ничего определенного о соответствующей колонке этого плана. Однако эта колонка включена в таблицу, так как требования, связанные с некоторыми планами эксперимента, заостряют или смягчают основную проблему. При использовании плана 4 в исследованиях социальных установок требуется настолько большая готовность к сотрудничеству со стороны части испытуемых или респондентов, что это вынуждает ограничить проведение исследований только подконтрольной аудиторией, отличающейся от обычных граждан, о которых хотели бы говорить на основании данных исследования. В такого рода исследованиях план 4 следует расценивать как план, имеющий «минус» по фактору отбора. Однако в педагогических исследованиях объектом внимания является подконтрольная популяция, и поэтому при использовании плана 4 могут быть получены высокорепрезентативные результаты.

*Другие взаимодействия с X.* Аналогичным образом взаимодействия  $X$  с другими факторами могут рассматриваться как угрозы *внешней* валидности. Дифференцированное *выбывание* является, видимо, следствием  $X$ , а не фактором, взаимодействующим с ним. Взаимодей-

ствие фактора *инструментальной погрешности* с  $X$  уже рассматривалось при обсуждении внутренней валидности, так как специфическое влияние этого фактора при предъявлении  $X$  бывает имитацией истинного влияния  $X$  (например, когда наблюдатель, который дает оценку, знает гипотезу и знает, кто из учащихся подвергался воздействию  $X$ ). Угрозой внешней валидности является вероятность эффектов, характерных для конкретных средств (тестов, наблюдателей, счетчиков), используемых в этом исследовании. Если в исследовании используется несколько наблюдателей или интервьюеров, то такие взаимодействия могут быть изучены непосредственно (Stanley [111]). *Регрессия* не выступает как фактор, взаимодействующий с  $X$ .

Действие фактора *естественного развития* связано с проблемой отбора испытуемых: результаты могут оказаться действительными только для данного возраста, данной степени утомления и т. д. Взаимодействие  $X$  с *фоном* означало бы, что эффект наблюдается только в данных условиях эксперимента, и, хотя для этого фона результаты валидны, они не будут получены в других условиях. Так, если эксперимент проводился во время войны или, скажем, сразу же после неудачной забастовки учителей, то он мог бы выявить повышенную реакцию на  $X$ , которая не обнаружилась бы в других обстоятельствах. Следуя принципам выборочного исследования, мы должны были бы воспроизвести наш эксперимент на случайной выборке прошлых и будущих ситуаций, что, конечно, невозможно. Кроме того, в нашей области, как и в других науках, мы руководствуемся эмпирическим допущением, согласно которому не существует действительно зависящих от времени законов, и обнаруживаемые эффекты *фона*, обусловленные конкретными сочетаниями условий стимуляции в данный момент, в конечном счете, должны быть учтены в едином, не зависящем от времени законе (Neuman [85]). (Применительно к космологическим теориям «расширяющейся Вселенной» это утверждение необходимо снабдить оговорками,

но не в контексте данного обсуждения.) Тем не менее, успешное воспроизведение результатов в разное время и в различных условиях повышает нашу уверенность в проводимом обобщении, уменьшая вероятность взаимодействия *фона* и *X*.

79

Эти несколько факторов не включены в табл. 1, ибо они не дают оснований для различения альтернативных типов экспериментов.

*Реакция испытуемых на эксперимент.* Весьма серьезным источником нерепрезентативности обычного психологического эксперимента и даже педагогического исследования является пресловутая искусственность экспериментальной обстановки и знание испытуемого о том, что он участвует в эксперименте. Для испытуемого создается проблемная ситуация повышенной сложности, а процедуры и экспериментальное воздействие на которые он реагирует, имеют не только значение стимулов, но играют роль ключевых признаков, используемых испытуемым с целью разгадать намерения экспериментатора. В такой обстановке для испытуемого характерно актерство, настороженность, установка на угадывание, он ощущает себя подопытным кроликом и т. д. Все это нетипично для нормальных школьных условий и, примешиваясь к эффекту *X*, сильно затрудняет обобщение результатов. Хотя подобные *реакции* неизбежны, внутренне валидные эксперименты этого типа все равно следует проводить. Но нужно предупреждать по возможности реакцию испытуемых на эксперимент. Тут мы должны отчасти согласиться с критикой экспериментов некоторыми административными работниками в сфере образования и педагогов, говорящих о тщетности «всех этих исследований». Но наша позиция более умеренна: мы призываем не прекратить, а усовершенствовать исследования. Ниже приводятся наши конкретные предложения.

*Реакцию на эксперимент* может вызвать любой аспект экспериментальной процедуры. Это относится и к процедуре предварительного тестирования (независимо от его содержания), и к его взаимодействию с *X*. Возможно, оказывает влияние и содержательная сторона процесса тестирования. Процедура рандомизации и распределения испытуемых по режимам эксперимента также может вызывать такую реакцию. Представьте себе, как подействует на класс то, что случайно отобранную половину учеников вдруг отправляют в отдельное помещение (Solomon [104]). Это, да еще появление незнакомых «учителей», конечно, должны вызвать у детей ожидание чего-то необычайного, активное желание

80

узнать, что же затевается. Аналогичную реакцию может вызвать введение какого-либо экстраординарного *X*. Предположительно в эксперименте по плану 6 с одним лишь итоговым тестированием подобный эффект может возникнуть при итоговом тестировании, причем вероятность эффекта тем больше, чем очевиднее связь между экспериментальным воздействием и характером тестирования.

Изучая перемены в общественном мнении, очень трудно избежать таких реакций. Но при изучении методов преподавания ученики могут и не подозревать, что идет какой-то эксперимент. (Хорошо было бы, если бы об этом ничего не знали также и учителя, как в случае с двойным слепым экспериментом в медицине, но обычно это неосуществимо.) Эксперимент можно замаскировать следующими приемами. Если *X-ы* — это варианты обычных мероприятий в классе, проводимых в соответствующее время и предусмотренных школьной программой, то уже многого можно достигнуть, не оповещая о введении экспериментальных воздействий. Далее, *O* можно проводить под видом регулярных контрольных работ, и тогда второе требование будет соблюдено. Если *X* предусматривает индивидуальное общение с испытуемым, то для рандомизированного разбиения класса на эквивалентные подгруппы вовсе не обязательно разводить их по разным комнатам и т. д.

Незаметное осуществление рандомизации является наиболее часто возникающей проблемой. Иногда можно воспользоваться тем, что в колледжах и больших средних школах учащиеся записываются на факультативные курсы, читаемые в разных помещениях в одни и те же часы. Контроль за процессом записи позволяет составлять рандомизированные эквивалентные группы. (Siegel, Siegel [100] показали возможность нетрадиционного использования процесса естественной рандомизации.) Правда, ввиду того, что в каждой группе складываются свои особые условия, группы со временем становятся все менее эквивалентными.

*Решением проблемы во всех отношениях является использование в качестве единиц рандомизации целых классов*, так что экспериментальная и контрольная группы в результате состоят из случайным образом отобранных классов (см. Lindquist [67, 68]). Обычно — но это

81

не столь существенно — вначале подбираются классы на основе таких признаков, как школа, учитель (когда один учитель ведет несколько классов), предмет, время дня, средний уровень умственного развития и т. д., а затем эти классы случайным образом распределяются по различным режимам эксперимента. Подобных исследований пока немного, но скоро они должны стать нормой. При проверке значимости эффекта в этом случае *нельзя* объединять данные по всем учащимся, поскольку не они подлежали рандомизации. Подробно это будет рассмотрено в следующем параграфе.

Наблюдая работу ряда экспериментаторов, которые настолько не ориентированы в жизни школы, что полученные ими результаты прямо-таки вводят в заблуждение, мы постепенно пришли к убеждению, что экспериментальные исследования в школе должны по возможности проводиться силами школьного персонала, особенно когда предполагается распространить полученные результаты на другие аналогичные ситуации.

В настоящее время в школе, видимо, практикуются два вида «экспериментов»: 1) исследования, «навязанные» школе людьми извне, которые преследуют свои собственные цели и не намерены сразу же предлагать школе какой-либо определенной программы действий, и 2) исследования, организованные «инициативными людьми», старающимися сделать так, чтобы сами учителя выступали в роли «экспериментаторов» в широком смысле этого слова. Результаты первого типа исследований могут быть строгими, но не находят применения. Во втором случае результаты непосредственно связаны с практикой, но, вероятно, не «верны», поскольку они получены в условиях полного отсутствия экспериментальной строгости. Альтернативный подход состоял бы в следующем: идеи школьного эксперимента должны исходить от учителей и других работников школы, затем схема исследования, направленного на верификацию этих идей, вырабатывается в ходе их совместной работы со специалистами в области методологии эксперимента, а основная экспериментальная работа проводится самими авторами идей. Соответствующий статистический анализ данных осуществляется специалистом-методологом, который возвращает результаты исследовательской группе через компетентного посредника

82

(учителя-методиста, руководителя исследований в системе школ и т. д.), постоянно исполняющего эту функцию. Такие результаты будут и нужными, и «правильными». Проблема в том, как направить *основные* исследования по такому руслу. При наличии компетентного посредника они отошли бы тогда от решения мелких злободневных проблем и приобрели бы большую теоретическую направленность.

Хотя обзор хороших и плохих примеров из литературы не входит в наши намерения, недавнее исследование Пейджа [87] настолько удачно выполнено с точки зрения

предотвращения реакции испытуемых на эксперимент, достижения репрезентативности выборок и исключения взаимодействия между тестированием и  $X$ , что оно заслуживает упоминания в качестве конкретной иллюстрации оптимальной практики. В этом исследовании показано, что если учитель, проверяя контрольные работы, пишет на полях краткие замечания, то его ученики лучше справляются с такими заданиями в дальнейшем. Этот результат был продемонстрирован на 74 учителях в 12 школьных системах по 6 старшим классам для 5 уровней выполнения ( $A, B, C, D, F$ ) задания и на самых различных испытуемых. Учителя и классы отбирались в случайном порядке. В качестве предварительных данных использовались результаты самой первой из регулярных контрольных работ, проводившихся в каждом классе. Решение о том, писать свои замечания или нет, учитель принимал в зависимости от результатов бросания особым образом размеченной игровой кости. Следующая по расписанию контрольная работа выполняла роль итогового тестирования. Насколько можно было заключить из дальнейших бесед, ни один из 2139 учеников не знал о проведении эксперимента. Лишь немногие школьные процедуры поддаются такой не вызывающей подозрения рандомизации, так как в классе обычно применяется устная речь, которая адресуется сразу всему классу, а не отдельным ученикам. (Письменное общение обеспечивает возможность рандомизации учеников по различным экспериментальным воздействиям, хотя остается проблемой, что ученики могут обнаружить варьирование воздействия.) Тем не менее, имея в виду сказанное выше, исследователи смогут предупреждать нежелательные

реакции испытуемых на эксперимент в большей мере, чем в настоящее время.

#### Проверка значимости эффекта в плане 4

Хороший экспериментальный план независим от способа статистической обработки результатов. Планирование эксперимента представляет собой искусство получения поддающихся интерпретации сравнений и как таковое необходимо и тогда, когда конечные результаты выражаются в виде процентных диаграмм, параллельного описания случаев, фотографий испытуемых, занятых выполнением задания, и т. д. Возможность истолкования результатов во всех этих случаях зависит от контроля рассмотренных нами факторов валидности. Если сравнение поддается интерпретации, мы обращаемся к статистической проверке значимости, позволяющей установить, превышает ли наблюдавшееся различие уровень случайных флюктуаций, которых можно ожидать для выборок данного объема при отсутствии подлинного различия. Применение этой статистической процедуры предполагает, но не доказывает и не гарантирует сопоставимости сравниваемых групп или возможности истолковать обнаруженное различие.

Нам хотелось бы научить строить план эксперимента на основе здравого смысла и соображений нематематического характера. Надеемся, что большая часть этой книги доступна исследователям, не прошедшим курса подготовки в области математической статистики. Но поскольку вопрос статистической обработки данных тесно связан с планированием эксперимента, мы выскажем некоторые соображения по этому поводу. (См. также Green, Tukey [45], Kaiser [56], Nunally [86], Rozeboom [96].)<sup>1</sup>

*Распространенные ошибки статистической обработки.* Хотя план 4 является стандартным и широко распространенным, даже в этом случае часто применяются неправильные, неполные или не соответствующие задаче статистические процедуры. Применяя обычный критерий

---

<sup>1</sup> См. также работы по планированию эксперимента и процедурам проверки статистической значимости: Х и к с Ч. Основные принципы планирования эксперимента. М., «Мир», 1967; Г л а с с Дж., С т э н л и Дж. Статистические методы в педагогике и психологии. М., «Прогресс», 1976. — *Прим. ред.*

Стьюдента (*t*-критерий), многие подсчитывают два значения *t*: одно — для различия в показателях предварительного и итогового тестирования по экспериментальной группе, а другое — для аналогичных результатов по контрольной группе. Если первое значение статистически значимо, а второе нет, то делается вывод о наличии эффекта *X* без прямого статистического сравнения экспериментальной и контрольной групп. Во многих таких случаях применение более адекватного критерия не обнаружило бы значимого различия между группами (например, когда значение *t*-критерия чуть выше критического в экспериментальной и почти достигает его в контрольной группе). Уиндл [135] и Кантор [18] показали, насколько распространена эта ошибка.

*Использование приращений и ковариационный анализ.* Наиболее широко используемая и приемлемая процедура состоит в определении для каждой группы приращения показателей от предварительного тестирования к конечному и вычислении *t*-критерия для приращений, наблюдавшихся в экспериментальной и контрольной группах. В большинстве случаев, однако, следует предпочесть ковариационный анализ, в котором показатели предварительного тестирования берутся в качестве сопутствующих переменных и которому предшествует процедура формирования рандомизированных блоков или «распределение по уровням» показателей предварительного тестирования. Использование этого более точного анализа представляется весьма желательным. По сравнению с трудом, затраченным на проведение эксперимента, усилия, связанные с надлежащим анализом, просто ничтожны. Детали обработки можно найти в стандартных примерах анализа данных по Фишеру. (См. также Сох [27, 28], Feldt [35], Lindquist [68].)

*Статистики для рандомизированного распределения целых групп по режимам эксперимента.* Обычные статистики применимы только тогда, когда назначение определенного режима производится в случайном порядке для каждого ученика в отдельности. Когда же режим определяется сразу для целого класса, ошибка, вычисленная согласно этим моделям, будет занижена, так как при рандомизации использовались более крупные единицы и выборка наблюдений содержала меньшее число случайных событий. Линдквист [68, с. 172—189]

предложил обоснование и формулы анализа, адекватного данной ситуации. В качестве исходных данных здесь используются средние значения класса, и эффект *X* должен быть выделен из колебаний именно этих величин. Они же должны фигурировать в роли сопутствующих переменных при проведении ковариационного анализа.

*Статистики достижения внутренней валидности.* Первые три пункта этого параграфа выражают ортодоксальную для статистиков установку в отношении планирования эксперимента. В данном разделе делается попытка расширить или скорректировать эту установку, при этом различие *внешней* и *внутренней валидности* распространяется на сферу статистики выборочного исследования. Все обсуждавшиеся выше статистики предполагают выборки из бесконечно большой популяции, что больше соответствует изучению общественного мнения, чем обычному лабораторному эксперименту. Лишь в редких случаях (как, например, в работе Пейджа [87]) выборки действительно берутся из большой, заранее определенной популяции, что делает пригодными обычные формулы. Диаметрально противоположны этому лабораторные эксперименты, описываемые, скажем, в «Journal of Experimental Psychology», в которых принимается во внимание только

*внутренняя валидность* и в которых *все* члены уникальной малой популяции входят в экспериментальные и контрольные группы. В таких экспериментах большое значение придается рандомизации, но не в целях обеспечения репрезентативности в отношении более широкой популяции. Вместо этого рандомизация служит единственной цели уравнивания экспериментальной и контрольной групп. Таким образом, рандомизация проводится в пределах весьма малой конечной популяции, которая, собственно говоря, есть сумма экспериментальных и контрольных групп.

Эта крайняя позиция в вопросе формирования выборки оправдана применительно к лабораторным процедурам такого типа: приглашаются добровольцы за вознаграждение — за деньги, за оценки, за зачетные очки по данному курсу — или без него, в порядке выполнения обязательных требований, которые рано или поздно все равно пришлось бы выполнить. Эти добровольцы в случайном порядке распределяются по группам, которые по-разному участвуют в эксперименте.

86

Как только обследовано определенное число испытуемых, эксперимент прекращается. Отсутствует даже случайный отбор из гораздо более длинного списка добровольцев. Первые добровольцы образуют смещенную выборку, а вся популяция, из которой комплектуется выборка, меняется изо дня в день, и требуется все больше усилий, чтобы набрать новых добровольцев, и т. д. В какой-то момент, когда все члены популяции уже использованы в том или ином режиме, процедуру прекращают. Отметим, что смещенность выборок здесь никак не влияет на случайную эквивалентность групп, а сказывается лишь на их репрезентативности.

В целом очевидно, что главная цель рандомизации в лабораторном эксперименте — это внутренняя, а не внешняя валидность. В связи с этим следует пользоваться меньшими значениями ошибки, отвечающими меньшим популяциям. Вслед за Кемпторном [59], а также Уилком и Кемпторном [132] мы считаем, что данной задаче соответствует не извлечение случайных выборок из какой-то бесконечной популяции, а модель урновой рандомизации. В данном случае возможна более подходящая, более точная непараметрическая модель, согласно которой берутся полученные показатели экспериментальной и контрольной групп и многократно распределяются в случайном порядке по двум «урнам». При этом эмпирически (или математически) генерируется распределение средних различий, возникающее целиком за счет случайного расклада этих конкретных показателей. Это распределение является критерием, с которым следует сравнивать фактически полученную разность групповых средних. Когда имеет место взаимодействие между испытуемыми и режимами эксперимента, это распределение характеризуется меньшей дисперсией, чем соответствующее распределение Стьюдента.

Мы не ожидаем, что приведенные выше замечания коренным образом изменят повседневную практику применения критерия значимости в педагогических исследованиях. Точные решения очень трудоемки и обычно недостижимы. Урновая рандомизация, например, требует наличия быстродействующих ЭВМ. Но направление ошибки известно: традиционные статистические методы слишком консервативны, слишком склоняют нас к ответу: «Никакого эффекта нет». Раз мы находим, что

87

наши публикации перегружены «ошибками первого рода», то есть констатациями эффектов, не подтверждаемых перекрестной валидизацией (это бесспорно в отношении экспериментальной и социальной психологии, и даже исследования обучения), то подобная ошибка вполне оправдана. Наибольшая недооценка значимости может иметь место, когда

сравниваются всего два вида экспериментальных условий, причем используются все имеющиеся в распоряжении испытуемые (Wilk, Kempthorne [132, с. 1154]).

### 5. План Соломона для четырех групп

Хотя план 4 используется чаще, план 5, предложенный Соломоном [104], заслуженно обладает большим признанием и является первой попыткой эксплицитно учитывать факторы *внешней валидности*. Этот план имеет следующий вид:

$R$	$O_1$	$X$	$O_2$
$R$	$O_3$		$O_4$
$R$		$X$	$O_5$
$R$			$O_6$

Путем параллельного использования элементов плана 4 ( $O_1 — O_4$ ), когда предварительное тестирование не проводится ни в экспериментальной, ни в контрольной группах, может быть определен не только главный эффект *тестирования*, но и его взаимодействие с  $X$ . Тем самым расширяются возможности обобщения результатов и, кроме того, эффект  $X$  воспроизводится четырьмя различными способами:  $O_2 > O_1$ ;  $O_2 > O_4$ ;  $O_5 > O_6$  и  $O_5 > O_3$ . Ввиду нестабильности экспериментальных условий, если все эти четыре соотношения выполняются, то правомерность вывода значительно возрастает. Обобщению экспериментальных результатов косвенно способствует и то, что применение плана 5 в какой-либо определенной области позволяет выяснить общую вероятность взаимодействия тестирования и  $X$ , благодаря чему облегчается интерпретация данных, полученных в прошлых и будущих экспериментах по плану 4. Точно так же сравнение  $O_6$  с  $O_1$  и  $O_3$  позволяет выявить комбинированный эффект естественного развития и фона.

88

#### Статистические модели для плана 5

Единой статистической процедуры, в которой бы одновременно использовались результаты всех шести наблюдений, не существует. Асимметрия этого плана исключает дисперсионный анализ приращений. (Предложения Соломона на этот счет неприемлемы.) Не обращая внимания на предварительные тестирования (отводя им роль лишь дополнительного параметра воздействия  $X$ ), можно обработать данные итогового тестирования согласно простой схеме  $2 \times 2$  дисперсионного анализа:

		$X$	
		Да	Нет
Предварительное тестирование	Да	$O_2$	$O_4$
	Нет	$O_5$	$O_6$

Средние по столбцам позволяют оценить главный эффект  $X$ , средние по строкам — главный эффект предварительного тестирования, а средние по ячейкам — взаимодействие этих двух факторов. Если главный эффект предварительного тестирования и эффект взаимодействия настолько малы, что ими можно пренебречь, желательно провести

ковариационный анализ  $O_4$  и  $O_2$ , используя результаты предварительного тестирования в качестве сопутствующей переменной.

### **6. План с применением контрольной группы и тестирования, осуществляемого только после воздействия**

Хотя идея предварительного тестирования глубоко проникла в сознание исследователей, работающих в области педагогики и психологии, оно не является неотъемлемой частью истинного эксперимента. По причинам психологического порядка трудно отказаться от того,

89

чтобы удостовериться в «равенстве» экспериментальной и контрольной групп до введения дифференцированного режима эксперимента. Тем не менее, наиболее адекватной во всех отношениях гарантией отсутствия изначального смещения между группами является рандомизация. В рамках доверительных пределов, устанавливаемых статистической моделью, рандомизации достаточно и без предварительного тестирования. Фактически почти все агробиологические эксперименты, выдержанные в традициях работ Фишера [37, 38], выполнены без предварительного тестирования. Проводя педагогические исследования, особенно в младших классах, часто приходится испытывать методы изложения ученикам совершенно нового учебного материала, и в этой обстановке предварительное тестирование в обычном смысле слова невозможно. Точно так же при исследовании влияния доводов адвокатов на решение присяжных неуместно определять априорные оценки виновности или невиновности подсудимого. Схема 6 годится для этих случаев, а также для тех, в которых можно было бы применить планы 4 или 5, позволяющие производить подлинную рандомизацию. Вот как выглядит план 6:

$$\begin{array}{r} R \quad X \quad O_1 \\ R \quad \quad O_2. \end{array}$$

Хотя этот план применялся еще в 20-х годах, он не рекомендуется большинством руководств по методам исследования в педагогике. Причина этого отчасти в неумении отличить его от плана 3, а отчасти в недоверии к рандомизации как средству уравнивания. Этот план представляет собой как бы половину плана Соломона, и, как нетрудно видеть, в нем контролируется как главный эффект тестирования, так и его взаимодействие с  $X$ , но в отличие от плана 5 они не измеряются. Однако такое измерение несущественно для решения центрального вопроса о наличии эффекта  $X$ . Таким образом, хотя план 5 предпочтительнее плана 6 по упомянутым причинам, выигрыш от плана 5 может не стоить большого труда, затрачиваемого на его реализацию. Аналогично план 6 обычно лучше плана 4, если только нет оснований сомневаться в подлинности рандомизации. Эксперимент типа 6 слишком редко применяется в педагогике и психологии.

90

Если, как это часто бывает при проведении педагогических исследований, доступны данные, полученные до введения  $X$ , ими, разумеется, следует воспользоваться формирования блоков, определения факторных уровней или как сопутствующими переменными. Этот совет основан на двух соображениях. Во-первых, статистические

модели, соответствующие плану 4, обладают большей мощностью, чем те, которые применимы для плана 6. Хотя значительные усилия, необходимые для реализации плана 4, в большинстве ситуаций делают подобный выигрыш неоправданным, положение меняется, когда нужные исходные показатели «лежат на поверхности». Во-вторых, доступность таких показателей позволяет проверить наличие взаимодействия  $X$  с начальным уровнем и тем самым составить более полное представление о возможности обобщения результатов. Правда, нечто подобное может быть выполнено и в рамках плана 6, если вместо предварительного тестирования воспользоваться другими имеющимися данными. Кроме того, для генеральной совокупности, на которую предполагается распространять выводы педагогического исследования, характерно частое тестирование. Последние два обстоятельства могут изменить ситуацию, делая план 6 предпочтительнее плана 4. Отметим также, что при заметном выбывании испытуемых между  $R$  и итоговым тестированием исходные данные в плане 4 обеспечивают больше возможностей для отклонения гипотезы о влиянии на эффект неравномерности выбывания из экспериментальной и контрольной групп.

Как бы то ни было, существует ряд проблем, применительно к которым предварительное тестирование невозможно, неуместно или легко провоцирует нежелательную реакцию испытуемых; для исследования этих проблем более приемлем план 6. Кроме упомянутого ранее изучения способов подачи нового учебного материала, можно привести много примеров, когда: 1)  $X$  и итоговое тестирование  $O$  могут быть предъявлены учащимся или группам естественным образом в единой композиции и 2) предварительное тестирование нельзя провести должным образом. Такие ситуации часто возникают при исследовании самой процедуры тестирования — например, при сравнении эффективности различ-

ных инструкций, бланков для регистрации ответов и т. д. Это относится и к сравнению эффективности обращений о наборе добровольцев и т. д. Когда необходимо сохранить анонимность респондентов, удобнее других оказывается план 6. В таких случаях рандомизация достигается путем перетасовки материалов, подлежащих распределению среди респондентов.

### Статистические модели для плана 6

Простейший способ обработки — это применение  $t$ -критерия. План 6, пожалуй, единственный, для которого он оптимален. Однако ковариационный анализ и формирование блоков по таким исходным показателям испытуемых (Underwood [129]), как успеваемость, оценки в тестах, профессия родителей и т. д., позволяют повысить мощность статистической процедуры примерно до того же уровня, что и при использовании предварительного тестирования. Идентичность предварительного и итогового тестов несущественна. Часто это различные формы «одного и того же» теста, и, следовательно, они менее идентичны, чем буквальное повторение предварительного теста. Точность возрастает с увеличением степени ковариации, и, хотя она обычно выше для эквивалентных форм «одного и того же» теста, чем для «различных» тестов, это чисто количественное различие, причем такой надежный комплексный показатель, как, скажем, средний балл успеваемости ученика, может оказаться лучше короткого предварительного тестирования. Отметим, однако, что использование среднего балла в качестве итогового показателя обычно нежелательно из-за его вероятной нечувствительности к  $X$  в сравнении с показателями, имеющими более тесную связь с содержанием и временем тестирования.

Неважно, как трактовать этот план с предварительным псевдотестированием — как план 6 или как план 4. Он обладает достоинствами плана 6, избавляя экспериментатора от проведения настоящего предварительного тестирования, от повторного применения

идентичного или весьма сходного по содержанию необычного материала (как при изучении изменений социальных установок). По этим причинам оценка плана 6 в отношении реакции на эксперимент будет несколько выше, чем для

планов 4 и 5. В социальных науках в целом это преимущество еще важнее, чем в педагогических исследованиях.

### Факторные планы

На концептуальной основе трех последних планов, особенно планов 4 и 6, добавляя к ним новые группы с новыми  $X$ , можно построить сложные факторные планы, типичные для фишеровской традиции. В обычном однофакторном дисперсионном анализе мы будем иметь несколько «уровней» воздействия ( $X_1, X_2, X_3$  и т. д., а также, возможно,  $X_0$  «без  $X$ »). Если отсутствие  $X$  считать одним из уровней («режимов») эксперимента  $X_0$ , то в планах 4 и 6 будет по одной группе на каждый уровень. По плану 5 мы будем иметь две группы (прошедшую и не прошедшую предварительное тестирование) для каждого режима, так что можно провести двухфакторный дисперсионный анализ. Нам неизвестно о применении более чем двухуровневых планов, аналогичных плану 5.

Обычно, если интерес представляет влияние предварительного тестирования, применяется план 6, так как иначе требуется большое число групп. Очень часто используются две и более независимых экспериментальных переменных, каждая из которых реализуется на нескольких «уровнях», для чего требуется ряд групп:  $X_{a1}X_{b1}$ ,  $X_{a1}X_{b2}$ ,  $X_{a1}X_{b3}$ ...  $X_{a2}X_{b1}$  и т. д.

Разработка соответствующих планов, сопровождаемая попытками экономить, устраняя из них некоторые из возможных перестановок из  $X_a$  и  $X_b$ , породила пугающие тайны факторного планирования (рандомизированные блоки, расщепленные делянки, греко-латинские квадраты, дробные реплики, смешивание и т. п.). В результате возникла пропасть между передовыми и традиционными методами педагогического исследования. Мы надеемся перебросить мост через эту пропасть, сохраняя непрерывную связь с традиционными методами и уповая на присущий каждому здравый смысл. По-видимому, большинство идей, необходимых для понимания планирования эксперимента, легче всего понять, если рассмотреть их на материале

планов с двумя экспериментальными режимами без дополнительных усложнений. Вместе с тем полное изложение проблем традиционного характера позволит достичь понимания необходимости и роли современных подходов. Уже в связи с поисками наиболее эффективного способа обработки результатов всеми признанного старого эксперимента по плану 4 мы познакомились с необходимостью ковариационного анализа, который пока почти не использовался в этих условиях. Точно так же при обсуждении плана 5, когда в связи с необходимостью контроля за некоторыми факторами мы столкнулись с задачей двух режимов, мы отказались от  $t$ -критерия в пользу соответствующей модели дисперсионного анализа.

Здесь мы не можем вдаваться в детали статистического анализа факторных планов или хотя бы привести их примеры. Элементарные аспекты этих методов с ориентацией на исследователей в области педагогики изложены Эдвардсом [33], Фергюсоном [36], Джонсоном и Джексонсом [54] и Линдквистом [68]. Надо надеяться, однако, что следующие несколько разделов позволят читателю составить себе известное представление о некоторых возможностях и трудностях, связанных с обсуждаемыми в этой главе вопросами

планирования эксперимента. Мы не станем рассматривать обычные соображения по поводу использования латинских квадратов и многих других неполных планов, где ради экономии в жертву приносится информация о факторных взаимодействиях. (Однако применение латинских квадратов взамен контрольных групп, когда невозможна рандомизация, будет рассмотрено ниже — см. квазиэкспериментальный план 11.) Причина, по которой мы решили не обсуждать неполные планы, состоит в том, что подробная информация о взаимодействиях имеет прямое отношение к проблеме внешней валидности, особенно в науке, где возникают трудности с воспроизведением результатов в разных условиях (см. Wilk, Kempthorne [132]). Мы хотели бы рассмотреть в этом разделе следующие понятия: взаимодействие, нестинг и перекрестная классификация, а также конечные, фиксированные, рандомизированные и смешанные факторные модели.

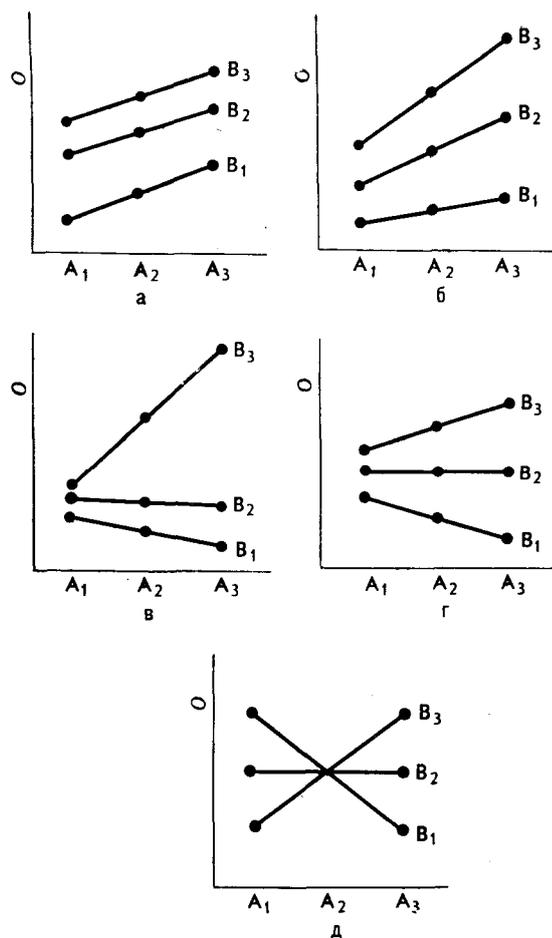
### Взаимодействие

Выше мы уже пользовались этим понятием, надеясь на то, что неподготовленный читатель поймет его значение из контекста. Как и ранее, мы делаем упор на то, как взаимодействие может отразиться на возможности обобщения результатов эксперимента.

Посмотрим, как выглядят на графике (рис. 2) пять возможных исходов эксперимента с тремя уровнями воздействий  $X_a$  и  $X_b$ , которые мы обозначим  $A$  и  $B$ . (Здесь мы пользуемся лишь одним из ряда возможных двумерных представлений задачи, в которой фигурируют три измерения  $A$ ,  $B$  и  $O$ ). На рис. 2а представлен случай, когда имеются значимые главные эффекты  $A$  и  $B$ , но отсутствует их взаимодействие. (Разумеется, происходит суммирование эффектов с максимумом для  $A_3 B_3$ , но без взаимодействия, ибо эффекты аддитивны.) Во всех остальных случаях имеется значимое взаимодействие помимо или взамен главных эффектов  $A$  и  $B$ . Иначе говоря, характер эффекта  $A$  зависит от конкретного значения  $B$ . В этом смысле взаимодействие представляет собой «правило специфичности эффекта» и, следовательно, имеет отношение к проблеме обобщения результатов.

Эффект взаимодействия, показанный на рис. 2г, хорошо иллюстрирует сказанное. Здесь отсутствует главный эффект  $A$  (то есть, подсчитав средние значения  $O$  для каждого  $A$  по всем  $B$ , мы получим горизонтальную линию). Но когда  $B$  находится на уровне 1, увеличение  $A$  приводит к снижению  $O$ , а когда  $B$  находится на уровне 3, увеличение  $A$  приводит к возрастанию  $O$ . Отметим, что, если бы экспериментатор менял только  $A$ , оставляя  $B$  постоянным на уровне 1, результаты, хотя и были бы внутренне валидными, подсказывали бы ошибочное обобщение на случаи  $B_2$  и  $B_3$ . Использование более чем одного фактора в плане позволяет провести изучение возможностей обобщения или внешней валидности любого итогового утверждения о главном эффекте  $A$ . Ограничение этих возможностей или специфичность эффектов обнаруживают себя при статистическом анализе в виде значимого взаимодействия.

На рис. 2д изображена еще более крайняя ситуация: ни  $A$ , ни  $B$  не имеют главного эффекта (нет об-



Р и с. 2. Некоторые возможные результаты факторного эксперимента, проведенного по схеме  $3 \times 3$ .

щих правил, позволяющих судить, какой из двух уровней лучше), но наблюдается сильное и определенное взаимодействие между  $A$  и  $B$ . Рассмотрим гипотетический результат такого рода. Предположим, три типа учителей (например, склонные к спонтанной импровизации, к тщательной подготовке к уроку и полной опеке учеников) в целом работают одинаково эффективно.

96

Точно так же три метода обучения (например, групповая дискуссия, лекционный метод и индивидуальная работа с учениками) в целом обладают одинаковой эффективностью. В таком случае даже в отсутствие «главных эффектов» для типа учителя или метода обучения эти два фактора будут сильно взаимодействовать друг другом: импровизатор лучше проведет групповую дискуссию и хуже всего организует индивидуальную работу, тогда как сторонник опеки учеников продемонстрирует максимум эффективности, работая по индивидуальному методу, но плохо справится с проведением дискуссии.

Следует различать типы обнаруживаемых значимых взаимодействий. Тут нам, вероятно, поможет понятие «монотонность взаимодействия». Отметим, что в случаях, соответствующих рис. 2а и 2б, имеет место главный эффект  $A$ , и  $B$ , причем главный эффект  $A$  направлен одну и ту же сторону для любого отдельного набора значений  $B$ . Таким образом, в новой ситуации мы с гораздо большей уверенностью можем ожидать увеличения  $O$  с ростом  $A$ , чем в случае 2, когда также могут иметь место значимые главные эффекты и значимое взаимодействие  $A$  и  $B$ . В случае 2б мы могли бы, собственно говоря, быть почти так же уверены в общем характере главного эффекта  $A$ , как и в случае 2а, когда взаимодействие отсутствует. Таким образом, выясняя возможность

обобщения результатов, мы должны построить и детально изучить соответствующий график. Некоторые «монотонные», или однонаправленные, взаимодействия почти или совсем не порождают ограничений специфичности эффекта (см. Lubin [72], где приводится широкое обсуждение данной проблемы).

### Нестинг<sup>1</sup>

Во всех приведенных до сих пор примерах каждый из классификационных критериев (например, различные

---

<sup>1</sup> В контексте теории эксперимента нет точного русского эквивалента английского слова «nesting». Буквально его следует понимать как набор однородных объектов внутри более крупных единиц (nests), наподобие, скажем, картотечных ящиков в секциях. Попытка перевести этот термин как «группировка» (см., например: Х и К с Ч. Основные принципы планирования эксперимента. М., «Мир», 1967) представляется не вполне удачной, так как это слово весьма многозначно и не содержит указаний на специфический ха-

*A* и *B*) «перекрещивался» со всеми остальными критериями, то есть каждый уровень *A* фигурировал на всех уровнях *B*. Однако дисперсионный анализ не ограничивается только этой ситуацией.

До сих пор в роли критериев классификации у нас выступали экспериментальные воздействия. Но во многих экспериментах могут быть использованы и другие признаки перекрестной классификации — например, пол или возраст учеников. Чтобы пояснить, в каких смыслах обычно употребляется термин «нестинг» применительно к классификации, нам придется обратиться к менее очевидным классификационным критериям.

Возьмем, к примеру, переменную «учитель». Представим себе эксперимент, в котором 10 учителей применяют два метода обучения данному предмету. В этом случае каждый отдельный учитель представляет собой конкретный «уровень». Полное перекрещивание переменной «учитель» с переменной «метод» будет иметь место, когда каждый учитель применяет оба метода в различных классах. Наличие «главного эффекта» фактора «учитель» свидетельствовало бы о том, что одни учителя работают лучше других, независимо от метода, который они используют. (Учащиеся или классы должны выбираться в случайном порядке, иначе произойдет смешивание двух факторов — особенностей учителей и состава классов.) Значимое взаимодействие между переменными «учитель» и «метод» означало бы, что для одних учителей более эффективен один метод, а для других — другой.

Предположим теперь, что, желая проследить далее такое взаимодействие, мы заинтересовались тем, кто из учителей — мужчины или женщины — работал лучше по данной методике. Разделив наших учителей на

---

рактер данной «группировки». Несколько более адекватным представляется использование в этой связи термина «иерархическая классификация». Однако и он является скорее формальным, нежели семантическим эквивалентом исходного термина, причем даже формальная эквивалентность лишь приближительна, ибо «nesting» соответствует только двухуровневой иерархии. По-видимому, мы здесь сталкиваемся с тем редким случаем, когда иноязычный термин лучше не переводить, а просто дать в русской транскрипции. — *Прим. перев.*

5 мужчин и 5 женщин, мы приходим к «нестингу»: теперь переменная «учитель», хотя и полезная по-прежнему, не перекрещивается с переменной «пол», то есть один и тот же учитель не может принадлежать к тому и другому полу, в то время как «учитель» и «пол» перекрещиваются с «методом». Такой нестинг требует несколько иного анализа

данных, чем в том случае, когда все классификации перекрещиваются друг с другом. (Соответствующие иллюстрации см. в: Green, Tukey [45], Stanley [111].) Кроме того, при нестинге исключены некоторые взаимодействия переменных. Так, взаимодействия «учитель» — «пол» и «учитель» — «пол» — «метод» нельзя рассчитать, и в концептуальном плане они не имеют смысла.

Распространим теперь данный эксперимент на несколько школ, так что «школа» становится переменной, главный эффект которой мог бы отражать различия в скорости усвоения материала учениками разных школ. В этих условиях, скорее всего, будет иметь место нестинг учителей по школам, так как каждый учитель обычно ведет уроки только в одной школе. Хотя в этом случае возможно взаимодействие «учитель» — «школа», его можно рассчитать только в том случае, если все учителя работают одновременно во всех школах, охваченных экспериментом. Но тогда нестинг уступает место «перекрещиванию» этих признаков.

Учащихся, или испытуемых в эксперименте, также можно рассматривать как признак классификации (переменную). При полном перекрещивании каждый учащийся проходит через все режимы эксперимента. Но часто случается так, что ученики проходят лишь через некоторые, а не все режимы, то есть имеет место нестинг. Такая ситуация часто возникает, например, при поэтапном изучении процесса научения. В этом случае можно получить кривые научения для каждого из испытуемых, распределенных по двум способам обучения. Переменная «испытуемый» перекрещивается с «порядковым номером пробы», но не со «способом обучения». Здесь мы можем изучать взаимодействие «испытуемый» — «номер пробы», но не «испытуемый» — «способ обучения». Точно так же нестинг возникнет в том случае, когда ученики расклассифицированы по признаку пола.

Большинство переменных, представляющих интерес в педагогическом эксперименте, может взаимно перекрещиваться, и нестинг для них не обязателен. Помимо уже названных случаев, исключение составляют: хронологический возраст, умственный возраст, школьный класс (первый, второй и т. д.) и социально-экономический уровень. Наблюдательный читатель, вероятно, уже заметил, что независимые переменные, или признаки классификации, бывают нескольких видов: 1) управляемые переменные (например, метод обучения, выбираемый по усмотрению экспериментатора); 2) потенциально управляемые переменные (например, школьные предметы, которые экспериментатор может выбирать случайным образом, но он редко пользуется этой возможностью); 3) относительно постоянные аспекты окружения (населенный пункт, школа, социально-экономический уровень), которые не находятся под непосредственным контролем экспериментатора, но которые в эксперименте служат четкими основами стратификации; 4) объективные («организмические») характеристики учеников (возраст, рост, вес, пол) и 5) характеристики учеников, проявляющиеся в их ответах (результаты различных тестов). Обычно первостепенный интерес представляют управляемые независимые переменные 1-го вида, хотя неуправляемые независимые переменные 3-го, 4-го и иногда 5-го вида служат повышению точности и выяснению, насколько эффекты управляемых переменных поддаются обобщению. Переменные 5-го вида обычно фигурируют в качестве сопутствующих или зависимых от других показателей переменных. Другой способ рассмотрения независимых переменных — это учет присущей им упорядоченности (школьный класс, социально-экономический уровень, рост, номер пробы и т. д.) или неупорядоченности (методика, учебный предмет, учитель, пол и т. д.). Эффекты упорядоченных переменных часто могут быть подвергнуты дальнейшему анализу для выяснения того, является ли тренд линейным, квадратичным, кубическим и т. д. (Grant [44], Myers [83]).

**Модели дисперсионного анализа.  
Конечные, рандомизированные,  
фиксированные и смешанные модели**

Под влиянием неопубликованной рукописи Тьюки (1949 г.) были разработаны модели дисперсионного анализа для конечного числа факторных уровней. Эти модели основываются на принципах, ранее хорошо разработанных для выборки из конечной популяции. Шеффе [98] дал исторический обзор работ в этом направлении. Математические ожидания средних квадратов, которые помогают определить соответствующее значение среднего квадрата ошибки, были получены для полностью рандомизированного трехфакторного плана (Stanley [107]). Такие модели особенно полезны, поскольку они могут быть немедленно распространены на случаи одного или большего числа факторов с фиксированными или случайными уровнями. Простое объяснение этого обобщения дано Фергюсоном [36].

Вместо того чтобы приводить формулы, мы дадим словесную иллюстрацию, показывающую, чем отличаются друг от друга конечный, случайный и фиксированный отборы уровней фактора. Предположим, что признак «учитель» — одно из нескольких оснований классификации (то есть независимых переменных) в эксперименте. Если у нас имеется 50 учителей, мы можем выбрать 5 из них *наугад* и использовать их в эксперименте. Тогда в некоторых из наших формул появится коэффициент выборки факторных уровней, равный  $(1 - 5/50)$ , то есть 0,9. Если в эксперименте принимают участие все 50 учителей, то мы будем иметь «фиксированные» уровни факторов, и этот коэффициент станет  $(1 - 50/50) = 0$ . Если бы, с другой стороны, существовала неограниченная популяция учителей, отобранные наугад 50 из них составили бы бесконечно малый процент, так что коэффициент приблизился бы к единице для каждого «случайного» эффекта. От значений этого коэффициента зависит вид формул для математического ожидания средних квадратов и, следовательно, для ожидаемых ошибок.

**Другие направления обобщения**

Прежде чем перейти к рассмотрению квазиэкспериментов, мы хотим обсудить другие возможности обобщения планов «истинных» экспериментов, исходящих из

того же простого ядра и действительных для всех обсуждаемых ниже типов экспериментов.

**Проверка наличия эффекта в различные моменты времени**

В исследованиях убеждающих воздействий, которые в известном смысле сродни педагогике, Ховленд и его сотрудники неоднократно обнаруживали, что долговременные эффекты отличаются от кратковременных не только количественно, но и качественно. Долговременные эффекты более выражены для общих и слабее для специфических социальных установок (Hovland, Lumsdain, Sheffield [51]). Речь дискредитированного оратора не производит немедленного убеждающего действия на слушателей, но может давать значительный эффект месяц спустя, если только слушателям не напоминать о первоисточнике (Hovland, Janis, Kelley [50]). Эти факты предупреждают нас о том, насколько рискованно ограничиваться в оценке методики обучения немедленным или вообще однократным тестированием после введения экспериментального воздействия. Хотя это очень затрудняет проведение исследования, можно лишь рекомендовать

включение в его программу тестирований, проводимых, скажем, через месяц, полгода и год. Правда, дело сводится к обычному учету (включая сюда и учет выбывших испытуемых), если в качестве итоговых показателей берутся годовые школьные оценки или результаты контрольных работ, которые все равно подлежат регистрации. Но если тестирование проводится экспериментатором, то, как считает большинство авторов, получение повторных данных на одних и тех же учениках еще опаснее, чем использование предварительного тестирования. Это мнение, разумеется, подтвердилось в исследовании памяти (см., например, Underwood [128]). Хотя Ховленд и его сотрудники, как правило, проводили предварительное тестирование (план 4), они также составляли отдельные группы — экспериментальные и контрольные — для каждого периода проведения последующего тестирования, например:

$$\begin{array}{cccc}
 R & O & X & O \\
 R & O & & O \\
 R & O & X & O \\
 R & O & & O.
 \end{array}$$

102

Подобное дублирование групп требовалось бы и для плана 5 или 6. Отметим, что этот план не обеспечивает полного контроля, если мы хотим установить, как экспериментальный эффект зависит от времени, истекшего после введения  $X$ : наблюдаемые различия могут объясняться взаимодействием между  $X$  и определенными фоновыми событиями, которые могли произойти в промежутке между «близким» и «отдаленным» последующими тестированиями. Полный контроль этой возможности требует разработки еще более сложных планов эксперимента. Подобные исследования обходятся чрезвычайно дорого. Исключение составляют случаи, когда тестирование является составной частью обычной деятельности исследуемого института. Поэтому те, кто располагает данными таких тестирований, должны воспользоваться этой возможностью для прослеживания экспериментального эффекта в различные периоды времени.

**Обобщение на другие  $X$ :  
вариативность реализаций  $X$**

Цель науки предполагает распространение результатов не только на другие популяции или периоды времени, но и на другие неидентичные случаи предъявления  $X$ , то есть на те воздействия, которые в теории будто идентичны  $X$ , но в действительности отличаются от  $X$  теоретически несущественными элементами. Эта цель вступает в противоречие с часто встречающимся широким толкованием необходимости экспериментального контроля, которое выражается в стремлении к *точному* воспроизведению  $X$  при каждом повторении эксперимента. Так, сравнивая эффекты эмоциональных и рациональных речей, можно было бы заставить одного оратора произнести все речи перед группами слушателей каждого типа. Есть даже еще лучшая возможность: записать речь на магнитную ленту так, чтобы все группы, проходящие тестирование по этому уровню воздействия, услышали «совершенно одно и то же». Такой способ проведения эксперимента может показаться более совершенным, чем привлечение нескольких ораторов, каждый из которых произнес бы одну-единственную речь. Действительно, в последнем случае мы «точно не знали бы», какой экспериментальный стимул предъяв-

103

ляется той или иной группе. Но ошибочно думать, что мы в состоянии ограничить экспериментальное воздействие рамками некоей абстрактной классификации и эффективно передавать эту информацию все новым испытуемым. В случае записанного на магнитофон интервью мы каждый раз воспроизводим множество специфических иррелевантных деталей. А кто знает, может быть, эти детали, а не сознательно выбранные нами компоненты воздействия и вызвали данный эффект. Если, однако, мы воспользуемся многочисленными независимыми реализациями воздействия, то специфические иррелевантные детали не будут воспроизводиться каждый раз в том же виде и наша интерпретация источника эффектов будет иметь больше шансов оказаться правильной.

Рассмотрим, например, исследование Гецкоу, Келли и Мак-Кичи [47], в котором сравнивалась эффективность опроса учеников и дискуссии. В эксперименте участвовало восемь учителей, каждый из которых интерпретировал оба метода по-своему. В результате нам лучше известно, каковы были экспериментальные воздействия (с точки зрения составления рекомендации учителям), чем, если бы использовался только один учитель или восемь учителей, усвоивших одни и те же детали, не включенные в абстрактное описание сравниваемых методов. (Этот акцент на различные реализации  $X$  должен по возможности сопровождаться, как у Гецкоу и др., применением обоих методов всеми учителями, участвующими в эксперименте, с тем, чтобы специфические побочные компоненты урока не смешивались с конкретным воздействием. Чтобы оценить значимость взаимодействия «учитель» — «метод» при использовании целых классов, каждый учитель должен дать по два урока, применяя каждый из двух методов.)

Еще более наглядный пример: при изучении влияния пола учителя на эффективность начального обучения арифметике следует привлечь к эксперименту многих учителей обоого пола, а не по одному на тот и другой пол. Хотя это положение очевидно, ему, как отмечает Хэммонд [48], следуют не всегда. Эта проблема является одним из аспектов проблемы репрезентативности эксперимента, рассматриваемой Брунsvиком [9]. Андервуд [129, с. 281—287] по аналогичным соображениям

выступал против точной стандартизации или точного воспроизведения аппаратуры в различных экспериментах, что отнюдь не противоречит его строгому операционализму.

**Обобщение на другие  $X$ :  
последовательная детализация  $X$  и  
новые контрольные группы**

Реальное воздействие ( $X$ ) в любом эксперименте является комплексным, включающим компоненты, которые могут быть концептуализированы как различные переменные. Когда обнаруживается сильный и отчетливый эффект, логика науки требует проведения все более тонких экспериментов, выделения тех аспектов, от которых больше всего зависит эффект. Это может достигаться путем повышения четкости в определении и предъявлении воздействий или с помощью новых контрольных групп, уравненных с экспериментальной группой по все большему числу показателей, благодаря чему различия сводятся к более специфическим признакам первоначального  $X$ . В качестве примера сошлемся на введение в медицинские исследования контрольных групп пациентов, принимающих плацебо или подвергающихся инсценированной операции. Первоначальные эксперименты обнаруживали внутренне валидный эффект, который, однако, мог быть вызван тем, что пациент знал о том, что он подвергается лечению, или хирургическим шоком, а не специфическими свойствами принимаемого лекарства или удалением мозговой ткани. Отсюда и использование особых контрольных групп, назначение которых — исключить подобные предположения. Процесс обобщения на другие  $X$  — направляемая теорией

поисковая экстраполяция, осуществляемая посредством проб и ошибок, и в этом процессе подобное уточнение  $X$  может играть важную роль.

### Обобщение на другие $O$

Подобно тому, как за любым  $X$  тянется шлейф его теоретически несущественных особенностей, так и любой конкретный измерительный инструмент есть комплекс, в котором релевантное содержание связано с конкретной инструментальной формой, детали которой не имеют отношения к теоретической цели. Так, если мы

105

пользуемся для регистрации ответов специальными карандашами и бланками фирмы IBM<sup>1</sup>, то это скорее по соображениям удобства, а не из-за того, что мы хотим включить в наши результаты дисперсию, обусловленную навыками работы с машинными бланками, знакомством с заполняемой формой, способностью следовать инструкции и т. д. Точно так же проверку знаний по пройденному материалу посредством письменных работ приходится осуществлять с учетом индивидуального стиля и лексики, и поэтому результирующие колебания показателей должны содержать дисперсию, обусловленную этими источниками, что, как правило, не входит в число наших целей. Ввиду комплексного характера любого  $O$  попытка распространить результаты на другие потенциальные  $O$  ставит нас перед проблемой: каким аспектом примененного в эксперименте  $O$  был обусловлен внутренне валидный эффект? Поскольку цель обучения — не просто подготовка к будущим сочинениям и контрольным работам, мы должны все время иметь в виду эту проблему внешней валидности, или обобщаемости результатов.

И снова с концептуальной точки зрения следует не уповать на «чистые» показатели, свободные от иррелевантных наслоений, а использовать множество разных измерений, в которых специфические средства реализации, конкретные иррелевантные детали были бы по возможности различными, а общее, существенное для нас содержание присутствовало бы в каждом из них. В рамках одного эксперимента в этом плане можно больше сделать применительно к  $O$ , чем к  $X$ , так как в один и тот же эксперимент обычно можно включить целый ряд мер эффекта (то есть зависимых переменных). В исследовании Гецкоу, Келли и Мак-Кичи [47] эффекты регистрировались не только в ходе экзаменов и проведения специальных тестов установок, но учитывались и такие последующие действия, как выбор профилирующего предмета и запись на продвинутый курс по данной проблематике. (Эти последние оказались столь же чувствительными к различиям в экспериментальных воздействиях, как и результаты тестов.)

---

<sup>1</sup> IBM (International Business Machines) — одна из крупнейших американских фирм, производящих вычислительную аппаратуру, в частности, для школ. — *Прим. перев.*

106

*Требование множественности  $O$  должно стать стандартом для любого исследования методов обучения.* В простейшем случае следует сочетать результаты письменных работ и экзаменов (см. Stanley, Веeman [115]), присоединяя к ним по возможности данные относительно активности учащихся на уроке. (Распространение этого подхода на вопросы валидности тестов содержится в: Campbell, Fiske [15]; Campbell [12]).

### Квазиэкспериментальные планы

Существует немало естественных социальных условий, в которых исследователь может использовать нечто вроде планирования эксперимента при сборе данных (то есть в

отношении того, *когда* и *на ком* производить измерения), даже если он не полностью контролирует порядок экспериментальных воздействий (то есть *когда* и *кому* предъявляются воздействия, рандомизацию предъявления), что делает возможным подлинный эксперимент. Такие исследования в целом могут рассматриваться как квазиэксперименты. Одна из целей настоящей главы — пробудить интерес к таким квазиэкспериментам и уточнить условия, в которых возникает возможность их проведения. Но как раз из-за отсутствия полного контроля экспериментальных параметров исследователь должен отдавать себе отчет в том, какие именно переменные в его конкретном плане не поддаются контролю. Собственно говоря, список источников невалидности, приведенный в табл. 1, 2 и 3, был разработан скорее ради оценки квазиэкспериментов, чем для понимания истинного эксперимента.

Ознакомившись с предыдущим параграфом, будущий исследователь, вероятно, почувствовал, что планирование эксперимента сопряжено с более многочисленными трудностями, чем он полагал ранее. Такой итог следует приветствовать, если результатом будет планирование и проведение более совершенных экспериментов и большая осмотрительность в выводах, которые делаются на основе полученных результатов. Однако мы вовсе не хотели бы, чтобы у читателя сложилось впечатление, будто контроль параметров эксперимента все равно недостижим, и прекратил направленные на это усилия,

107

обратившись к еще более неформальным методам исследования. Кроме того, этот внушительный список источников невалидности может с еще большей вероятностью отпугнуть от проведения квазиэкспериментов, раз уж с самого начала, очевидно, отсутствие полного контроля за экспериментальными параметрами. Подобный итог был бы прямо противоположен нашим намерениям.

С точки зрения конечной интерпретации эксперимента и попыток вписать его в развивающуюся науку, каждый эксперимент несовершенен. Польза от списка факторов валидности могла бы состоять в том, чтобы экспериментатор лучше понимал недостатки своего плана, которые он не может устранить, и соответственно знал бы о возможной альтернативной интерпретации полученных данных. Конечно, он должен планировать самый лучший эксперимент, какой только возможен в данной ситуации. Он должен сознательно выискивать такие искусственные и естественные лаборатории, которые обеспечивают наилучшие возможности контроля. Но после этого ему нужно провести эксперимент и интерпретировать результаты, полностью сознавая, в каких пунктах эти результаты двусмысленны. Такое понимание важно и для осуществления экспериментов с «полным» контролем, однако оно абсолютно необходимо в случае квазиэкспериментов.

Осуществляя этот общий замысел, мы в этом разделе дадим обзор сильных и слабых сторон разнородных планов квазиэкспериментов, каждый из которых заслуживает использования, *если применение лучшего плана невозможно*. Сначала будут обсуждены три плана эксперимента с одной группой. Затем будут представлены пять общих типов экспериментов с многими группами. Отдельный параграф отводится вопросам корреляции, планам *ex post facto*, панельным исследованиям и т. д.

### **Некоторые предварительные замечания о теории экспериментирования**

Этот параграф написан, прежде всего, для специалиста в области прикладной науки, который хочет перенести свое исследование из лаборатории в «производственные» условия. Авторы не могут не сознавать, что

108

психологи-экспериментаторы с большим подозрением относятся к любой попытке санкционировать исследования, которые не предполагают полного контроля экспериментальных параметров. Следующие общие соображения, касающиеся роли эксперимента в науке, отчасти приводятся ради оправдания квазиэксперимента в глазах таких ревнителей чистоты исследования. Мы надеемся, что эти соображения согласуются с большинством современных воззрений в области философии науки и исходят из того, что могло бы стать в будущем общей психологией индуктивных процессов (Campbell [11]).

Наука, как и другие процессы познания, связана с выдвижением теорий, гипотез, моделей и т. д. и с принятием или отклонением их на основе некоторых внешних критериев. Экспериментирование принадлежит к этой второй фазе — к фазе прореживания, отклонения, редактирования. Мы можем предположить существование своего рода экологии науки, в которой число потенциально позитивных гипотез значительно превышает число гипотез, которые, пройдя проверку, окажутся в дальнейшем совместимыми с нашими наблюдениями. *Задача сбора данных для испытания теории — это преимущественно задача отбрасывания несостоятельных гипотез.* Для решения этой задачи полезен любой способ организации наблюдений, определенные результаты которых опровергают теорию, в том числе квазиэксперименты, обладающие меньшей эффективностью, чем истинные эксперименты.

Но позволительно спросить: не приведут ли такие несовершенные эксперименты к незаконному подтверждению ошибочной теории, не направят ли по ложному пути наши дальнейшие поиски и не поглотят ли напрасно в наших журналах место десятки статей, посвященных опровержению ошибочно принятой, но броско поданной гипотезы? Риск серьезный, но на него мы должны пойти. Это риск того же рода, если не в такой же степени, как и в случае «истинных» экспериментов, проводимых по планам 4, 5 и 6. Дело в том, что результаты эксперимента никогда не «подтверждают», не «доказывают» теорию. Скорее, успешная теория проверяется и избегает опровержения. Слово «доказывает» из-за его частого употребления для обозначения дедук-

тивной валидности приобрело в нашем поколении значение, не соответствующее ни его прежнему употреблению, ни его применению в индуктивных процедурах, таких, как экспериментирование. Результаты эксперимента «опробывают» (probe), а не «доказывают» теорию. Адекватная гипотеза — это гипотеза, которая выдержала неоднократно такие проверки, но она всегда может быть отвергнута новым испытанием.

Теперь уже все понимают, что «нулевая гипотеза», часто используемая как удобный способ формулировать гипотезу эксперимента, никогда не может быть «принята» в свете полученных данных. Она может быть только или «отвергнута», или «не отвергнута». Так же обстоит дело и с гипотезами в более общем плане: они формально никогда не «подтверждаются». Если мы из удобства и пользуемся этим термином, то, скорее, имеем в виду, что гипотеза была подвергнута критической проверке, но не опровергнута. Эта точка зрения согласуется со всеми положениями юмовской философии науки, которые подчеркивают невозможность дедуктивного доказательства индуктивных законов. Недавно Хэнсон, [49] и Поппер [90] особенно ясно высказались по этому вопросу. Многие массивы данных, полученные в педагогических исследованиях, мало или вообще не пригодны для проверки гипотез, а многие системы гипотез столь тесно связаны между собой, что их нельзя опровергнуть доступными нам средствами проверки. Мы не намерены ратовать за такие псевдоисследования. Обсуждаемые ниже планы экспериментов, как мы надеемся, обладают достаточной эффективностью в качестве инструмента проверки гипотез, однако ими следует пользоваться лишь в том случае, когда *недоступны более эффективные средства проверки.*

Мнение, что эксперимент никогда не «подтверждает» теорию, хотя и правильно, но настолько противоречит нашим установкам и опыту ученых, что является почти нетерпимым. Оно кажется особенно неудовлетворительным, когда знакомишься с изящными, поразительными подтверждениями теории, нередко встречающимися в физике и химии, где результаты эксперимента могут до мельчайших деталей совпадать в многочисленных точках измерения со сложной кривой, предсказанной теорией. И такое представление становится феноменологи-

чески неприемлемым для большинства из нас, если его распространить на индуктивные процессы зрения. Так, трудно свыкнуться с мыслью, что столы и стулья, которые мы «видим» перед собой, не «подтверждены», не «доказаны» визуальными данными, но суть «всего лишь» гипотезы относительно внешних объектов, пока еще не опровергнутые многочисленными проверками в процессе функционирования зрительной системы<sup>1</sup>. В этом нашем внутреннем противодействии есть зерно истины.

Степень «подтверждения» определяется для той или иной теории числом *правдоподобных конкурентных гипотез*, которые могут быть привлечены для объяснения результатов. Чем меньше остается таких правдоподобных конкурентных гипотез, тем больше степень «подтверждения». Надо полагать, на каждой стадии накопления данных, даже в случае самой развитой науки, существует множество совместимых с результатами теорий, особенно если допустить все теории, включающие сложные условные данные. Однако у «вполне установленных» теорий и теорий, полностью опробованных сложными экспериментами, остается мало или вовсе не остается серьезных конкурентов. Эпистемологически это соответствует подтверждению теории в результате изящных экспериментов. Столь же малое число конкурентных гипотез имеет место в знании, которое позитивно в феноменальном плане; очевидно, такое знание дает зрение в отличие, например, от относительной неоднозначности слепого тактильного обследования.

В этом плане список источников невалидности, контролируемых в экспериментальных моделях, можно рассматривать как перечень часто возникающих правдоподобных гипотез, конкурирующих с гипотезой об эффекте, вызванном экспериментальной переменной. План эксперимента, ставящий какой-либо побочный фактор «под контроль», просто делает соответствующую конкурентную гипотезу маловероятной, даже если при некотором стечении обстоятельств этот фактор все еще способен вызвать полученный в эксперименте эффект.

---

<sup>1</sup> См. также: К э м п б е л л Д. Т. Слепые вариации и селективный отбор как главная стратегия процессов познания. — В кн.: Самоорганизующиеся системы. М., «Мир», 1964. — *Прим. ред.*

«Правдоподобные конкурентные гипотезы», которые делают необходимым традиционное использование особых контрольных групп, имеют статус «вполне установленных» эмпирических законов. Это эффект тренировки, требующий введения контрольной группы в план 2, внушаемость (контрольные группы здесь вводятся для приема плацебо), хирургический шок (контрольные группы пациентов, подвергаемых ложным операциям)<sup>1</sup>. Конкурентные гипотезы остаются правдоподобными, пока мы склонны приписывать им статус эмпирических законов. Если какой-либо параметр не контролируется в квазиэксперименте, необходимо при интерпретации результатов тщательно выяснить вероятность их объяснения за счет неконтролируемых факторов. Чем невероятнее такое объяснение, тем «валиднее» эксперимент.

Как отмечалось при обсуждении плана Соломона для четырех групп (план 5), чем многочисленнее и независимее способы, с помощью которых демонстрируется

экспериментальный эффект, тем менее правдоподобна любая отдельно взятая гипотеза, ставящая под сомнение валидность эксперимента, и тем меньше число таких гипотез. Здесь мы обращаемся к принципу экономичности. «Валидность» эксперимента сводится к относительной вероятности конкурирующих теорий — теорий, объясняющих эффект фактором  $X$ , и теорий, приписывающих эффект действию неконтролируемых факторов. Если некоторые различия могут быть полностью объяснены единственной гипотезой о том, что эффект вызван именно  $X$ , в то время как для каждого наблюдаемого отклонения нужно каждый раз строить новые предположения о действии неконтролируемых факторов, то эффект фактора  $X$  становится наиболее убедительным. К такой логике вывода часто прибегают в обзорах литературы по экспериментам, в которых отсутствует полный контроль. Так, Уотсон [131, с. 296] нашел убедительной гипотезу об отрицательном действии длительного отсутствия матери, ибо это подкреплялось целым рядом разнородных данных, тогда как специфические неадекватные детали в этих данных были раз-

---

<sup>1</sup> Этот вопрос подробнее рассматривается в работе «Научный вывод, артефакты и контроль» (см. перевод в данной книге). — *Прим. ред.*

личными в разных исследованиях. Гликман [40], несмотря на наличие правдоподобных конкурентных гипотез в каждом из рассмотренных им исследований, считал данные в пользу процесса консолидации убедительными именно потому, что правдоподобная конкурентная гипотеза менялась от исследования к исследованию. Этот подход, обычно применяемый при комбинировании выводов нескольких исследований, сознательно введен в некоторые планы квазиэкспериментов, особенно в «лоскутные» («patched-up») планы вроде плана 15.

Принцип экономичности не столько оправдан с дедуктивной точки зрения, сколько является общим предположением о природе мира, лежащим в основе почти всех теоретических построений в науке, даже, несмотря на его частую несостоятельность в конкретных приложениях. С ним связана другая аргументация правдоподобия, к которой мы обратимся главным образом при обсуждении широко распространенного плана 10 (хороший план квазиэксперимента, который нередко путают с экспериментом по плану 4, принадлежащим к разряду истинных экспериментов). Эта аргументация исходит из того, что главный эффект одной переменной считается более вероятным, чем взаимодействие двух других переменных, или в более общем виде — главный эффект более вероятен, чем эффект взаимодействия. В пределе мы можем отметить, что если каждое взаимодействие высшего порядка значимо, если каждый эффект специфичен для определенных значений по всем другим потенциальным факторам, то наука невозможна. Если мы все-таки можем делать обобщения, то только потому, что множеством потенциально определяющих факторов можно пренебречь. Андервуд [129, с. 6] ссылаясь на это как на постулат конечной каузальной связи.

Следующие страницы посвящены экспериментам, в которых изучается одна-единственная группа. Начиная с 1920 гг. самыми распространенными планами экспериментов в психологии и педагогике стали планы с применением контрольной группы (см. планы 4, 6 и особенно план 10, о котором пойдет речь ниже). В социальных науках и полевых исследованиях планы экспериментов, включающие контрольные группы, заняли столь господствующее положение, что, похоже, сделались

для многих синонимом эксперимента вообще. В результате возникла опасность того, что ряд научных работников откажется от процедур, сходных с экспериментированием, если нельзя будет сформировать контрольные группы, и в итоге исследование проиграет в

точности больше, чем было бы необходимо. Существует, однако, несколько разновидностей квазиэкспериментальных планов, применяемых при обследовании отдельных групп, и этими планами с успехом можно пользоваться, сохраняя логику экспериментального подхода и возможность интерпретации результатов во многих случаях, когда обращение к контрольной группе невозможно. Так, эксперимент часто проходит в производственных условиях — учитель имеет в своем распоряжении собственный класс, директор средней школы может проводить периодические опросы и т. д. В таких случаях дифференцированное предъявление воздействия различным участникам обследования (что необходимо для эксперимента с использованием контрольных групп) часто невозможно по причинам административного порядка, а если даже возможно, то нежелательно из-за возникновения реакции испытуемых на эксперимент. В таких условиях стоит взвесить целесообразность ограничиться планами эксперимента с использованием одной группы испытуемых.

### 7. Эксперимент по плану временных серий

Суть эксперимента данного типа состоит в осуществлении серии периодических замеров на некоторой группе или индивиде с введением экспериментального воздействия посреди серии. На эффект воздействия указывает нарушение непрерывности результатов измерения, регистрируемых в этой серии. Этот план может быть представлен схематически следующим образом:

$$O_1 O_2 O_3 O_4 X O_5 O_6 O_7 O_8.$$

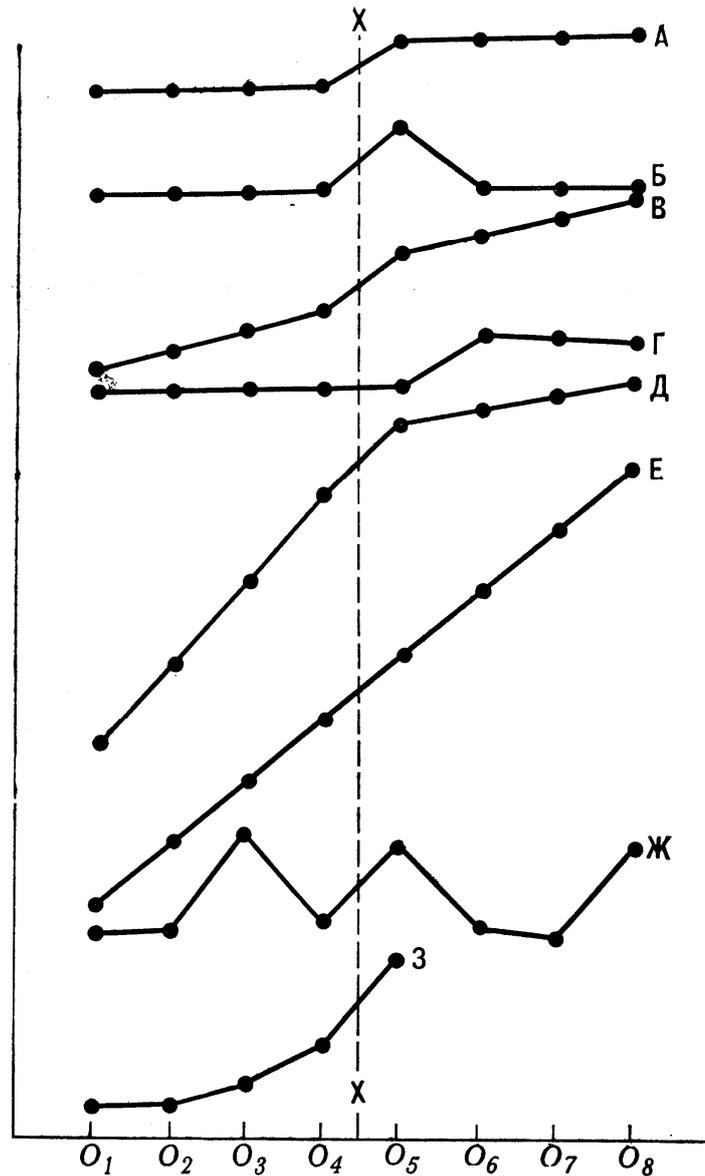
Этот экспериментальный план типичен для многих классических исследований, проводившихся в XIX в. в физических науках и биологии. Так, если железный брусок, вес которого оставался неизменным в течение не-

скольких месяцев, окунуть в азотную кислоту, а затем вынуть из нее, то вывод, связывающий пребывание бруска в азотной кислоте и уменьшение его веса, будет основываться, в общем, на той же логике. На полке могли бы находиться «контрольные группы» брусков, вес которых оставался бы неизменным, однако вряд ли кто-нибудь счел бы нужным их взвешивать. Вероятно, этот тип эксперимента признается валидным в более успешных науках, тогда как он редко удостоивался упоминания в списках экспериментальных планов, применяемых в социальных науках (см., однако, Maxwell [73]; Underwood [129, с.133]). Различное отношение к этому типу эксперимента имеет свои основания, тщательное рассмотрение которых позволит лучше понять, когда он может осмысленно применяться в социальных науках в отсутствие более полного контроля параметров эксперимента. Данный план типичен для классических процедур, применявшихся Британской комиссией по исследованию производственного утомления при изучении факторов, влияющих на выпуск продукции (например, Farmer, Brooks, Chambers [34]).

На рис. 3 изображены некоторые возможные результаты серии периодических замеров, то есть так называемые временные ряды. Момент включения экспериментального воздействия всюду отмечен вертикальной линией  $X$ . Предположим, что исследователь будет склонен делать вывод о наличии эффекта  $X$  в случаях А, Б, а также, вероятно, в случаях В, Г и Д и не будет склонен делать его в случаях Е, Ж и З, даже если скачок в результатах замера от  $O_4$  к  $O_5$  в них столь же велик и статистически стабилен, как и для А и Б, например. Отложив пока обсуждение статистической обработки, отметим лишь, что проблема внутренней валидности сводится к вопросу о правдоподобных конкурентных

гипотезах, которые дают вероятные альтернативные объяснения сдвигу во временном ряду за счет факторов, отличных от  $X$ . Факторы, которые в оптимальных условиях контролируются рассматриваемым планом эксперимента, приводятся в табл. 2. Сильные стороны этого типа эксперимента особенно заметны на фоне плана 2, с которым он имеет внешнее сходство, выражающееся в отсутствии контрольной группы и применении тестирования до и после  $X$ .

115



Р и с. 3. Некоторые возможные результаты введения экспериментального воздействия в точке  $X$  для серии периодических замеров  $O_1 — O_8$ . За исключением случая Г, увеличение от  $O_4$  к  $O_5$  одинаково для всех временных рядов, хотя право мерность вывода о наличии эффекта весьма различна: она максимальна для кривых А и Б и совершенно отсутствует в случаях Е, Ж и 3.

Просматривая список источников угроз внутренней валидности в табл. 2, мы видим, что слабым местом эксперимента по плану 7 является отсутствие контроля фона. Иначе говоря, возможна конкурентная гипотеза о том, что сдвиг в результатах вызван не  $X$ , а другими

Таблица 2

**Источники невалидности для квазиэкспериментальных  
планов 7—12**

	Источники невалидности											
	внутренней							внешней				
	фон	естественное развитие	эффект тестирования	инструментальная погрешность	регрессия	состав групп	выбывание	взаимодействие состава групп с естественным развитием и др. факторами	взаимодействие тестирования и X	взаимодействие состава групп и X	реакция испытуемых на эксперимент	взаимодействие между разными X
<i>Квазиэкспериментальные планы</i>												
7. Эксперимент по плану временных серий $O O O O X O O O O$	-	+	+	?	+	+	+	+	-	?	?	
8. Серии временных выборок $X_1 O X_0 O X_1 O X_0 O$ и т. д.	+	+	+	+	+	+	+	+	-	?	-	-
9. Серии эквивалентных воздействий $M_a X_1 O M_b X_0 O M_c X_1 O M_d X_0 O$ и т. д.	+	+	+	+	+	+	+	+	-	?	?	-
10. План с неэквивалентной контрольной группой $\frac{O \quad X \quad O}{O \quad \text{---} \quad O}$	+	+	+	+	?	+	+	-	-	?	?	



событиями, происшедшими примерно в то же самое время. Именно от правомерности устранения гипотезы о роли фона зависит достоверность интерпретации такого эксперимента. Взять, к примеру, эксперимент, предусматривающий повторные измерения для выяснения влияния документального фильма на оценку школьниками вероятности войны. Здесь отсутствие полного контроля над происходящими параллельно событиями будет иметь серьезные последствия, так как очевидно, что, помимо воздействий, контролируемых экспериментатором в классе, дети ежедневно подвергаются многим другим, потенциально имеющим отношение к вопросу войны и мира. Конечно, даже при таком изобилии неэкспериментальных воздействий, которые могут оказывать влияние на результаты опыта, возможна их правдоподобная интерпретация, оправдывающая постановку данного эксперимента. Как отмечалось выше, фактор *фона* порождает условия, противоположные тем, которые в физической или биологической лаборатории были бы названы *экспериментальной изоляцией*. Вероятность *фоновых воздействий* в качестве источника наблюдаемого сдвига, вроде того, что мы находим на рис. 3 (кривые А и Б), в значительной мере зависит от степени экспериментальной изоляции, которую в состоянии обеспечить экспериментатор. Павловские исследования условных рефлексов у собак, по существу являющиеся экспериментами «на одной группе» или «на одном животном», не служили бы в такой мере подтверждению его теорий, если бы опыты велись не в звуконепроницаемой лаборатории, а на оживленном перекрестке. Что конкретно понимать под экспериментальной изоляцией, зависит от изучаемой проблемы и от характера применяемой измерительной процедуры. Обеспечение экспериментальной изоляции при исследовании элементарных частиц в камере Вильсона или счетчиками сцинтилляций требует принятия более серьезных мер предосторожности, чем в воображаемом эксперименте с железным бруском, погружаемым в азотную кислоту. Во многих ситуациях, в которых применим план 7, экспериментатор мог бы с большой степенью уверенности говорить о наличии экспериментальной изоляции в том смысле, что он знал, какие конкурентные события

могли бы обусловить наблюдавшийся эффект, и сумел существенно снизить вероятность их действия.

К *фону* можно из соображений удобства отнести влияние погоды и времени года. Так, при исследовании производительности труда рабочих может иметь место смешивание экспериментального фактора с сезонными изменениями в освещенности, погоде и т. д. Поэтому соответствующие эксперименты обычно проводятся в различное время года. Пожалуй, легче всего контролировать действие циклических факторов *фона*, связанных с институциональными обычаями в жизни группы (недельные циклы труда, выдачи зарплаты, периоды экзаменов, каникулы, школьные праздники), вызывающих периодические изменения зависимой переменной. В известном смысле эти факторы близки фактору *естественного развития*. При проведении серии наблюдений нужно стремиться к тому, чтобы известные циклы оставались постоянными или серия по длительности охватывала несколько таких циклов.

Продолжим анализ факторов, которые подлежат контролю. *Естественное развитие*, когда результаты наблюдений отвечают кривым А и Б на рис. 3, не может служить правдоподобным объяснением скачка, имевшего место между  $O_4$  и  $O_5$ , но отсутствовавшего в другие периоды времени. (Однако естественное развитие не всегда бывает гладким, регулярным. Так, внезапное наступление первой менструации у школьниц способно сместить значения физиологических показателей, создавая в случае плана 7 ложный экспериментальный эффект.) Точно так же не является допустимой конкурентной

гипотезой для скачка между  $O_4$  и  $O_5$  и влияние тестирования. Но, располагая, как в эксперименте по плану 2, только этими наблюдениями (в точках  $O_4$  и  $O_5$ ), мы не можем считать неправдоподобными эффекты естественного развития и тестирования. В этом большое преимущество данного плана перед планом 2.

Аналогично для плана 7, в отличие от рассмотренных ранее планов, возможные ссылки на *инструментальную погрешность* будут лишены конкретных оснований, поскольку нет причин ожидать, что ошибка прибора будет иметь место именно в этом замере. Однако вопро-

120

сительный знак в табл. 2 требует обратить внимание на ситуации, в которых изменение калибровки средства измерения может быть неправильно истолковано как эффект  $X$ . Если в измерительной процедуре используются оценки, даваемые лицами, знакомыми с замыслом эксперимента, то может иметь место ложное подтверждение гипотезы из-за того, что испытуемые представляют себе, каковы должны быть ожидаемые результаты. Так, экспериментальное воздействие в виде назначения нового директора школы может повлиять на регистрацию нарушений школьной дисциплины, а не на число самих нарушений. При использовании плана 7 для определения эффекта значительных изменений в административной политике необходимо избегать одновременного перехода к новой процедуре измерения. В большинстве случаев, чтобы сохранить возможность интерпретировать результаты серии замеров, лучше бывает продолжать использование устаревшей процедуры, а не переходить к новой.

Эффект *регрессии* обычно выражается функцией с отрицательной второй производной по времени (negatively accelerated function) и поэтому не даст правдоподобного объяснения большего эффекта  $O_5$ , чем в точке  $O_2$ ,  $O_3$  и  $O_4$ . Фактор *состава экспериментальной группы* как источник главного эффекта отпадает так же, как и для эксперимента типа 2, если все замеры производятся на одних и тех же лицах. Если данные по группе регистрируются в основном в индивидуальном порядке, то, как и в плане 2, отпадает фактор выбывания из групп. Однако если наблюдения относятся к группе в целом, то необходимо параллельно регистрировать изменения в составе группы, следя за тем, чтобы совпадения в изменении личного состава не давали оснований для конкурентной гипотезы.

Перейдем теперь к внешней валидности. Ясно, что экспериментальный эффект может быть характерен только для тех членов популяции, которые подвергаются повторной проверке. Это вряд ли составляет ограничение исследований в области школьного обучения, если эксперимент не включает процедуры тестирования ( $O$ ), необычные для школьной обстановки. Более того, эксперимент данного типа обычно подходит для тех учреждений, в которых сбор и регистрация данных ве-

121

дутся постоянно и составляют естественную часть окружения. Годовые проверки уровня знаний в средних школах, отметки о болезни и т. д. обычно не сопряжены с реакцией на эксперимент в том смысле, что они типичны для генеральной совокупности, на которую хотят распространить сделанные выводы. Взаимодействие между *составом группы* (отбором испытуемых) и  $X$  означало бы, что полученный эффект является особенностью данной выборки и что он не типичен для изучаемой генеральной совокупности, относительно которой сформированная естественным путем экспериментальная группа была бы смещенной выборкой. Так, специфика данных может вынудить исследователя ограничиться результатами только тех учеников, которые в течение долгого времени не пропустили ни одного обследования, то есть явно отобранного подмножества. Кроме того,

если использовались новые измерительные процедуры  $O$ , то повторные случаи этого могли провоцировать уклонение от обследований.

Если такие временные серии интерпретируются как эксперименты, важно, чтобы экспериментатор заранее уточнил, сколько времени должно пройти между введением экспериментальной переменной и проявлением ее эффекта. При наличии этих данных форма кривой  $\Gamma$  на рис. 3 будет почти столь же определенно указывать на наличие эффекта, как и кривая  $A$ . Массовые обследования, в которых решения основываются на интерпретациях отсроченных эффектов, нуждаются в перекрестной валидации. При увеличении времени отсрочки растет также вероятность появления эффектов, вызванных фактором фона.

Столь же важно, чтобы исследователь определил  $X$  еще до того, как он приступит к анализу временной серии. Попытки установить *post hoc*, какое  $X$  предшествовало наиболее значительному сдвигу, следует исключить на том основании, что допускаемый при этом произвол затрудняет, а то и делает вовсе невозможным проверку значимости эффектов.

Преобладание экспериментов рассматриваемого типа в более успешных науках должно внушать определенное уважение к нему. Однако нам нужно помнить, что благодаря «экспериментальной изоляции» и «постоянству условий» там имеется больше возможностей для

интерпретации результатов. Следует также иметь в виду, что там никогда не делается вывод по данным одного эксперимента. План 7 воспроизводится в самых различных условиях разными исследователями, прежде чем устанавливается закон, хотя они могут никогда не обращаться к контрольным группам. Мы также должны следовать этому правилу. *Если лучший контроль невозможен*, мы будем пользоваться этим планом эксперимента. Нам нужно так организовать дело, чтобы получать как можно больше временных серий, и постараться более детально, чем раньше, изучить влияние административных изменений и других внешних внезапных и случайных событий в качестве  $X$ . Но результаты не будут считаться достоверными, пока они не будут многократно воспроизведены в различных условиях<sup>1</sup>.

### **8. План с эквивалентными временными выборками**

Эксперимент в его наиболее типичной форме предполагает использование эквивалентной выборки испытуемых, чтобы получить «базовую линию», с которой сравнивается действие экспериментальной переменной. В отличие от этого рекуррентная форма эксперимента с одной группой использует две эквивалентные выборки ситуаций, в одной из которых экспериментальное воздействие имеет место, а в другой — нет. Схема такого эксперимента имеет следующий вид (хотя предпочтительнее случайное, а не регулярное чередование):

$$X_1O \quad X_0O \quad X_1O \quad X_0O.$$

Этот эксперимент можно рассматривать как разновидность временной серии с повторным введением экспериментального воздействия. Достоинства такого экспери-

---

<sup>1</sup> В последнее время было разработано много приемов, которые могут быть использованы для статистической обработки результатов экспериментов по этому плану. См., например: Box G. E. P., Tiao G. C. *Intervention analysis with applications to economic and environmental problems*. — «Journal of American Statistical Association» 1975, v. 70, p. 70—92; Box G. E. P., Tiao G. C. *A change in level of a non-stationary time series*. «Biometrika», 1965, v. 52, p. 181—192.

мента наиболее очевидны, когда ожидается, что эффект воздействия будет носить преходящий или обратимый характер. Хотя этот эксперимент может трактоваться как расширенный вариант временной серии, методика статистического анализа более сходна с обработкой результатов эксперимента, использующего две группы, в котором применяется критерий значимости различий между средними двух наборов измерений. Обычно изменения точно совпадают с экспериментальными воздействиями, часто сопутствуя им, как в исследованиях научения, производительности труда, выработки условных связей, физиологических реакций и т. д. Среди ранних примеров применения этого типа эксперимента, пожалуй, самыми типичными были исследования Олпорта [1] и Сорокина [105], в которых изучалась эффективность деятельности учащихся в различных условиях, причем сравнивалось не наличие и отсутствие экспериментального воздействия, а два различных экспериментальных воздействия, то есть  $X_1$  и  $X_2$ . В большинстве случаев простое чередование условий и использование постоянных промежутков времени нежелательны, особенно если имеется опасность смешивания эффекта  $X$  с влиянием дневных, недельных или месячных циклов или когда очевидная периодичность приводит к выработке нежелательной реакции на временной интервал, что в свою очередь может усилить различие между двумя воздействиями. Так, Сорокин позаботился о том, чтобы каждое из экспериментальных воздействий имело место с одинаковой частотой в утреннее и послеобеденное время.

В большинстве экспериментов, проведенных до сих пор по данному плану, каждое экспериментальное воздействие повторялось сравнительно небольшое число раз. Однако тип обобщения теории выборочного исследования, представленный в работе Брунсвика [9], заставляет обратить внимание на необходимость применения больших, репрезентативных и эквивалентных случайных выборок временных интервалов. Керр [63], по-видимому, очень близко подошел к этому идеалу в своих исследованиях влияния музыки на производительность труда в промышленности. В каждой серии, его опытов использовалась единственная экспериментальная группа и применялись рандомизированные выборки

дней в течение каждого месячного периода. Так, в одном эксперименте он сравнил 56 «музыкальных» дней с 51 днем без музыки, а в другом — влияние трех различных видов музыкальных произведений, каждый из которых был использован в эквивалентных выборках по 14 дней.

В том виде, в каком план 8 применялся Керром, он в целом представляется внутренне валидным. Фактор *фона*, самое слабое место временной серии, контролируется благодаря предъявлению  $X$  в большом числе отдельных случаев, что делает крайне маловероятным любое конкурентное объяснение результатов за счет одновременного действия побочных явлений. Другие источники невалидности находятся под контролем в силу тех же соображений, которые были подробно приведены для плана 7. Что касается внешней валидности, то обобщение возможно только по отношению к часто обследовавшимся популяциям. Реакция испытуемых на эксперимент, осведомленность об экспериментировании — особенно уязвимое место данного плана. Если отдельные воздействия  $X$  применяются последовательно к различным группам, то испытуемые (особенно в плане 6) могут даже не подозревать, что проводится опыт и сравниваются различные воздействия. Иное дело, если исследуется одна-единственная группа, для которой повторно создаются те или иные условия — например, применяется то один, то другой способ расчета зарплаты, как в опытах Сорокина; или то одни, то другие условия работы, как у Олпорта; или то один, то другой вид вентиляции (Wyatt, Fraser, Stock [138]; или то один, то другой

музыкальный жанр (хотя Керр принял меры предосторожности, чтобы разнородная программа стала естественной составной частью рабочего окружения). Что касается взаимодействия *состава групп* с  $X$ , то здесь, как и обычно, существуют ограничения в распространении обнаруженного эффекта на всю популяцию.

В плане 8 кроется риск для внешней валидности. Его можно обнаружить во всех тех экспериментах, упоминаемых в данной работе, в которых на одной и той же группе испытывается действие нескольких уровней  $X$ . Этот эффект был обозначен как «межуровневая интерференция  $X$ ». Эффект  $X_1$ , который в простей-

125

шем случае сравнивается с действием  $X_0$ , может быть распространен только на условия повторного и разделённого промежутками времени предъявления  $X_1$ . Нет оснований распространять его на возможные ситуации непрерывного действия  $X_1$  или на случаи, когда  $X_1$  появляется один-единственный раз. Кроме того, условие  $X_0$ , то есть отсутствие  $X$ , не типично для периодов отсутствия  $X$  вообще. Оно репрезентативно только для отсутствия  $X$ , перемежающегося с его наличием. Если  $X_1$  обладает некоторым последствием, распространяющимся на периоды без  $X$ , как это обычно представляется вероятным, применение плана 8 может привести к недооценке эффекта  $X_1$  в сравнении, скажем, с планом 6. Кроме того, сам факт частых изменений может увеличить значение  $X$  как стимула по сравнению с тем, что было бы при непрерывном, гомогенном воздействии. Гавайская музыка в опытах Керра могла бы совсем по-иному влиять на работу испытуемых, если бы она не перемежалась с другими жанрами, а составляла постоянный «музыкальный рацион». Эксперименты Эббингауза [32], очевидно, в основном принадлежат к этому типу, и, как отмечал Андервуд [128], установленные им законы распространяются только на тех, кто усвоил десятки других весьма сходных списков. Многие из его результатов фактически недействительны для людей, запоминающих единственный список бессмысленных слогов. Таким образом, хотя этот план эксперимента внутренне валиден, его внешняя валидность значительно ограничена некоторыми типами содержания (см. также Kempthorne [58, гл. 29]).

Отметим, однако, что действие многих факторов школьного обучения, представляющих интерес для экспериментального исследования, во всех практически важных отношениях ограничено временем фактического присутствия  $X$ . В таких случаях данная схема может оказаться весьма ценной. Предположим, что учитель стремится выяснить значение чтения вслух по сравнению с индивидуальным чтением про себя. Применяя в серии уроков то один, то другой способ, он может провести эксперимент, результаты которого поддаются интерпретации. Этим путем можно исследовать и эффект присутствия кого-либо из родителей в классе во время добровольной дискуссии учеников. Знание такого плана

126

сделает экспериментальную проверку альтернатив доступной для учителя. Это позволит проводить пробные испытания методик, которые в случае положительного исхода могут быть исследованы посредством более обширных и более скоординированных экспериментов.

Этот подход может быть применен в отношении выборок ситуаций при исследовании одного испытуемого. Примером может служить рекуррентный план эксперимента, реализуемый (обычно без проверки статистической значимости) в физиологических исследованиях, в которых животному повторно предъявляют стимулы, стремясь избежать какой бы то ни было периодичности стимуляции, что соответствует требованию рандомизации, следующему из логики эксперимента. Вместо простой рандомизации могут также быть использованы латинские квадраты (см., например, Cox [26]; Maxwell [73]).

## Проверка значимости для плана 8

Для данного конкретного типа эксперимента нам опять необходима соответствующая процедура проверки статистической значимости экспериментального эффекта. Отметим, что предполагается два типа обобщения: по ситуациям и по испытуемым. В случае если используется лишь один испытуемый, проверка значимости эффекта, очевидно, будет ограничена обобщениями в отношении только данного индивида и будет предполагать обобщение ситуаций (реализации  $X$ ). Для этой цели подходит  $t$ -критерий с числом степеней свободы, равным числу ситуаций (реализаций) минус 2. При наличии индивидуальных результатов по ряду испытуемых, которым как членам одной группы предъявлялось одно и то же экспериментальное воздействие, данные позволяют осуществить обобщение в отношении испытуемых. В этой ситуации обычно применяются две стратегии. Ошибочная стратегия состоит в получении для каждого испытуемого по одному численному результату при каждом воздействии и затем в проверке значимости различий между средними, при использовании коррелированных данных. Этой логике следовали Олпорт и Сорокин, хотя они фактически не прибегали к проверке значимости эффекта. Но если каждая экспериментальная ситуация воспроизводится всего один или

127

два раза, ошибка выборки по ситуациям может быть весьма значительной или контроль фоновой стимуляции может оказаться весьма слабым. И то, что при таком способе анализа будет представляться значимым различием, в какой-то мере может быть вызвано случайными ошибками выборки. Такая ошибка станет весьма серьезной, если велик и статистически значим обучающий эффект тестирования. Так, согласно этой логике можно получить весьма значимые различия между  $X_1$  и  $X_2$ , когда каждое из этих воздействий применялось только однажды и когда одно из них совпало с появлением побочной стимуляции, оказавшей заметное влияние на результат. Поэтому важно, чтобы каждое воздействие ( $X_1$  и  $X_2$ ) имело по крайней мере две реализации и тем самым были представлены степени свободы случаев (ситуаций), сгруппированных по данному воздействию. Это требование, вероятно, легче всего удовлетворить путем первоначальной проверки значимости различия между средними для обоих воздействий, сравнивая ее с величиной ошибки реализации каждого воздействия. Установив таким путем значимость эффекта экспериментального воздействия, можно определить, для какой части испытуемых он имеет место, и тем самым получить сведения о воспроизводимости (обобщаемости) эффекта на разных испытуемых. Повторные измерения и использование выборок ситуаций ставят на повестку дня многие статистические проблемы, часть которых все еще не решена (Collier [24], Cox [26], Kempthorne [58]).

## 9. План с сериями эквивалентных воздействий

С планом эквивалентных временных выборок тесно связан план 9, основанный на эквивалентности выборок материала, которые используются в сравниваемых режимах эксперимента. При этом всегда или почти всегда в эксперименте используются и эквивалентные временные выборки, но они могут быть так тонко и запутанно переплетены друг с другом, что весь эксперимент практически однороден во времени. Применение эквивалентного материала при повторной реализации  $X$  в од-

128

ногрупповом эксперименте требуется в том случае, когда эффект воздействия оказывается стойким, вследствие чего различные воздействия и их повторные реализации должны осуществляться с использованием неидентичного материала. Схема такого эксперимента выглядит следующим образом:

$$M_a X_1 O \quad M_b X_0 O \quad M_c X_1 O \quad M_d X_0 O \quad \text{и т. д.}$$

Буква *M* обозначает здесь конкретный материал, причем выборка *M<sub>a</sub>*, *M<sub>c</sub>* и т. д. с точки зрения выборочного анализа эквивалентна выборке *M<sub>a</sub>*, *M<sub>b</sub>* и т. д. Важность выборочной эквивалентности двух комплектов материала более отчетливо выступает при таком изображении плана эксперимента:

$$\begin{array}{l} \text{Один испытуемый} \\ \text{или группа} \end{array} \left\{ \begin{array}{l} \text{Выборка материала} \quad A(O)X_0O \\ \text{Выборка материала} \quad B(O)X_1O. \end{array} \right.$$

Буква *O* в скобках обозначает наличие предварительного тестирования в некоторых вариантах плана.

Отличной иллюстрацией служат ранние эксперименты Йоста [55], сравнивавшего запоминание в режимах концентрированного (*K*) и распределенного (*P*) заучивания. Для третьего эксперимента были более или менее случайным образом составлены 12 списков из 12 бессмысленных слогов в каждом. Шесть списков предъявлялось в режиме *K* и шесть в режиме *P*. Все 12 списков заучивались одновременно в течение семи дней, причем в расписании опыта они всячески перемежались друг с другом, что позволяло контролировать утомление и другие посторонние факторы. Всего в период с 6 ноября 1895 г. по 7 апреля 1896 г. было использовано 7 наборов по 12 списков, так что в конце эксперимента Йост располагал результатами по 40 спискам слогов, заучивавшимся в режиме *K*, и по 40 спискам, заучивавшимся в режиме *P*. Возможность интерпретации различий, обнаруженных на одном испытуемом, профессоре Г. Э. Мюллере, зависит от выборочной эквивалентности применявшихся неидентичных списков. При этих ограничениях эксперимент, по-видимому, обладает внутренней валидностью. Результаты, конечно, относятся лишь

к психологии профессора Г. Э. Мюллера в 1895—1896 гг. и к совокупности мнемического материала, из которого составлялись выборки. Для того чтобы можно было делать выводы относительно свойств человеческой памяти вообще, необходимо было, конечно, повторить этот эксперимент со многими другими испытуемыми.

Другой пример взят из ранних исследований конформизма. Так, Мур [81] получил «контрольную» оценку стабильности ответов на вопросник при повторном предъявлении одной серии вопросов, а затем сравнил с ней изменения, возникшие при повторном предъявлении другой серии вопросов, во время которого испытуемым сообщалось мнение большинства. Рассмотрим исследование, в котором учащихся просят изложить свои мнения по длинному списку вопросов. Этот список затем подразделяется на две по возможности эквивалентные части. Затем вопросники, в которых по каждому вопросу указано «мнение большинства», возвращаются испытуемым. «Групповые суждения» сфабрикованы так, чтобы для двух выборок вопросов они были противоположны друг другу. В заключение ученики снова отвечают на все вопросы списка. Поскольку обе половины вопросника эквивалентны друг другу, различия в смещении между двумя экспериментальными воздействиями служили бы определенной экспериментальной

демонстрацией влияния сообщаемого мнения большинства даже в отсутствие какой-либо контрольной группы испытуемых.

План 9, как и план 8, обладает внутренней валидностью по всем пунктам, причем, в общем, по тем же причинам. Что же касается внешней валидности, то для плана 9, как и для всех экспериментов с повторными измерениями, результаты нередко имеют силу только в отношении данных испытуемых. В экспериментах по научению измерения обычно до такой степени составляют часть обстановки опыта (это не обязательно относится к методу Йоста, предполагавшему заучивание списка путем многократного его чтения), что подобное ограничение обобщения становится несущественным. План 9, по-видимому, вызывает меньшую реакцию испытуемых на эксперимент, чем план 8, ввиду разнородности материала и большей возможности скрыть от

130

испытуемых применение различных воздействий в разное время по разным пунктам. Такая низкая реактивность была бы обнаружена в исследованиях конформизма, но не в экспериментах Йоста. Интерференция между уровнями экспериментальной переменной или между различными материалами является, вероятно, наиболее уязвимым местом данного типа эксперимента, как и в случае плана 8.

На конкретном примере опытов Йоста виден характер ограничений, связанных с обобщением результатов. Сам он считал, что распределенное заучивание эффективнее концентрированного. Мы видим, что по условиям эксперимента он имел право распространять этот вывод только на тех лиц, кому приходится заучивать много списков, то есть на случаи высокого общего уровня интерференции. Современные же исследования показывают, что превосходство распределенного заучивания ограничивается именно этими ситуациями и что в случаях заучивания совершенно нового материала данный способ не обладает никакими преимуществами (Underwood, Richardson [130]).

#### Статистика для плана 9

Вопросы валидности и доказательности эксперимента, очевидно, возникают и в связи с применением выборок экспериментального материала. Число  $N$  при вычислении значимости различий между средними по группам, которым предъявлялось воздействие, должно было быть равным числу списков в эксперименте Йоста (или числу вопросов анкеты при изучении конформизма), так, чтобы оно представляло соответствующее пространство выборки. К этому необходимо добавить основание для обобщения по испытуемым. Вероятно, лучшим в настоящее время является поэтапный подход: сначала производится обобщение по выборке списков или вопросов, после чего для каждого испытуемого вычисляется величина экспериментального эффекта и полученные значения используются как основа для обобщения по испытуемым. (Связанные с этим меры предосторожности излагаются в литературе, указанной при обсуждении плана 8.)

131

### 10. План с неэквивалентной контрольной группой

В педагогических исследованиях наибольшее распространение получил эксперимент, в котором экспериментальная и контрольная группы подвергаются тестированию до и после изучаемого воздействия, причем обе группы не эквивалентны друг другу до эксперимента. Скорее, они представляют собой естественно сложившиеся коллективы (например, школьные классы), по возможности сходные, но не настолько, чтобы различие между ними

не обнаруживалось во время предварительного тестирования. Считается, что выбор, к какой из групп должно быть применено воздействие  $X$ , является случайным и подлежит контролю экспериментатора.

$$\begin{array}{ccc} O & X & O \\ \hline O & & O \end{array} .$$

В отношении такого плана экспериментов нужно ясно представлять себе две вещи. Во-первых, его не следует смешивать с планом 4, в котором испытуемые из общей популяции распределяются по обеим группам *в случайном порядке*. Во-вторых, данный план эксперимента все же стоит применять во многих случаях, когда использование планов 4, 5 или 6 невозможно. В частности, следует признать, что добавление даже неуравненной или неэквивалентной контрольной группы значительно снижает неоднозначность интерпретации результатов по сравнению с тем, что имеет место при плане 2. Чем больше сходства в составлении экспериментальной и контрольной групп, чем больше их сходство подтверждается результатами предварительного тестирования, тем эффективнее становится это средство контроля. Если это требование внутренней валидности в какой-то степени выполнено, то можно считать, что в данном типе эксперимента контролируются главные эффекты фоновой стимуляции, естественного развития, тестирования и инструментальной погрешности в том смысле, что различия в результатах тестирования в экспериментальной группе до и после введения  $X$  (если они выше, чем те же различия для контрольной группы) нельзя объяснить за счет главных эффектов этих пере-

менных, оказывающих влияние, как на экспериментальную, так и на контрольную группу. (Следует, однако, серьезно отнестись к замечанию, сделанному в связи с планом 4 по поводу посторонних воздействий в промежутках между сериями эксперимента.)

Попытки объяснить наблюдавшийся сдвиг в результатах экспериментальной группы за счет таких побочных факторов, как фоновая стимуляция, естественное развитие или влияние тестирования, должны предполагать взаимодействие между этими переменными и конкретными различиями в составе двух групп. Хотя, вообще говоря, такие взаимодействия и маловероятны, существует ряд ситуаций, в которых подобное предположение допустимо. Пожалуй, чаще всего имеют место взаимодействия с фактором *естественного развития*. Если экспериментальная группа состоит из пациентов, проходящих курс психотерапии, а роль контрольной группы играет какая-либо другая доступная совокупность испытуемых, то смещение показателей экспериментальной группы можно интерпретировать как процесс спонтанной ремиссии, являющейся особенностью именно данной необычной группы и возможной даже без  $X$ . Такое взаимодействие факторов состава групп с естественным развитием (или с фоновыми событиями, или с эффектом тестирования) может быть ошибочно принято за эффект  $X$  и, следовательно, составляет угрозу для *внутренней* валидности эксперимента. По этому критерию (см. 8-ю колонку табл. 2) план 10 отличается от плана 4.

Поясним это на конкретном примере педагогического исследования. К типу 10 принадлежит проведенное Санфордом и Хемфиллом [97] изучение влияния преподавания психологии в университете города Аннаполис, Экспериментальная группа состояла из второкурсников, а контрольная — из студентов третьего курса. Более высокие показатели изменений, зарегистрированные в экспериментальной группе, могли объясняться отчасти за счет тех сложных процессов, которые происходят скорее в течение первых двух лет обучения, чем на третьем и четвертом курсах, что было бы проявлением различия во

взаимодействиях между факторами состава групп и естественного развития, а не за счет экспериментальной программы. Примененная Санфордом и

133

Хемфиллом контрольная группа допускает проверку данной конкурентной интерпретации (в духе рассматриваемого ниже плана 15). На основе гипотезы о взаимодействии факторов состава группы и естественного развития можно было бы предсказать, что превосходство третьекурсников (контрольной группы) над второкурсниками при первоначальном испытании окажется примерно того же порядка, что и разница в результатах предварительного и итогового тестирования испытуемых экспериментальной группы. К счастью, в целом это было не так. Направление и величина различий между курсами при предварительном тестировании в большинстве случаев были иными, чем различия в результатах предварительного и итогового тестирования в экспериментальной группе. Однако обнаруженный авторами значимый прирост показателей доверия в экспериментальной группе, по данным анкеты социальных ситуаций, мог оказаться артефактом взаимодействия факторов состава группы и естественного развития. Для экспериментальной группы этот показатель увеличился с 43,26 до 51,42, тогда как для третьего курса эти же цифры соответственно равны 55,82 и 56,78.

Гипотеза о взаимодействии между факторами состава группы и естественного развития иногда может оказаться пригодной, даже если группы имеют идентичные показатели по данным исходного теста. Наиболее типичным является случай, когда естественное развитие (или независимое изменение) в одной группе протекает скорее, чем в другой. План 14, являющийся усовершенствованным вариантом плана 10, позволяет контролировать это взаимодействие.

Эффект регрессии составляет другую важную проблему с точки зрения внутренней валидности экспериментов типа 10. Вопросительный знак в табл. 2 указывает, что проблемы можно избежать, хотя чаще она создается самим исследователем. Вообще говоря, если каждая из сравниваемых групп составлена по признаку максимальных и минимальных результатов или коррелированных с ними величин, то различная степень изменения показателей от предварительного тестирования к тестированию после воздействия может быть проявлением скорее эффекта регрессии, чем  $X$ . Такая возмож-

134

ность становится все более характерной для исследований в области педагогики, где попарное уравнивание в силу устойчивой и дезориентирующей традиции считается адекватной и достаточной процедурой установления доэкспериментальной эквивалентности групп. Эта ошибка сопровождается непониманием того, что между планами 4 и 10 существует серьезное различие и что попарное уравнивание по исходным показателям играет в этих случаях разную роль. В плане 4 попарное уравнивание может служить полезным добавлением к рандомизации, но не заменять ее. Вся популяция испытуемых, привлекаемых к эксперименту, разбивается на пары, тщательно уравненные по показателям предварительного тестирования или по связанным переменным. Затем члены каждой пары *случайным образом* включаются в экспериментальную или контрольную группу. Такое уравнивание с последующей рандомизацией обычно дает план эксперимента, обеспечивающий большую точность результатов, чем одна рандомизация.

С этим идеалом нельзя смешивать предпринимаемые в рамках плана 10 попытки компенсировать различия между неэквивалентными группами путем попарного уравнивания, когда рандомизированное распределение по группам невозможно. Если в плане 10 средние групп существенно отличаются друг от друга, то уравнивание пар не только не обеспечивает искомого равенства, но еще и вызывает появление нежелательного

эффекта регрессии. Можно с уверенностью предсказать, что обе группы будут отличаться по результатам итогового тестирования независимо от наличия эффекта  $X$  и что различие будет прямо пропорционально различию между популяциями, из которых набирались испытуемые, и обратно пропорционально тест-ретестной (test-retest) корреляции<sup>1</sup>.

Можно также предсказать направление эффекта регрессии. Рассмотрим психотерапевтический экспери-

---

<sup>1</sup> Некоторые специальные данные о применении в этой связи ковариационного анализа и об опасности простых ковариаций содержатся в более поздней работе автора, написанной совместно с Боруком [13]. Особое внимание в ней уделяется надежности скорректированной ковариации и множественной регрессии. В этой работе рассматриваются также условия, при которых корреляция предварительных и последующих тестирований более соответствует принятым коэффициентам надежности. — *Прим. ред.*

мент, в котором в качестве  $O$  используется балльная оценка степени неудовлетворенности собой. Предположим, экспериментальная группа состоит из обратившихся за помощью к психотерапевту, а уравненная с ней контрольная группа — из «нормальных» людей. Тогда контрольная группа будет представлена крайне низкими показателями нормальной группы (образованной на основе этих крайних показателей) и при итоговом тестировании будет отмечена регрессия этих данных к средним для нормальной группы. Менее вероятно, что это приведет к выявлению значимого эффекта терапевтического воздействия, скорее, создаст ложное впечатление действительности терапевтической процедуры. Пример с психотерапевтической группой иллюстрирует также возможную несостоятельность предположений об однородности регрессии и о принадлежности выборок к одной и той же генеральной совокупности (за исключением случаев, когда используются лишь экстремальные показатели). Контрольные группы здоровых людей находят применение в психотерапевтических исследованиях, но при интерпретации результатов таких исследований необходимо соблюдать крайнюю осторожность.

Важно различать две разновидности плана 10 и определять их различный статус в качестве аппроксимаций истинного эксперимента. С одной стороны, в некоторых случаях экспериментатор располагает двумя естественными группами (например, двумя классами) и может свободно решать, какая из них подвергнется  $X$ , или, по крайней мере, у него нет оснований подозревать, что отбор испытуемых в группы, подвергаемые  $X$ , осуществляется особым образом. Даже при различии исходных средних по  $O$  данное исследование может приближаться к истинному эксперименту. С другой стороны, в некоторых случаях применения плана 10 испытуемые экспериментальной группы сами изъявляют желание подвергнуться  $X$ , но нет контрольной группы, которая состояла бы из лиц, выразивших желание участвовать именно в контрольной группе. В этом случае посылка об однородности регрессии между экспериментальной и контрольной группами становится менее достоверной и повышается вероятность взаимодействий между факторами состава группы и естественного развития (а также других взаимодействий с фактором состава группы).

План 10 с участием добровольцев намного слабее, но и он дает информацию, которая во многих случаях позволит отклонить гипотезу об эффективности  $X$ . Контрольная группа, даже если она значительно отличается от экспериментальной по способу подбора и по среднему уровню, все же помогает интерпретации данных.

Угроза внешней валидности со стороны эффекта тестирования та же, что и для плана 4. Знак вопроса против взаимодействия фактора состава группы и  $X$  напоминает нам, что действие  $X$  может быть специфичным для лиц, отобранных по принципу, который приме-

нялся при отборе испытуемых в данном эксперименте. Но поскольку план 10 меньше, чем план 4, ограничивает нашу свободу в выборе состава групп, эта особенность будет сказываться в меньшей степени, чем в лабораторных условиях.

Опасность реакции на эксперимент имеет место, но, видимо, в меньшей степени, чем для большинства истинных экспериментов вроде плана 4. Выбирая между планом 10 при сохранении двух классов в неизменном виде и планами 4, 5 или 6, согласно которым для различных экспериментальных воздействий используются случайные выборки учеников из разных классов, нужно учитывать, что вероятность реакции на эксперимент во втором случае будет почти наверняка большей, потому что ученики будут больше осведомлены об эксперименте и будут чувствовать себя подопытными кроликами и т. д.

Исследования Торндайка в области формального обучения и переноса (например, E. L. Thorndike, Woodworth [123]; Brolyer, Thorndike, Woodyard [7]) представляют собой пример применения плана 10 в случае, когда  $X$  не контролируется экспериментатором. Эти исследования, по крайней мере, отчасти избежали ошибки эффекта регрессии, связанной с простым уравниванием групп, но их нужно тщательно проанализировать в свете современных методов. Так, использование ковариационного анализа дало бы, вероятно, сильные доводы в пользу эффекта переноса слов из латинского языка в английский.

Кроме того, наблюдаемый обычно положительный, хотя и слабовыраженный эффект переноса мог бы быть объяснен за счет отбора на курс латыни тех учеников, у которых рост словарного запаса происходил быстрее,

чем в контрольной группе, даже без обучения латинскому языку. В наших терминах это следовало бы назвать взаимодействием состава группы и естественного развития. Во многих школьных системах эта конкурентная гипотеза могла бы быть проверена путем увеличения числа процедур предварительных тестирований перед курсом латыни, как это предусмотрено планом 14. Эти эксперименты представляли собой грандиозные усилия по внедрению экспериментального мышления в полевое исследование. Они заслуживают того, чтобы на них снова обратили внимание и продолжили их современными методами.

## 11. Сбалансированные планы

Под этой рубрикой объединены все те планы, в которых для достижения контроля экспериментальных параметров или повышения точности результатов предусматривается предъявление всем испытуемым (или использование во всех ситуациях) всех экспериментальных воздействий. Такие планы обозначались как «ротационные эксперименты» (McCall [74]), «сбалансированные планы» (например, Underwood [126]), «перекрестные планы» (например, Cochran, Cox [22], Cox [28]), «планы с переключением» (Kempthorne [58]). Для построения сбалансированного плана обычно используется латинский квадрат. Такой латинский квадрат в качестве плана квазиэксперимента представлен в нижеследующей схеме, в которой четыре экспериментальных воздействия в квазислучайном порядке последовательно применяются к четырем естественным образом составленным группам или даже к четырем испытуемым (например, Maxwell [73]):

Группа испытуемых	Порядковый номер воздействия (серии экспериментов) ( $t$ )			
	1	2	3	4
<i>A</i>	$X_1O$	$X_2O$	$X_3O$	$X_4O$
<i>B</i>	$X_2O$	$X_4O$	$X_1O$	$X_3O$
<i>C</i>	$X_3O$	$X_1O$	$X_4O$	$X_2O$
<i>D</i>	$X_4O$	$X_3O$	$X_2O$	$X_1O$

138

В план включены только последующие тестирования, поскольку он находит применение особенно в тех случаях, когда предварительное тестирование неосуществимо и невозможно воспользоваться планами, подобными плану 10. План включает 3 переменные [группы ( $g$ ), порядковый номер воздействия ( $t$ ) и экспериментальные воздействия ( $X$ )]. Каждая переменная «ортогональна» двум другим в том смысле, что каждый уровень одной переменной одинаково часто (один раз для латинского квадрата) сочетается с каждым уровнем любой другой. Легко видеть, что каждое воздействие ( $X$ ) фигурирует в каждой строке и в каждом столбце по одному, и только одному, разу. Тот же латинский квадрат можно переписать так, чтобы столбцы были разобраны по  $X$ :

Группа испытуемых ( $g$ )	Экспериментальные воздействия ( $X$ )			
	$X_1$	$X_2$	$X_3$	$X_4$
<i>A</i>	$t_1O$	$t_2O$	$t_3O$	$t_4O$
<i>B</i>	$t_3O$	$t_1O$	$t_4O$	$t_2O$
<i>C</i>	$t_2O$	$t_4O$	$t_1O$	$t_3O$
<i>D</i>	$t_4O$	$t_3O$	$t_2O$	$t_1O$

Таким образом, суммы результатов измерений по  $X$  сравнимы друг с другом, так как в каждой сумме представлены все группы ( $g$ ) и все серии экспериментов ( $t$ ). Различия между этими суммами нельзя считать просто артефактами первоначальных межгрупповых различий, эффекта научения, фона и т. д. Аналогично можно сравнивать суммы по строкам, отражающим различия между группами, а также суммы по столбцам данных первого предъявления в различных ситуациях. В терминах дисперсионного анализа можно сказать, что рассматриваемый план позволяет получать данные по трем главным эффектам при числе ячеек исходной таблицы, отвечающем полной двухфакторной схеме. Ясно также, какую цену приходится платить за повышенную эффективность: то, что кажется значимым главным эффектом какой-либо из трех переменных, может на самом деле

быть значимым взаимодействием двух других переменных (Lindquist [68, с. 258—264]). Кажущееся различие между эффектами  $X$ -ов, в частности, может оказаться специфическим эффектом комплексного взаимодействия между групповыми различиями и ситуациями (сериями экспериментов). Вывод относительно эффектов  $X$  будет зависеть от правдоподобности этой конкурентной гипотезы, и мы обсудим его подробнее.

Отметим, во-первых, что гипотеза о таком взаимодействии применительно к рассмотренным квазиэкспериментам более правдоподобна, чем в описываемых обычно случаях использования латинского квадрата для планирования подлинного эксперимента. В факторе «группы» ( $g$ ) смешаны два потенциальных источника систематических эффектов. Во-первых, это факторы систематического отбора, связанные с естественным формированием групп. Можно ожидать, что эти факторы не только дают главный эффект, но и взаимодействуют с фоном, естественным развитием, эффектами упражнения и т. д. Если бы эксперимент, в котором имеется полный контроль параметров, был организован подобным образом, каждый испытуемый был бы включен в свою группу независимым и случайным образом, и этот источник, как главного эффекта, так и эффекта взаимодействия был бы устранен или, по крайней мере, сведен до уровня ошибки выборки. В квазиэксперименте же уравнивание производится для осуществления некоторого уравнивания групп как раз потому, что невозможно случайное их комплектование. (В отличие от этого в полностью контролируемых экспериментах латинский квадрат применяется по соображениям экономии или для преодоления трудностей, подобных тем, которые возникают при выборочном исследовании земельных участков.)

Второй возможный источник указанных эффектов связан с последовательностью воздействий. Если бы все повторения *истинного* эксперимента следовали одному и тому же латинскому квадрату, то этот источник главных эффектов и эффектов взаимодействия также давал бы о себе знать. Однако в типичном истинном эксперименте повторные группы испытуемых распределяются по различным латинским квадратам, благодаря чему устраняется систематический эффект определенной по-

следовательности. Это также исключает возможность того, что за главный эффект  $X$  будет принят эффект взаимодействия фактора последовательности с другими факторами.

Порядковый номер воздействия вполне может вызвать главный эффект, связанный с повторным тестированием, естественным развитием, упражнением, накоплением опыта, переносом навыков. Фоновые явления также могут оказывать влияние на результаты последовательно применяемых воздействий. Правда, латинский квадрат препятствует контаминации главного эффекта  $X$  со стороны главных эффектов указанных факторов. Но там, где главные эффекты свидетельствуют о существенной неоднородности, видимо, больше оснований подозревать существование значимых взаимодействий, чем при отсутствии главных эффектов факторов  $g$  и  $t$ . Так, эффект тренировки, например, может быть монотонным, но, по всей вероятности, нелинейным и может вызвать как главный эффект, так и эффекты взаимодействия. Многочисленные примеры применения латинского квадрата в истинных экспериментах (например, в агробиологии) обычно не сопряжены с повторными измерениями, и там, как правило, систематический эффект по столбцам (см. схемы) отсутствует. Однако эксперименты перекрестного типа в этом отношении столь же уязвимы, как и квазиэксперименты.

Приведенные соображения ясно показывают чрезвычайную важность повторения квазиэксперимента с использованием различных латинских квадратов. При достаточном числе таких повторений квазиэксперимент превратился бы в истинный эксперимент. Они

должны были бы, вероятно, включать также достаточное число групп, чтобы сделать возможным случайное распределение целых групп по воздействиям (это средство контроля обычно предпочтительнее других). Но когда это невозможно, одиночный латинский квадрат интуитивно представляется удовлетворительным планом квазиэксперимента, ибо он позволяет продемонстрировать все эффекты на всех сравниваемых группах. Отдавая себе отчет в возможных ошибках интерпретации, эксперимент этого типа стоит применять, когда лучший контроль параметров эксперимента невозможен. Подчеркнув

141

серьезные недостатки данного плана, рассмотрим и оттеним теперь его сравнительные достоинства.

Как и в других квазиэкспериментах, преимущества эксперимента этого типа определяются непротиворечивостью результатов повторений эксперимента. Чтобы выявить наличие такой воспроизводимости, нужно устранить главные эффекты факторов  $g$  и  $t$ , записывая в каждой ячейке таблицы отклонения от средних по строкам ( $g$ ) и по столбцам ( $t$ ):  $M_{gt} - M_{g.} - M_{.t} + M_{..}$ . Перегруппируем теперь данные так, чтобы получилась таблица  $g \times X$  (вторая схема). Допустим, полученная картина обладает желаемой однородностью, то есть какое-то воздействие дало максимальные результаты по всем четырем группам и т. д. Каковы шансы, что имеет место не подлинный эффект  $X$ , а взаимодействие между факторами  $g$  и  $t$ ? Мы можем отметить, что наиболее вероятные взаимодействия этих факторов уменьшили бы или затушевали явный эффект  $X$ . Взаимодействие, имитирующее главный эффект  $X$ ,— явление редкое, и его вероятность снижается с увеличением размера латинского квадрата.

План 11 особенно привлекателен в условиях, когда при очень малом числе естественных групп (например, школьных классов) есть возможность планировать порядок воздействий, но нельзя случайным образом разделить группы на эквивалентные подгруппы для предъявления  $X$  или проведения тестирования. В тех случаях, когда возможно предварительное тестирование, годится также план 10, для которого также характерен риск смешивания эффекта  $X$  с эффектами взаимодействия состава групп и порядка предъявления воздействия. Этот риск, вероятно, окажется меньшим для сбалансированных планов, поскольку на каждой группе можно сравнить действие всех  $X$  и, следовательно, для имитации экспериментального эффекта потребовалось бы совмещение нескольких взаимодействий.

В то время как в экспериментах других типов особая реакция одной из групп на действие факторов фона или естественного развития может имитировать эффект  $X_1$ , в сбалансированном плане подобное явление должно было бы наблюдаться в различных ситуациях поочередно в каждой группе. Предполагается, конечно, что мы не будем расценивать главный эффект

142

$X$  как значимый, если изучение таблицы результатов показывает, что статистически значимый главный эффект имеет место в первую очередь за счет весьма сильного эффекта в единственной группе (см. также Wilk, Kempthorne [134]; Lubin [72] и Stanley [106]).

## **12. План с предварительным и итоговым тестированием на различных выборках**

При изучении больших популяций, таких, как города, предприятия, школы, воинские подразделения и т. д., часто может оказаться, что, хотя рандомизированное выделение подгрупп для дифференцированного экспериментального воздействия и невозможно, все

же удается осуществить нечто вроде полного контроля над тем, *когда* и *на ком* производить тестирование, применяя процедуры рандомизированного распределения испытуемых по группам. Такой контроль позволяет реализовать план 12:

$$\begin{array}{cc} R & O(X) \\ R & XO. \end{array}$$

В этой схеме строки представляют случайным образом выделенные эквивалентные подгруппы, а скобки при  $X$  означают, что результаты предъявления  $X$  по данной группе не используются. Одна выборка проходит тестирование до  $X$ , а другая — после  $X$ . Этот план не относится к числу сильных, на что указывает его графа в табл. 2. Тем не менее часто это единственная возможность, и ею стоит воспользоваться. План 12 применялся в социальных исследованиях, которые остаются лучшими в своей области (см., например, Star, Hughes[118]). Хотя его называли «имитацией» плана с начальным и конечным тестированием (Selltiz, Jahoda, Deutsch, Cook [99, с. 116]), он на самом деле имеет преимущество по сравнению с обычным планом с предварительным и последующим тестированием (то есть планом 2), поскольку обеспечивает контроль над главным эффектом тестирования и взаимодействием этого фактора с  $X$ . Главным недостатком этого плана является неконтролируемость фактора фона (см. Star, Hughes [118]).

143

В этой работе мы стремимся повысить у читателя интерес к «лоскутным» схемам, в которых средства контроля отдельных факторов вводятся более или менее постепенно (в отличие от более изящных «истинных» экспериментов, в которых контроль всех факторов риска для внутренней валидности обеспечивается одной контрольной группой). Повторное применение плана 12 в различных условиях и в разное время, как в плане 12а (см. табл. 2), обеспечивает контроль фонового фактора: если эффект обнаруживается снова, вероятность того, что он является результатом действия сопутствующих фоновых событий, снижается. Но неконтролируемыми конкурентными объяснениями все еще остаются сезонные циклы и некоторые другие фоновые изменения. Повторение эффекта в других условиях может снизить вероятность того, что он свойствен лишь первоначально взятой популяции. Однако если условия исследования допускают применение плана 12а, то применим также и план 13, которому в целом следует отдать предпочтение.

Ссылка на естественное развитие вряд ли может быть конкурентным объяснением даже в продолжающемся несколько месяцев исследовании общественного мнения. Но при выборочном исследовании общественного мнения или даже в некоторых студенческих аудиториях выборки достаточно велики и в возрастном отношении достаточно разнородны, чтобы группу, которая подвергается предварительному тестированию, можно было разделить на подвыборки, различающиеся по фактору естественного развития (по признаку возраста, курса и т. д.), и сравнивать эти подвыборки. Фактор естественного развития и, вероятно, более серьезная угроза медленных и сезонных изменений могут также контролироваться в плане 12б, где вводится дополнительная группа, которая подвергается более раннему предварительному тестированию. Это сближает данный план с планом временных серий, хотя и без повторного тестирования. Для популяций, подобных упомянутым в примере с пациентами психотерапевтического кабинета, где возможна спонтанная ремиссия, может оказаться несостоятельным имплицитное предположение о линейности естественного развития. Скорее, оно будет представлять собой постепенно замедляю-

144

щийся процесс, так что естественное увеличение от  $O_1$  к  $O_2$  будет большим, чем от  $O_2$  к  $O_3$ , что препятствует выявлению эффекта  $X$ .

Инструментальная погрешность представляет опасность, если план 12 применяется в массовом обследовании. Если одни и те же интервьюеры проводят предварительное и итоговое тестирование, обычно оказывается, что для многих из них начальное интервью было первым в их жизни, а к конечному они набираются некоторого опыта и, пожалуй, цинизма. Если каждая волна опросов проводится разными, причем немногими, людьми, происходит смешивание личных особенностей интервьюеров с экспериментальной переменной. Если интервьюерам известна гипотеза, то независимо от наличия или отсутствия воздействия  $X$  их ожидания могут породить различия, как это экспериментально показали Стэнтон и Бейкер [117], а также Смит и Хаймен [102]. Идеальным было бы использовать в каждой волне опросов различные эквивалентные случайные выборки интервьюеров и держать их в неведении относительно замысла эксперимента. Кроме того, состав интервьюеров может зависеть от времени года, например вследствие того, что в летние месяцы можно использовать студентов. Летом число отказов меньше и интервью длится дольше, чем зимой. В случаях, когда респонденты сами заполняют анкеты в классе или аудитории, подобная инструментальная погрешность менее вероятна, хотя изменение установки в отношении анкеты может оказаться таким, что его будет правильнее считать инструментальной погрешностью, чем влияниями  $X$  на  $O$ .

Если между предварительным и итоговым тестированием проходит несколько месяцев, то в эксперименте типа 12 возникает проблема выбывания. Если обе выборки составлены одновременно (этап  $R$ ), популяция со временем может меняться: на различных этапах исследования утрачивается доступ ко все большему числу респондентов и из популяции выбывают наиболее мобильные. Сигналом такой опасности служит различие между группами по числу лиц, с которыми интервьюерам не удалось встретиться.

По-видимому, в длительном исследовании выборки, на которых проводятся предварительное и итоговое

тестирования, должны составляться независимо друг от друга и в разное время, хотя это в свою очередь служит источником систематической погрешности вследствие возможных изменений в данной части генеральной совокупности. В условиях, например, школы учет посещаемости позволяет из результатов предварительного тестирования изъять те, которые принадлежат отсутствовавшим при итоговом тестировании ученикам. Это облегчает сравнение результатов обоих тестирований. Чтобы сделать возможной такую коррекцию в массовом обследовании, равно как и для дополнительного подтверждения наличия эффекта, на который не оказывал бы влияния процесс выбывания, группа, выбранная для предварительного тестирования, может быть подвергнута повторному тестированию (как в плане 12в, где различие  $O_1 - O_2$  следует подтвердить путем сравнения показателей  $O_1 - O_3$ ). Так было построено исследование Дункана и сотрудников [31], посвященное влиянию вводного курса психологии на преодоление ошибочных взглядов. (В этом плане эксперимента наличие группы, прошедшей оба тестирования, не обеспечивает возможности анализа приращений зависимой переменной по отдельным испытуемым ввиду отсутствия контрольной группы для учета влияния регрессии.)

Для экспериментов данного типа характерно то, что они переносят лабораторные методы в полевые условия, на которые исследователь хочет распространить свои выводы, испытывая действие  $X$  в естественной обстановке. Вообще говоря, как показано в табл. 1 и 2, планы 12, 12а, 12б и 12в по обеспечению внешней валидности или возможностям обобщения результатов могут превосходить планы 4, 5, 6 «истинных» экспериментов. Они предъявляют к респондентам так мало требований в отношении сотрудничества,

пробы в определенное время в определенном месте и т. д., что можно формировать репрезентативные выборки из заранее определенных популяций.

Планы 12 и 13 (и, конечно, некоторые варианты планов 4 и 6, в которых  $X$  и  $O$  осуществляются в условиях индивидуального контакта с обследуемыми) действительно обеспечивают возможность пользоваться репрезентативными выборками. Плюсы в колонке «взаимодействие состава групп и  $X$ » весьма

146

относительны и могли бы, справедливости ради, быть заменены вопросительными знаками, поскольку, как правило, респонденты выбираются не по теоретическим соображениям, а по признаку готовности участвовать и доступности, что нередко делает их нетипичными представителями генеральной совокупности, на которую предполагается распространить результаты. Но величина смещения, обусловленного составлением групп по признаку доступности испытуемых, намного меньше, чем в планах экспериментов, предъявляющих к ним более высокие требования, так что на этом фоне плюс представляется оправданным.

### 13. План с контрольными выборками для предварительного и итогового тестирования

Предполагается, что план 12 может быть использован в условиях, когда  $X$  (если он вообще имеет место) предъявляется всей группе в целом. При наличии сравнимых (или даже эквивалентных) групп, в которых  $X$  не вводится, можно в плане 12 добавить контрольную группу, получив тем самым план 13:

$RO$	$X$
$R$	$XO$
-----	
$RO$	
$R$	$O$

Этот план сходен с планом 10, если не считать того, что предварительное и итоговое тестирование проводится на разных лицах, благодаря чему исключается возможное взаимодействие тестирования и  $X$ . Как и для плана 10, в этом случае существует опасность (что и является недостатком плана в отношении внутренней валидности) принять за эффект  $X$  частную тенденцию, характерную для данной экспериментальной группы и фактически не связанную с воздействием. Этот источник невалидности можно устранить, увеличивая число вовлеченных в исследование социальных единиц (школ, городов, предприятий и т. д.) и предписывая им посредством рандомизации экспериментальный или контрольный режим. Это позволяет получить план истинного

147

Таблица 3

Источники невалидности для квазиэкспериментальных планов 13—16

	Источники невалидности											
	внутренней							внешней				
	фон	естественное развитие	тестирование	инструментальная погрешность	регрессия	состав групп	выбывание	взаимодействие состава групп с естественным развитием и др. факторами	взаимодействие тестирования и X	взаимодействие состава групп и X	реакция испытуемых на эксперимент	взаимодействие между разными X
<i>Квазиэкспериментальные планы (продолжение)</i>												
13. План с контрольными выборками для предварительного и итогового тестирования	+	+	+	+	+	+	+	-	+	+	+	
	$\begin{matrix} R & O & (X) \\ R & & X & O \\ \hline R & O & & \\ R & & & O \end{matrix}$											
13а.	$\left\{ \begin{matrix} R & O & (X) \\ R & & X & O \\ \hline R & O & (X) \\ R & & X & O \\ \hline R & O & & \\ R & & & O \\ \hline R & O & & \\ R & & & O \\ \hline R & O & & \\ R & & & O \end{matrix} \right.$	+	+	+	+	+	+	+	+	+	+	

	Источники невалидности											
	внутренней							внешней				
	фон	естественное развитие	тестирование	инструментальная погрешность	регрессия	состав групп	выбывание	взаимодействие состава групп с естественным развитием и др. факторами	взаимодействие тестирования и X	взаимодействие состава групп и X	реакция испытуемых на эксперимент	взаимодействие между разными X
14. Множественные серии замеров	+	+	+	+	+	+	+	+	-	-	?	
$O O O X O O O$												
$O O O O O O$												
15. Рекуррентный институциональный цикл («лоскутные» планы)												
Класс A X O <sub>1</sub>												
Класс B <sub>1</sub> R O <sub>2</sub> X O <sub>3</sub>												
Класс B <sub>2</sub> R X O <sub>4</sub>												
Класс C O <sub>5</sub> X												
Контр. гр. для кл. B O <sub>6</sub> <sup>1</sup>												
Контр. гр. для кл. C O <sub>7</sub> <sup>1</sup>												
$O_2 < O_1$	+	-	+	+	?	-	?		+	?	+	
$O_5 < O_4$												
$O_2 < O_3$	-	-	-	?	?	+	+		-	?	+	
$O_2 < O_4$	-	-	+	?	?	+	?		+	?	?	
$O_6 = O_7$												
$O_{2y} = O_{2o}$	+						-					
16. План, в котором нарушение непрерывности регрессии выступает как признак экспер. эффекта	+	+	+	?	+	+	?	+	+	-	+	+

<sup>1</sup> Контрольная группа для класса B (соответственно C), взятая из генеральной совокупности.

эксперимента, подобного плану 4, с тем, однако, исключением, что здесь мы избегаем повторного тестирования одних и тех же испытуемых. Такой план можно обозначить как план 13а. Его схематическое представление (см. табл. 3) усложняется двумя уровнями эквивалентности (достигаемой путем рандомизации). На уровне испытуемых внутри каждой социальной единицы существует эквивалентность отдельных выборок, используемых для предварительного и итогового тестирования (что отмечено пунктом рандомизации — R). Между отдельными социальными единицами, которые подвергаются одному из этих воздействий, такой эквивалентности нет, на что указывают пунктирные линии, R' обозначает уравнивание экспериментальной и контрольной групп путем рандомизированного распределения этих многочисленных социальных единиц по двум режимам.

Как видно из табл. 3, план 13а оказывается безукоризненным в отношении как внутренней, так и внешней валидности (последнее уже обсуждалось при анализе плана 12), причем здесь благодаря наличию многочисленных социальных единиц (в отличие от использования только одной из них) добавляется контроль взаимодействия между

факторами состава групп и  $X$ . Насколько известно, эта отличная, но требующая больших затрат схема эксперимента еще не была реализована.

#### 14. План с множественными сериями измерений

При изучении эффектов крупной административной перестройки по данным серии измерений исследователь поступит правильно, отыскав аналогичное учреждение, не подвергшееся  $X$ , и взяв оттуда данные для «контрольной» серии (в идеале учреждение, подвергающееся  $X$ , должно быть, выбрано наугад):

$$\begin{array}{cccccccc} O & O & O & O & O & X & O & O & O & O & O \\ \hline O & O & O & O & O & & O & O & O & O & O \end{array} .$$

Этот план включает план 10, то есть план с неэквивалентной контрольной группой ( $X$  в окружении  $O$ ), по сравнению с которой он выигрывает в достоверности

150

интерпретации благодаря использованию многочисленных измерений, причем экспериментальный эффект как бы демонстрируется дважды — один раз на фоне контрольных данных, а другой — в сравнении со значениями до  $X$  внутри той же самой серии данных, как в плане 7. Кроме того, взаимодействие фактора состава групп с естественным развитием контролируется постольку, поскольку при большей общей скорости изменения зависимой переменной в экспериментальной группе это должно проявиться также в  $O$ , предшествующих  $X$ . В табл. 2 и 3 этот добавочный выигрыш представлен слабо. Он нашел отражение лишь в последней колонке внутренней валидности — «взаимодействие состава групп с естественным развитием». Поскольку естественное развитие контролируется как в экспериментальной, так и в контрольной серии, то по соображениям, приведенным выше при обсуждении серии периодических измерений (план 7), различие в составе групп, подвергаясь влиянию со стороны факторов естественного развития, инструментальной погрешности или регрессии, вряд ли может дать заметный эффект. Однако взаимодействие фактора состава групп с фоном может сказываться на результатах.

Как для плана 7, поставлен минус в колонке внешней валидности — «взаимодействие тестирования и  $X$ », хотя, как и план 7, данный план используется преимущественно в условиях, когда тестирование не вызывает такого эффекта. Табл. 3 показывает также, что надо учитывать возможность того, что полученный эффект  $X$  может оказаться особенностью исследуемой популяции. Что касается проверки значимости эффекта, то предлагается анализировать различия между результатами соответственных измерений в экспериментальной и контрольной группах, как данные, полученные по плану 7. Эти различия более вероятно могут оказаться линейными, чем данные исходных временных рядов.

В общем, это отличный квазиэксперимент — пожалуй, лучший из наиболее доступных. Как видно из сделанных выше замечаний и из обсуждения плана 10, он обладает явными преимуществами перед экспериментами типа 7 и 10. Доступность повторных измерений делает его особенно пригодным для исследовательской работы в школе.

151

## 15. Рекуррентный институциональный цикл — «лоскутный» план

План 15 является примером стратегии полевого исследования, при котором, начав с неадекватного плана, затем добавляют к нему специфические меры контроля того или иного рекуррентного источника невалидности. Результатом часто оказывается неуклюжее нагромождение проверок, которое — хотя ему и недостает внутренней симметрии истинных экспериментов — все же приближается к экспериментальному исследованию. Приняв эту стратегию, экспериментатор должен быть начеку относительно конкурентных интерпретаций (отличных от констатации эффекта  $X$ ), от которых план не застрахован, и искать такие способы анализа или обобщения данных, которые исключали бы подобные интерпретации. Другая черта, часто присущая данному плану, состоит в выявлении эффекта  $X$  несколькими различными способами. Это очень важно, когда каждое отдельное сравнение результатов само по себе не позволяет сделать однозначного вывода.

Такой «лоскутный» план, о котором идет речь, применим лишь к ограниченному кругу проблем и условий и должен строиться исходя из особенностей этих условий. Основная его идея становится яснее при рассмотрении третьей и четвертой граф табл. 1. Как легко заметить, ряды стоящих там плюсов и минусов для плана 2 и 3 по большей части дополняют друг друга, а удачное сочетание этих двух неполноценных планов может оказаться весьма эффективным. Данный план применим в тех случаях, когда определенный аспект институционального процесса циклически реализуется каждый раз для новой группы респондентов, как это имеет место, скажем, в школе, в системе воспитания, обучения ремеслу и т. д. Если в этих условиях интерес представляет оценка результатов такого всеобъемлющего и сложного воздействия, как общая программа обучения и социализации, то план 15 является, видимо, наилучшим из всех пока разработанных.

Этот план был впервые концептуально разработан применительно к изучению влияния командирской и летной подготовки курсантов военно-воздушной школы

(в ходе 14-месячного учебного цикла) на их социальные установки в отношении начальников и подчиненных и на функции лидерства в группе (Campbell, McCormack [16]). Проведение истинного эксперимента затруднялось отсутствием контроля над тем, кто будет подвергнут экспериментальному воздействию, а кто — нет. Действительно, как разделить принятых курсантов на две уравненные группы, одна из которых прошла бы годичную программу обучения, а другая — отправлена «на гражданку»? Даже если бы такой эксперимент и удалось поставить (непредвиденные сокращения бюджета сделали бы в отдельных случаях возможным проведение такого опыта), то все равно из тех, кто был принят, прошел отсев, доставлен на авиабазу, а затем отправлен домой, вряд ли получилась бы идеальная контрольная группа ввиду реакции испытуемых на эксперимент и нарушения ритма их жизни. Различие между ними и экспериментальной группой, проходящей подготовку, вряд ли могло бы служить хорошей основой для распространения результатов на нормальные условия набора и обучения курсантов. Экспериментатор, однако, мог выбирать, *когда и на ком* производить наблюдения. Это, а также тот факт, что экспериментальная переменная носила рекуррентный (циклический) характер, то есть оказывала воздействие каждый раз на новую группу респондентов, обеспечивало какую-то степень контроля параметров эксперимента. В этом исследовании можно было проводить два типа сравнений, относившихся к влиянию опыта военной службы на социальные установки. Каждое из них было совершенно неадекватно в плане экспериментального контроля, но если оба они свидетельствовали об одном и том же, они взаимно подкрепляли друг друга, ибо их слабые места оказывались различными. Сравнения первого типа производились по результатам, полученным в одно и то же время на группах,

отличавшихся сроком службы, а сравнения второго типа — внутри каждой группы между показателями первой недели и после 13 месяцев обучения. В идеализированной форме эта схема выглядит так:

Класс	$A$	$X$	$O_1$ ,	
Класс	$B$		$O_2$	$X$
				$O_3$

153

В этой схеме объединены «лонгитюдинальный» подход и «сравнительное изучение поэтапных срезов» — процедуры, обычно используемые в исследованиях развития. При этом предполагается, что одновременно измерения могут производиться на группе, уже подвергшейся  $X$ , и на группе, которая вскоре должна подвергнуться  $X$ . Сравнение данных  $O_1$  и  $O_2$ , следовательно, отвечает плану 3 (сравнение статических групп). Повторные измерения в классе  $B$  спустя один цикл позволяют воспроизвести план 2. В табл. 3 первые две строчки, относящиеся к плану 15, показывают особенности этих сравнений. Результат сравнения в поперечном срезе,  $O_1 > O_2$ , нельзя отнести за счет фоновой стимуляции или эффекта повторного тестирования. Однако его можно объяснить тем, что из года в год состав курсантов меняется (на что указывает минус в колонке «состав группы») или что респонденты стали на год старше (минус по фактору «естественное развитие»). Если тестирования проведены в один и тот же период времени, то маловероятно смешивание эффектов  $X$  и фактора инструментальной погрешности, или изменения характера средств измерения. Эффект выбывания обычно также является конкурентным объяснением при сравнении социальных установок новичков и слушателей второго года обучения: результаты  $O_1$  и  $O_2$  могут отличаться друг от друга из-за того, что лица, подобные отсеявшимся из класса  $A$ , еще представлены в классе  $B$ . Этого недостатка можно избежать, если регистрировать индивидуальные результаты и отложить анализ данных до того момента, когда закончится воздействие  $X$  на класс  $B$ , что позволит устранить из  $O_2$  результаты тех респондентов, которые не завершили курса подготовки. Поскольку эта процедура часто отсутствует, в колонке «выбывание» поставлен вопросительный знак. Другой знак вопроса в колонке «регрессия» предупреждает о возможности ложных эффектов в случае, если в эксперименте используются показатели, по которым производился отбор курсантов для обучения. В этих условиях следует ожидать систематического смещения результатов, которое нельзя приписать воздействию  $X$ . Если сравнение предварительного и итогового тестирования при  $O_2$  и  $O_3$  показывает, что изменение переменной от  $O_2$  к  $O_3$  того же порядка, что и при переходе

154

от  $O_1$  к  $O_2$ , есть основания отклонить конкурентную гипотезу о том, что это изменение обусловлено различием в отборе слушателей для данных двух классов, а также исключить возможность использования в качестве объяснения фактора выбывания. Однако если пользоваться одним лишь сравнением  $O_2$  —  $O_3$ , то возникает опасность конкурентного объяснения результатов фоновой стимуляцией и эффектом тестирования.

В тех случаях, когда исследуется эффект годичного обучения, наибольших затрат требует проведение двух измерений, разделенных промежутком времени в один год. Если на это уже имеются средства, то проведение дополнительных измерений во второй период обследований потребует лишь незначительных добавочных затрат. Принимая это во внимание, можно расширить план данного эксперимента согласно схеме, приведенной в табл. 3. Пользуясь правом решать, кто и когда будет проходить измерительную процедуру, можно разбить класс  $B$  на две уравненные выборки, одна из которых обследуется и до и

после воздействия, а другая — только после него ( $O_4$ ). Эта вторая группа обеспечивает сравнение тщательно уравненных выборок по первому измерению, проводимому до и после  $X$ , причем точность сравнения здесь выше, чем для  $O_1 — O_2$  с точки зрения фактора состава групп, и выше, чем для  $O_2 — O_3$  в плане эффекта повторного тестирования. Эффект  $X$ , таким образом, должен быть зафиксирован в трех различных сравнениях:  $O_1 > O_2$ ,  $O_2 < O_3$  и  $O_2 < O_4$ .

Отметим, однако, что  $O_2$  фигурирует во всех трех сравнениях и все они могут подтверждать эффект только благодаря необычному характеру этой серии измерений. Введение  $O_5$  и вместе с ним класса  $C$ , проходящего обследование во время второй волны тестирования перед  $X$ , позволяет получить данные для сравнения с  $O_4$  и  $O_1$  и т. д., что обеспечивает необходимую избыточность. Разбиение класса  $B$  на две части обеспечивает бóльшую четкость сравнения  $O_4 — O_5$ , чем  $O_3 — O_5$ . Но разбиение класса на обследуемую и необследуемую группы нередко провоцирует реакцию испытуемых на эксперимент. Поэтому в строке  $O_2 < O_4$  в табл. 3 для этого фактора поставлен вопросительный знак. Возникновение указанной реакции зависит от

155

конкретных условий. Если испытуемые тянут жребий, а затем половину класса просят пройти в другую комнату, то опасность реакции на эксперимент велика (см., например, Duncan et al. [31], Solomon [104]). Если, как это бывает в ряде исследований, контакт с испытуемым устанавливается в индивидуальном порядке, класс удастся разделить на две уравненные половины, не вызывая подозрений респондентов. В тех случаях, когда поток слушателей состоит из ряда групп, каждая из которых работает по своему расписанию, есть возможность, не разрушая эти группы, разделить их на те, которые проходят и не проходят предварительное тестирование (см., например, Novland, Lumsdaine, Sheffield [51]). В рамках одной аудитории анкеты или тесты разного содержания можно раздать всем испытуемым таким образом, что одна случайно выбранная половина слушателей составит группу, данные которой можно использовать в качестве предварительного тестирования, а другая, обследуемая с помощью иного инструментария, может служить группой для сравнения. При этом такое разделение будет провоцировать нежелательную реакцию не больше, чем обследование всей аудитории.

В плане 15 (в той его части, которая представлена измерениями  $O_1 — O_5$ ) не контролируется фактор естественного развития. Серьезность этого ограничения зависит от объекта исследования. Если речь идет об усвоении специфических (эзотерических) навыков или знаний, конкурентная гипотеза о том, что испытуемый покажет более высокие результаты, становясь старше и приобретая жизненный опыт, представляется крайне маловероятной.

Однако в рассмотренном выше исследовании отношения к старшим и младшим по званию (Campbell, McCormack [16]) изменение показателей было таким, что вполне могло объясняться приобретением жизненного опыта людьми данного возраста и происхождения просто в силу взросления или пребывания вдали от дома. В таких условиях контроль естественного развития представляется весьма важным. Поэтому к плану были добавлены измерения  $O_6$  и  $O_7$ , обеспечивавшие проверку гипотезы естественного развития в поперечном срезе во второй период обследования. Это предпо-

156

лагало тестирование двух групп (из общей популяции), отличающихся друг от друга только по возрастным показателям. Но эти показатели были сходны с теми, которые характеризовали классы  $B$  и  $C$  в период их обследования. Для подтверждения гипотезы о наличии эффекта  $X$  результаты  $O_6$  и  $O_7$  должны быть одинаковыми или по крайней мере

отличаться друг от друга меньше, чем данные тестирований, проведенных до и после  $X$ . Составление контрольных групп из общей популяции зависит от особенностей гипотезы. Учитывая то значение, которое повсеместно придается социальному происхождению и образованию, можно было бы подобрать группы таким образом, чтобы они были сопоставимы с институциональными группами по показателю принадлежности к тому или иному социальному классу и по предварительному образованию. Следовало бы также составлять контрольные группы из лиц призывного возраста, проживших вдали от дома один год (группа  $O_6$ ) или только что покинувших родные места (группа  $O_7$ ). Но и тогда эти контрольные группы, в которых учитывался бы фактор возраста, были бы в какой-то степени неудовлетворительным средством контроля, а их обследование составляло бы самую большую статью расходов, так как обычно легче провести тестирование в рамках одной организации, чем отбирать респондентов из общей популяции. По этим соображениям  $O_6$  и  $O_7$  были приурочены ко второй волне тестирования, поскольку если бы эффект  $X$  не был обнаружен при сравнении результатов  $O_1$  и  $O_2$ , то указанные дорогостоящие мероприятия, скорее всего, оказались бы неоправданными (иное дело, когда гипотеза состоит, скажем, в том, что действие институционального по своему характеру  $X$  подавляет естественный процесс развития).

Возможен и другой подход с использованием поперечного среза для контроля фактора естественного развития, если популяция, которая вовлекается в институциональный процесс, неоднородна по показателям возраста (или числа лет, проведенных вне дома, и т. д.). Это характерно для многих ситуаций — например, когда изучается эффект какого-либо лекционного курса в колледже. В этом случае измерения  $O_2$  делят на две части, относящиеся к старшей ( $O_{20}$ ) и младшей ( $O_{2y}$ )

157

подгруппе, и выясняют, отличаются ли друг от друга эти подгруппы в той же степени, как и  $O_2$  от  $O_1$  (правда, здесь возникает опасность все той же отрицательной корреляции между возрастом и способностями в рамках одного года обучения). Вместо контрольных групп, отобранных из общей популяции по признаку возраста и зрелости, иногда лучше для сравнения пользоваться контингентом другого института. Например, новобранцев военно-воздушных сил сравнить с первокурсниками колледжа. Если проводится сравнение такого типа, то экспериментальная переменная сводится лишь к тем характеристикам, которые не являются общими для обоих типов институтов или организаций. В этих случаях обычно могут быть использованы в целом более эффективные планы 10 и 13.

Формальные свойства плана 15 как будто делают его применимым даже в сфере психотерапии. Здесь, однако, обнаруживается, насколько трудно осуществить надлежащий контроль фактора естественного развития. Независимо от того, как испытуемые контрольной группы отбираются из общей популяции, они будут существенно отличаться от членов экспериментальной группы, если только сами не являются пациентами психотерапевта. Даже если они точно так же больны, как и пациенты, подвергающиеся психотерапевтическому воздействию, их информированность, мнение о психотерапии, вера в нее будут почти наверняка иными. У группы больных, настроенных оптимистически в отношении психотерапии, скорее активизируются потенциальные силы организма для борьбы с болезнью, что будет нетипичным для любой сопоставимой с нею группы, которую мы могли бы, вероятно, составить, и, таким образом, взаимодействие состава групп и естественного развития могло бы быть ошибочно истолковано как эффект  $X$ .

Когда объектом изучения становится процесс развития как таковой, отсутствие контроля фактора естественного развития, конечно, не является недостатком, поскольку этот фактор находится в фокусе исследования. В исследовании процессов развития следовало бы систематически использовать сочетание лонгитюдинальных и поэтапно-срезовых сравнений. В поперечном срезе естественное развитие смешано с фактором состава

ва и выбывания из групп. В лонгитюдных исследованиях эффект естественного развития смешивается с эффектами тестирования и фоновых воздействий. Само по себе лонгитюдное исследование, вероятно, не лучше поэтапных срезов, хотя большие затраты придают ему больший престиж. Идеальным представляется сочетание того и другого с использованием повторных поперечных срезов, выполненных в разное время.

При построении плана 15 предполагалось, что итоговое тестирование одной группы проводится одновременно с предварительным тестированием другой. Однако это возможно не во всех случаях, когда мы хотели бы применить этот план. Типичную ситуацию исследования, проводимого в условиях школы, более точно отражает следующая схема:

Класс <i>A</i>	<i>X</i>	<i>O</i> <sub>1</sub>			
Класс <i>B</i> <sub>1</sub>		<i>R</i>	<i>O</i> <sub>2</sub>	<i>X</i>	<i>O</i> <sub>3</sub>
Класс <i>B</i> <sub>2</sub>		<i>R</i>	<i>X</i>	<i>O</i> <sub>4</sub>	
Класс <i>C</i>				<i>O</i> <sub>5</sub>	<i>X</i>

В этой схеме нет четкого контроля фактора фона вследствие того, что сравнения *O*<sub>1</sub> и *O*<sub>2</sub>, а также *O*<sub>4</sub> и *O*<sub>5</sub> проводятся не одновременно. Однако когда эффект обнаруживается в обоих сравнениях, то ссылка на посторонние воздействия вряд ли будет обоснованной, разве только предположить, что имел место чрезвычайно сложный ряд совпадений.

Отметим, что общие исторические изменения, влияние которых мы действительно обнаруживаем при изучении социальных установок, не смешиваются с четкими экспериментальными результатами. Любые такие изменения могут вызвать *O*<sub>2</sub>, занимающее промежуточное положение между *O*<sub>1</sub> и *O*<sub>3</sub>, тогда как по гипотезе о наличии эффекта *X* они должны быть одинаковыми, а *O*<sub>2</sub> отличаться от обоих (*O*<sub>1</sub> и *O*<sub>3</sub>) в одном и том же направлении. В целом при повторении эксперимента в различных условиях смешивание с эффектами фоновых воздействий маловероятно для данного варианта этого плана. Но если продолжительность институционального цикла менее года, становится возможным

смешивание эффекта *X* с сезонными колебаниями социальных установок, морального состояния, оптимизма, уровня интеллекта — всего, чего угодно. Предположим, *X* — курс, читаемый только в осенний семестр, а в период с сентября по январь под влиянием сезонных климатических факторов возрастают враждебность и пессимизм. Тогда все проявления этой рекуррентной сезонной тенденции будут приписываться к эффектам *X*. В подобных случаях следует предпочесть планы 10 и 13.

Если сравнение данных лонгитюдного исследования и исследования поэтапных срезов выявляет сопоставимые эффекты *X*, это нельзя приписать взаимодействию между естественным развитием и различиями в составах классов. Но, поскольку этот контроль не отражен в схеме, представленной в табл. 3, данная колонка оставлена незаполненной. Оценки по критериям внешней валидности в целом следуют образцу, характерному для предыдущих планов, содержащих те же элементы. Знак вопроса в колонке «Взаимодействие состава групп и *X*» предупреждает о том, что результаты действительны только для институционального цикла, изучавшегося в эксперименте. Ввиду сложности *X*

исследование обычно подходит скорее для практических, а не теоретических целей, и, вероятно, в данном случае выводы и предназначаются для одного исследуемого института.

Последним в этом ряду квазиэкспериментальных планов вводится план 16, связанный с анализом нарушения непрерывности регрессии.

### **16. План, в котором нарушение непрерывности регрессии выступает как признак экспериментального эффекта<sup>1</sup>**

Основные идеи плана с нарушением непрерывности регрессии будет легче понять, если мы сначала рас-

---

<sup>1</sup> В данном разделе использованы материалы, опубликованные в работах: Campbell D. T. Quasi-Experimental Designs. — In: Social Experimentation: A method for planning and evaluating social innovations. Ed. by H. W. Riecken and R. F. Boruch. N. Y., Academic Press, 1974. © Academic Press, Inc., 1974, Campbell D. T. Reforms as Experiments. — In: Reading in evaluation research. Ed. by F. G. Caro, N. Y., Russell Sage Foundation, 1977. © Sage Publications, Inc, 1977. — *Прим. ред.*

160

смотрим сильную и слабую формы одного истинного эксперимента, применимого к ситуациям того же типа, что и обсуждаемый квазиэксперимент. Напомним, что одним из неизменных условий проведения социального эксперимента является невозможность предоставления всем желающим определенных благ, выступающих в роли экспериментальных воздействий. Это гарантирует, что число экспериментальных единиц (испытуемых, школ, городов и т. д.), готовых подвергнуться воздействию, превышает число таких единиц, которые фактически могут быть им охвачены. Данное условие, как правило, выполняется для пилотажных программ, которые применяются к подгруппам популяции, выделенным по принципу больших заслуг или большей нуждаемости. Если число кандидатов превышает количество предусмотренных программой мест, то обеспечивается выполнение одного из условий для формирования контрольной группы из части кандидатов.

В этих условиях лучшая стратегия состояла бы в проведении истинного эксперимента — в составлении экспериментальной и контрольной групп путем рандомизации, охватывающей *весь ранговый ряд* «достойности». Рассмотрим, например, программу профессиональной подготовки полностью или частично безработных молодых людей для получения ими определенной хорошо оплачиваемой работы. Критерием отбора может служить (и одно время действительно служил) доход на одного члена семьи кандидата. Если число «достойных» кандидатов, подавших заявление о приеме, больше числа вакансий, то администратор, ориентированный на эксперимент, мог бы отобрать будущих участников из всех «достойных» случайным образом. Учитывая выборочные данные эквивалентных оставшихся кандидатов для формирования из них контрольной группы, он мог бы в дальнейшем получить сведения о заработках тех, кто прошел и кто не прошел подготовку, и измерить ее эффект.

Но во многих случаях рандомизированный отбор «достойных» считается неприемлемым. При этом ссылаются на то, что признак, по которому производится отбор, не дихотомичен, а имеет различные степени выраженности. Если возможности слишком ограничены, чтобы принять всех кандидатов, то места должны быть

161

предоставлены наиболее нуждающимся и наиболее «достойным». Этот довод против рандомизации часто приводится, как если бы существующие процедуры распределения

кандидатов *действительно* удовлетворяли требованиям равенства, в свете которых рандомизация якобы оказывается несостоятельной. Тщательное изучение работы соответствующих учреждений показало бы в большинстве случаев, что они не удовлетворяют этим требованиям. Процедура ранжирования «достойных» кандидатов по степени «нуждаемости», на которой основывалось бы решение о приеме, обычно отсутствует. Случайность процедур отбора и распределения делает возможным то, что кандидаты могут зачисляться в значительной степени по соображениям удобства администрации или просто по знакомству. Чтобы избежать чрезмерного потока заявлений, вновь открывающимся возможностям дается минимальная огласка. Если применяется правило «первым пришел — первым обслужен», то наиболее нуждающиеся редко оказываются среди тех, кто внимательно следит за новыми возможностями, а «пришедшие первыми» узнают о них по неофициальным каналам, еще до опубликования соответствующего объявления. В отличие от существующей практики широкая огласка программы, приводящая к избытку кандидатов, которые затем в случайном порядке отбираются для участия в программе или входят в контрольную группу, представляла бы собой высокоморальную процедуру, помимо и сверх того, что она делала бы возможным экспериментальное оценивание.

Однако если нельзя преодолеть сопротивление полной рандомизации, то все же можно провести слабую форму этого эксперимента, а именно эксперимент с «*рандомизацией, разрывающей ранговую связку*», в котором особое внимание уделяется степени достоинства.

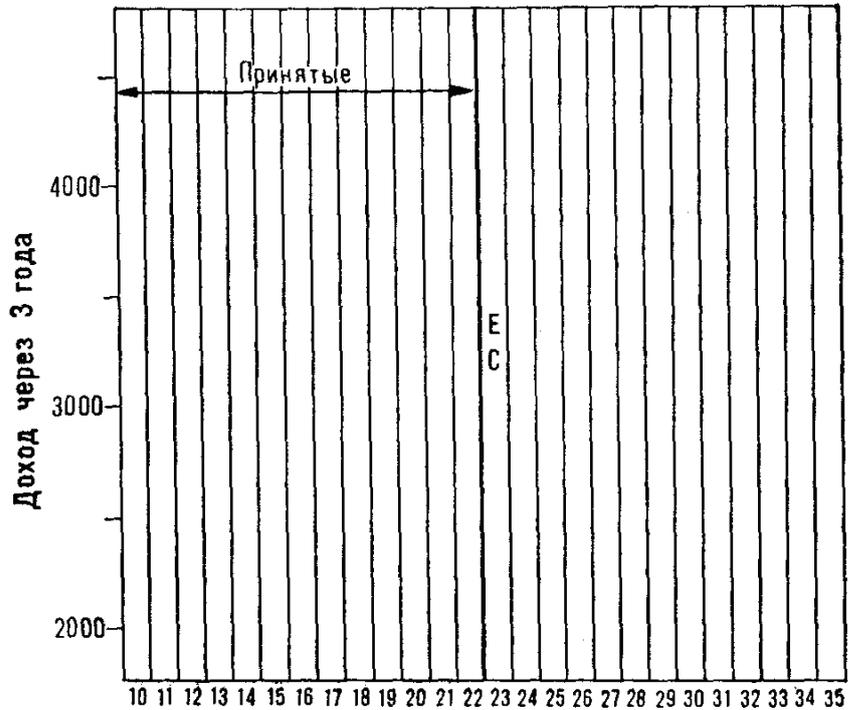
Предположим, что отбор участников программы производится не из всех кандидатов, а сконцентрирован на наиболее нуждающихся, причем мест хватает как раз для приема всех тех, чей недельный доход на каждого члена семьи составляет 22 доллара и ниже, плюс еще половину тех, для кого эта цифра равна 23 долларам. Теперь, не нарушая ограничения, согласно которому программа предназначается наиболее нуждающимся,

## 162

можно рандомизировать отбор из тех, кто попал в ранговую связку «23 доллара». Таким образом удастся провести небольшой истинный эксперимент. Правда, это слабый эксперимент в двух отношениях. Во-первых, в нашем распоряжении оказывается очень малое число случаев, и, следовательно, различие, отражающее подлинный эффект, может оказаться статистически незначимым или даже иметь противоположный знак из-за флюктуаций выборочных показателей. Естественно, возникает желание максимально увеличить число лиц, приходящихся на пограничную ранговую связку, для чего можно расширить интервал значений критерия отбора, внутри которого они считаются практически равными. Так, в настоящем примере интервал в 2 доллара дает большее число связанных рангов, чем, скажем, интервал в 50 центов, и к тому же позволяет избежать чрезмерной и бессмысленной точности в последнем случае. Другая слабость этого эксперимента состоит в том, что он исследует эффект воздействия только для узкого класса значений критерия отбора и поэтому позволяет лишь в ограниченной степени судить об эффекте по всему диапазону значений этого критерия (см. рис. 4). Правда, исследуемый узкий интервал вряд ли сильно отличается от соседних интервалов, которые могли бы стать граничными в случае возможного расширения или сокращения программы, и, следовательно, получаемые данные существенны для принятия важных административных решений.

При рассмотрении рис. 4 возникает вопрос: какова дальнейшая судьба тех, для кого значение критерия отбора ( $k$ ) слегка отличается от граничного? Считая программу подготовки эффективной, можно было бы ожидать, что полностью включенная в эксперимент категория лиц с показателем  $k$ , равным 22 долларам, будет в дальнейшем иметь доход, весьма близкий к доходу экспериментальной ( $E$ ) подгруппы с  $k = 23$  долларам

— возможно, чуть ниже, соответственно несколько меньшему начальному доходу, но все же выше, чем для контрольной группы (С) с  $k = 23$ . Аналогично дело будет обстоять и для  $k$ , равных 21, 20 и ниже. С другой стороны, дальнейшие доходы лиц с  $k = 24$  (никто из них не участвовал в эксперименте) должны быть примерно такими же, как и у членов контрольной группы

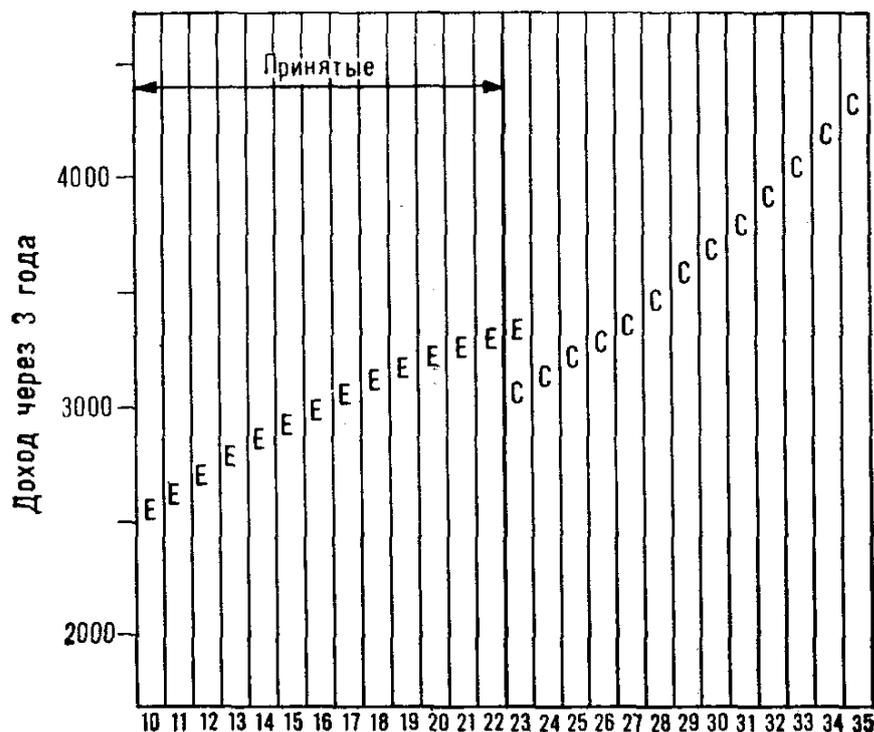


Недельный доход на одного члена семьи

Р и с. 4. Результат гипотетического эксперимента по плану с разрывом ранговой связки, в котором одни кандидаты с недельным доходом 23 доллара на одного члена семьи принимаются в случайном порядке на курсы профессиональной подготовки, а другие образуют контрольную группу. Все кандидаты с доходом 22 доллара и ниже проходят подготовку. Средний заработок испытуемых спустя 3 года отмечен буквами *E* и *C* соответственно для экспериментальной и контрольной группы.

с  $k = 23$  — чуть выше, но не как у членов экспериментальной группы с  $k = 23$  и т. д. На рис. 5 представлены гипотетические результаты такого прослеживания последующих доходов по всем значениям  $k$ .

Сравнение рис. 5 и 4 наводит на мысль о том, что результаты гипотетического эксперимента с рандомизацией ранговой связки можно вывести из анализа эффектов при неслучайном назначении воздействия по всему диапазону значений критерия отбора  $k$ . Так, если принятыми оказываются все лица с доходом в 22 доллара и ниже, а категория «23 доллара» целиком остается вне воздействия, то, получив результат, изображенный на рис. 6, мы могли бы с полной уверенностью считать, что эксперимент с рандомизацией ранговой связки привел бы к результатам, показанным на рис. 4 и



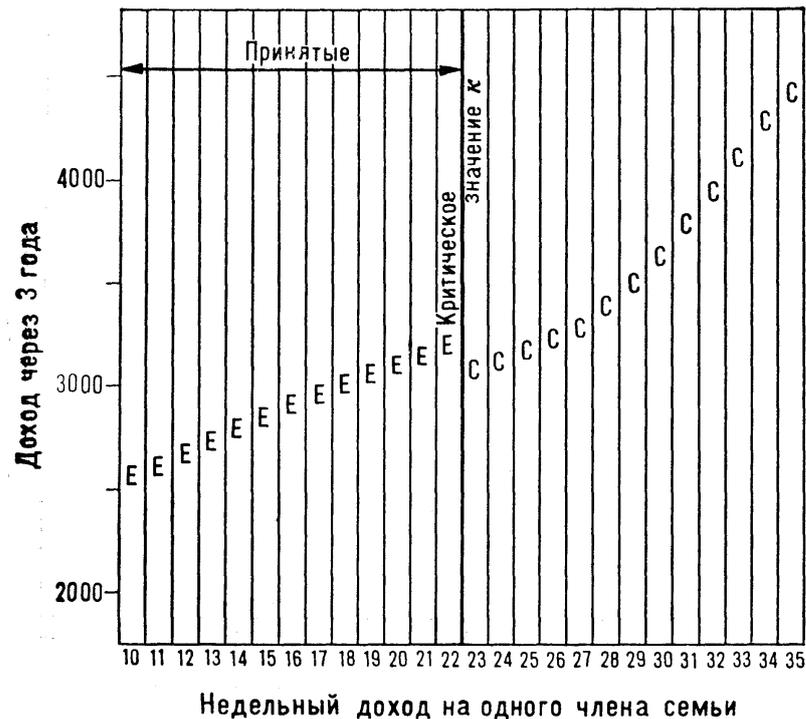
Недельный доход на одного члена семьи

Р и с. 5. Данные гипотетического эксперимента по плану с разрывом ранговой связки для различных категорий доходов на члена семьи (включая данные, представленные на рис. 4). Все кандидаты с доходом 22 доллара и ниже получили подготовку на данных курсах, и никто не был принят, если доход на одного члена семьи составлял 24 доллара и выше.

5. В то же время результаты, представленные на рис. 7, указывали бы на отсутствие эффекта воздействия.

Рис. 6 и 7 иллюстрируют использование нарушения непрерывности регрессии в качестве признака экспериментального эффекта<sup>1</sup>. Этот квазиэксперимент может заменить эксперимент с разрывом ранговой связки, но не с рандомизацией, распространенной на весь диапазон значений  $k$ . Результаты анализа не предназначены для оценки эффектов при любых значениях критерия

<sup>1</sup> См., например, работы: Thistlethwaite D. L. and Campbell D. T. Regression-discontinuity analysis: An alternative to the ex post facto experiment. — «Journal of Educational Psychology», 1960, v. 51, p. 309—317; Campbell D. T. Reforms as experiments. — «Amer. Psychologist», 1969, v. 24, p. 409—429; Goldberger A. S. Selection bias in evaluating treatment effects: Some formal illustrations. Madison, Institute for Research on Poverty, Univ. of Wisconsin. 1972.

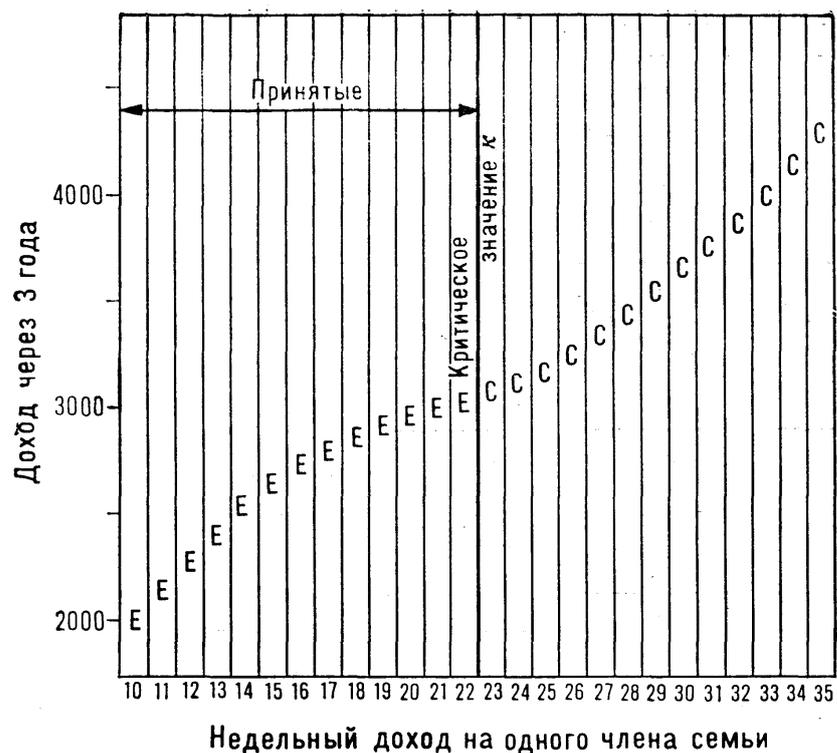


Р и с. 6. Гипотетические данные эксперимента по плану, в котором используется нарушение непрерывности регрессии в качестве показателя экспериментального эффекта. Эффект курсов профессиональной подготовки тот же, что и на рис. 4 и 5. Рис. 6 в принципе идентичен рис. 5, с той лишь разницей, что здесь отсутствует рандомизируемая категория кандидатов и рандомизация ранговой связки. Вместо этого на курсы приняты те, чьи доходы составляли 22 доллара на одного члена семьи и ниже.

отбора и лишь служат основой для экстраполяции результатов гипотетического эксперимента с разрывом ранговой связки в данной критической точке.

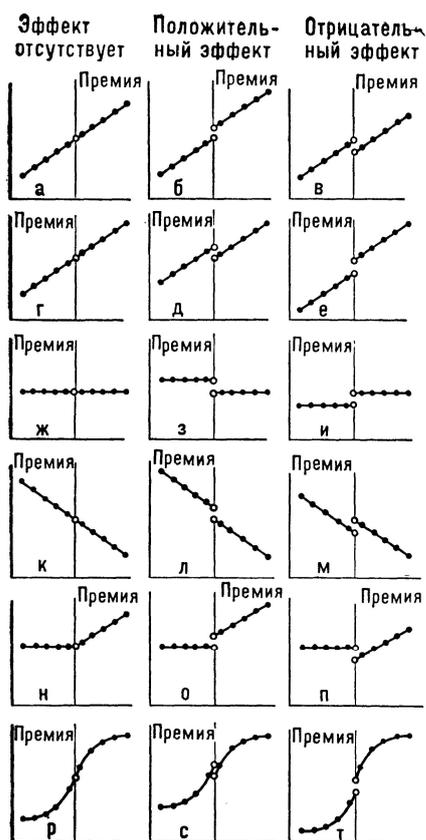
План 16 является квазиэкспериментальным в том смысле, что по сравнению со случаем рандомизации ранговой связки при интерпретации данных приходится делать предположения, менее доступные проверке. Например, необходимо предположить однородность единиц измерения по обе стороны от критической точки. Требуется также сделать ряд предположений о виде функции регрессии. Предлагаемый способ статистического анализа<sup>1</sup> состоит в подборе линий регрессии от-

<sup>1</sup> Разработан, в частности, Д. А. Суином в диссертации, подготовленной в Северо-Западном университете.



Р и с. 7. Гипотетический результат эксперимента по плану с нарушением непрерывности регрессии в условиях, аналогичных рис. 5 и 6, для случая, когда эффект воздействия полностью отсутствует.

дельно для двух сегментов, данных слева и справа от критической точки, с последующим определением величины скачка путем экстраполяции каждой кривой за критическую точку. Разность между двумя экстраполированными значениями, очевидно, будет зависеть от предполагаемой формы каждой кривой. В случае двух сигмоид, изображенных на рис. 6 и 7, использование линейной функции регрессии при наличии данных, представленных на рис. 7, даст псевдоэффект. Наряду с численной обработкой данных рекомендуется также представлять данные в виде графика и не доверять результатам статистических расчетов, если вид его указывает на правдоподобность непрерывной функции, не имеющей разрыва в критической точке. Нельзя, конечно, исключить случайного совпадения независимого скачка линии регрессии с критической точкой. Но это маловероятно, если обосновано предположение о равенстве интервалов шкалы измерения и частоты внутри интервалов распределены равномерно.



На рис. 8 приведены всевозможные гипотетические результаты применения плана 16, причем графики, помещенные слева, отвечают отсутствию эффекта, а те, что в центре и справа, — наличию положительного или отрицательного эффекта. Словом «премия» повсюду обозначена область значений критерия отбора  $k$ , соответствующая экспериментальной группе. Если премия назначается по принципу больших заслуг (рис. 8а, б), то сравнение показателей в области критической точки без учета регрессии данных конечного тестирования по  $k$  приводит к оптимистическим псевдоэффектам: на рис. 8а получившие премию в дальнейшем больше преуспевают, но не благодаря ей. Ситуация помощи нуждающимся обычно соответствует рис. 8 г, д. Здесь, если не принять во внимание наличие регрессии, программа может показаться вредной в отсутствие эффекта или неэффективной, если имеется реальный эффект.

План 16, разумеется, работает ничуть не хуже, если критерий отбора  $k$  статистически не связан с данными, полученными после воздействия (премии), как в примерах рис. 8 ж, и. В таких случаях признак, по которому производится отбор, является функциональным экви-

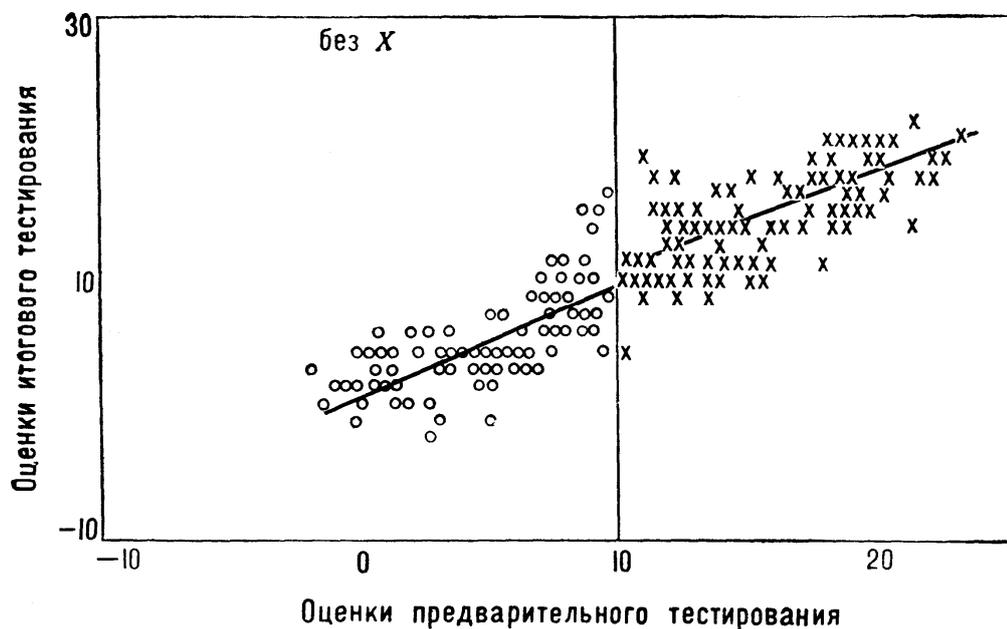
Р и с. 8. Примерные данные экспериментов типа 16.

валентом рандомизации. Очевидно, возможна и отрицательная регрессия (рис. 8к, м). Рис. 8 н, п приведены с целью подчеркнуть тот факт, что именно скачок регрессии в критической точке является признаком эффекта, тогда как различие в наклоне кривой без такого скачка еще не свидетельствует о наличии эффекта. Это становится более очевидным, если мы вспомним, что в случаях типа  $n$  эксперимент с рандомизацией ранговой связки показал бы отсутствие эффекта. Криволинейная регрессия (рис. 8 р—т) может вызвать дополнительные трудности в интерпретации результатов, поскольку из-за выборочной ошибки случай, показанный на рис. 8 р, может быть принят за случай 8 б.

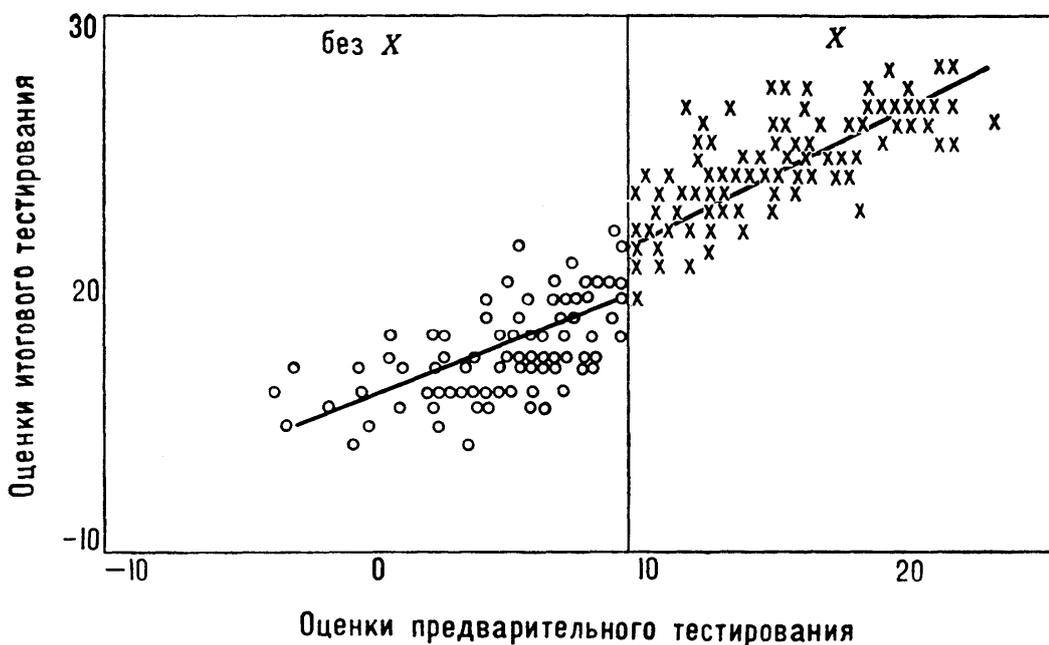
Следующая иллюстрация, рис. 9, содержит данные машинного эксперимента — имитированные индивидуальные наблюдения и построенные для них линии регрессии — в отсутствие эффекта и представляет собой более подробный вариант рис. 8 а. Рис. 10 показывает наличие эффекта. Данные предварительного тестирования генерировались путем приписывания каждому индивиду взвешенной нормальной случайной величины, игравшей роль «истинного значения», к которому добавлялась взвешенная независимая случайная «ошибка». «Истинное значение тестового показателя» в сумме с другой независимой «ошибкой» имитировали результаты итогового тестирования для случаев отсутствия эффекта (рис. 9). Эффект воздействия имитировался введением соответствующей добавки для тех «испытуемых», которые «подверглись воздействию», то есть для значений  $k$ , превышающих критическое значение.

Хотя приведенные иллюстрации относятся к случаю одной зависимой и одной независимой переменной, часто предпочтительны многомерные варианты данного эксперимента, если при этом не утрачивается четкость правила отбора. Многомерные переменные, характеризующие последующее состояние испытуемых, можно использовать разными способами. Так, можно ввести статистические поправки для одной из зависимых переменных (такой, как более поздние заработки) в целях устранения дифференциальных

эффектов различных социально-экономических фоновых переменных, кроме тех, которые были использованы при выработке количественного критерия отбора. Вместо этого можно



Р и с. 9. Данные машинного эксперимента типа 16: эффект отсутствует.



Р и с. 10. Данные машинного эксперимента типа 16: эффект имеет место.

представить результат в виде составной переменной и применить технику множественной регрессии.

Бросим теперь взгляд на строку табл. 3, соответствующую плану 16. Ввиду синхронности экспериментального и контрольного режимов фоновая стимуляция и естественное развитие

находятся под контролем. Главный эффект тестирования контролируется, поскольку оно осуществляется одинаково как в экспериментальной, так и в контрольной группе. Инструментальные погрешности могут создать известные трудности, если итоговое тестирование проводится под эгидой той же организации, которая ведала отбором кандидатов: благодарность счастливых и горечь обиденных может привести к различиям в выражении ими своих установок, к различной степени преувеличения собственных успехов. Этот недостаток присущ также истинному эксперименту с разрывом ранговой связки. Этот фактор можно проконтролировать, поручив продолжение исследования другой организации. Мы считаем, что, согласно приведенным выше соображениям, эффекты регрессии и состава групп не могут повлиять на интерпретацию результатов, даже, несмотря на неэквивалентность групп и наличие регрессии. Оба фактора контролируются путем детального представления соответствующих зависимостей, а не посредством уравнивания групп. Фактор выбывания будет сказываться, если последующее тестирование и распределение благ производится одной и той же организацией: получившие премию или место на курсах будут, по всей видимости, более охотно сотрудничать, чем те, кто оказался в контрольной группе. Между прочим, похвальное в обычных условиях стремление к репрезентативному опросу всей намеченной выборки может оказаться здесь дезориентирующим. Если, например, рассылка анкет со штампом другой организации ведет к снижению готовности сотрудничать с 90 до 50%, то экспериментатор может отнестись к этому отрицательно, так как он стремился к 100%-ному охвату получивших премию. Он готов забыть, что его действительная цель — получить поддающиеся интерпретации данные, что никакие цифры нельзя интерпретировать изолированно и что использовать данные по премированным можно лишь при наличии сопоставимой контрольной группы. По

171

этой причине, а также из-за опасности инструментальной погрешности привлечение независимой организации с научной точки зрения предпочтительнее, а 50%-ный возврат анкет из обеих групп лучше, чем 90%-ный из экспериментальной и 50%-ный из контрольной группы. Проблема выбывания остается в силе и для истинного эксперимента с разрывом связки. Для обоих экспериментальных планов угроза внутренней валидности со стороны взаимодействия состава групп и естественного развития находится под контролем. В случае квазиэксперимента это объясняется невозможностью отнести явный скачок регрессии на счет такого взаимодействия. Угроза внешней валидности со стороны взаимодействия между эффектом тестирования и  $X$  контролируется постольку, поскольку основные измерения, используемые для принятия решения об отборе кандидатов, являются частью генеральной совокупности, на которую предполагается распространить выводы.

Внешняя валидность истинного эксперимента с рандомизацией ранговой связки и нарушением непрерывности регрессии особенно зависит от взаимодействия между составом групп и  $X$ . Иначе говоря, эффект демонстрируется только для очень узкого класса значений  $k$ , равных критическому. Однако в случае квазиэксперимента возможности делать заключения, по-видимому, шире.

### **Корреляционные планы и планы EX POST FACTO**

Одним из признаков квазиэксперимента, все отчетливее выступавшим при обсуждении последних девяти планов, является то, насколько экспериментатор может управлять воздействием  $X$ , то есть вторгаться в нормальный ход событий. Разумеется, чем больше он контролирует  $X$ , тем больше он приближается к осуществлению истинного эксперимента. Это особенно справедливо в отношении планов 7 и 10. Планы 7, 10, 12, 13 (но не 13а) и 14

применимы как в случае естественно возникающих воздействий, так и в случае воздействий, намеренно вводимых экспериментатором. Там, где  $X$  остается вне контроля, эксперимент вызывал бы больше подозрений, и те, кто еще согласился бы назвать квази-

172

*экспериментом* его варианты с воздействием, контролируемым экспериментатором, не захотели бы применять этот термин к случаям неуправляемого воздействия. Не вдаваясь в обсуждение этого вопроса, подчеркнем только ценность анализа результатов неуправляемого воздействия по типу анализа экспериментальных данных по сравнению с некоторыми вводящими в заблуждение оценочными описаниями и рассуждениями, к которым слишком часто прибегают в подобных случаях.

План 15 прямо предназначен для изучения эффектов естественно возникающих  $X$ , но планы, обсуждаемые в этом параграфе (их, правда, чаще именуют не планами квазиэксперимента, а планами анализа данных), еще более интегрированы в естественной обстановке. Мы начнем наше изложение с простого корреляционного анализа, затем перейдем к двум, в общем, приемлемым планам, а в конце подробнее остановимся на исследованиях *ex post facto*, которые даже в лучшем исполнении признаются неудовлетворительными.

### Корреляция и каузальная связь

План 3 является весьма слабой разновидностью корреляционного плана, так как он предполагает сравнение лишь двух естественных единиц, различающихся не только наличием и отсутствием  $X$ , но и огромным числом иных признаков. Каждый из этих признаков может оказывать влияние на результаты тестирования и, следовательно, давать основания для гипотезы, конкурирующей с гипотезой о том, что эффект вызван  $X$ . И нам не остается ничего другого, как еще раз признать невозможность интерпретации различий между двумя естественными объектами. Распространим теперь это сравнение на большее число независимых естественных случаев наличия и отсутствия  $X$  и связанных с ними различий в  $O$ . Поскольку каждая естественная реализация  $X$  отличается от других по своим признакам, эти признаки уже в меньшей степени оправдывают соответствующие конкурентные гипотезы. Таким образом, могут быть установлены довольно внушительные корреляции — как, например, между курением и

173

раком легких. Могут ли такие данные, подобно данным эксперимента, свидетельствовать о наличии каузальной связи?

Прежде всего, отметим один положительный момент. Такие данные имеют отношение к гипотезам о каузальной связи постольку, поскольку они могут их опровергнуть. Нулевая корреляция снижает правдоподобие гипотезы. При высокой корреляции оно возрастает, ибо гипотеза избежала опровержения. Иначе говоря, корреляция не обязательно указывает на наличие причинной связи, но каузальный закон, поскольку он связан с различием между средними значениями, предполагает наличие корреляции. В любом эксперименте, где  $X$  привело к увеличению  $O$ , будет обнаружена положительная бисериальная корреляция между наличием / отсутствием  $X$  и данными итогового тестирования или приращением показателей (от предварительного обследования к последующему). Отсутствие такой корреляции может опровергнуть много простых, общих каузальных гипотез, гипотез о главных эффектах  $X$ . В этом смысле относительно недорогой корреляционный подход может служить для предварительного испытания гипотез, и те, которые его выдержат, могут быть затем подвергнуты более дорогой экспериментальной проверке. Кац, Маккоби

и Морз [57], обосновав этот подход, провели исследование, в котором влияние лидерства на производительность труда было вначале изучено корреляционным методом, а затем основная гипотеза была подвергнута экспериментальной проверке (Morse, Reimer [82]).

Внимательное ознакомление с исследованиями в педагогике показывает, что корреляционные данные скорее истолковываются в пользу существования каузальной связи, что правдоподобные конкурентные гипотезы нередко выпадают из поля зрения исследователя и что для установления временной последовательности каузальных связей важно если не управляемое воздействие  $X$ , то хотя бы развернутые во времени наблюдения. Если, например, поведение учителя и ученика скоррелировано, то в соответствии с нашими культурными стереотипами мы почти никогда не принимаем в расчет возможность влияния поведения ученика на поведение учителя. Даже когда в естественных условиях как будто выявляется внутренне присущая событиям времен-

174

ная последовательность, избирательность нашей памяти может отобразить эту причинно-следственную связь в обратном направлении. Допустим, например, было установлено, что администраторы лучших школ имеют лучшее образование и что в школах, где администраторы часто меняются, низок моральный дух персонала. Почти неизбежно мы заключим из этого, что более высокий уровень образования администраторов и стабильное руководство ведут к более высокому уровню школ. Но каузальная цепочка может быть прямо противоположной: лучшие школы (неважно, по каким причинам лучшие) могут вызывать у хорошо образованных людей желание остаться, тогда как, оказавшись в более слабой школе, они пытаются искать работу в другом месте. Точно так же лучшие школы могут создавать подходящие условия для более продолжительного пребывания администратора на своей должности. Еще чаще, чем пресловутая обратная корреляция, нас может вводить в заблуждение корреляция с третьей переменной: лица, обладающие правом решать, кто будет подвергнут  $X$ , пользуются им так, что высокие результаты достигаются и без всякого  $X$ . К этим случаям мы еще вернемся в последнем разделе, посвященном плану *ex post facto*.

Истинный эксперимент тем и отличается от условий корреляционного исследования, что процесс рандомизации разрушает любую закономерную связь между характеристиками учеников (предшествующее событие) и предъявлением им  $X$ . При наличии предварительного тестирования и возможности четко определить, кто будет подвергнут  $X$ , а кто нет, эксперименты, проведенные по планам 10 и 14, могут оказаться убедительными даже без рандомизации. Но для естественного осуществления эксперимента, в котором не проводится предварительное тестирование (имитируемого планом 6), требуются совершенно особые условия, которые почти никогда не реализуются. Но и здесь в соответствии с нашей общей установкой следует внимательно присматриваться к ситуациям, в которых можно получать поддающиеся интерпретации данные. Это ситуации, в которых  $X$  реализуется произвольно, вне всякой закономерности или связи с предшествующими событиями. В идеале решения о произвольном воздействии должны

175

быть многочисленными и взаимно независимыми. Более того, они должны подкрепляться какими-либо дополнительными данными, пусть самыми слабыми, вроде тех, которые, скажем, удастся получить посредством ретроспективной методики предварительного тестирования. Саймон [101, с. 10—61] и Уолд [137] привели доводы в пользу того, что каузальная интерпретация простой или частичной корреляции зависит от наличия правдоподобной каузальной гипотезы и отсутствия правдоподобных конкурентных гипотез, объясняющих корреляцию на другой основе.

В одном таком корреляционном исследовании настолько удачно использованы обстоятельства, что о нем следует упомянуть. Барч, Трамбо и Нэнгл [4] пользовались в качестве  $X$  наличием или отсутствием сигнала поворота от впереди идущего автомобиля, а роль зависимой переменной играл факт включения или невключения такого сигнала следующей машиной. Был продемонстрирован значимый эффект имитации, моделирования, или конформизма, что находилось в согласии с результатами многих лабораторных исследований. В отсутствие какого-либо предварительного тестирования интерпретация результатов зависит от допущения, что влияние поведения первой машины на водителя второй — единственная возможная тенденция. Опубликованные данные выглядят довольно убедительно. Отметим, однако, что любая третья переменная, влияющая на частоту сигналов, подаваемых обоими водителями, может стать правдоподобной конкурентной гипотезой. Так, если погода, условия видимости, цели поездки, обусловленные временем дня, наличие полицейских машин и т. д. воздействуют на обоих водителей и если данные собираются в условиях, различных по таким третьим переменным, корреляцию можно объяснить и независимо от влияния сигналов, исходящих от впереди идущего автомобиля. Еще лучше поддается интерпретации как «естественный эксперимент по плану б» исследование Брима [6] о влиянии пола одного ребенка на личность другого ребенка в семье с двумя детьми. Пол определяется почти что случайно. Насколько известно, он не коррелирует ни с семейными ни с социальными, ни с генетическими детерминантами личности. Одновременная детерминация пола одного

176

ребенка и личности другого ребенка третьей переменной, как и обратная каузальная связь личности ребенка с полом его брата или сестры, не представляют собой правдоподобных объяснений, конкурирующих с основной каузальной интерпретацией этих интересных данных.

### **Ретроспективное предварительное тестирование**

Во многих воинских формированиях в военное время лица одного и того же звания и специальности получают различные назначения случайно, без учета их особых привилегий, предпочтений или способностей. Поэтому сравнение социальных установок белых, попавших в подразделения со смешанным расовым составом, с установками тех, кто оказался в частях с однородным составом, может представлять интерес для установления соответствующих каузальных отношений. Мы, разумеется, должны не пренебрегать этими данными, а искать дополнительные, чтобы отклонить правдоподобные конкурентные гипотезы, отдавая себе отчет в остающихся источниках невалидности. В этом примере итоговые интервью содержали не только вопросы об отношении к неграм в данное время (в смешанных подразделениях оно оказалось более благоприятным), но предполагали припоминание установок, имевших место до получения данного назначения. Это «ретроспективное предварительное тестирование» обнаружило отсутствие различия между двумя группами, что увеличило вероятность отсутствия различий и до назначения.

Подобный анализ сыграл важную роль в исследовании Дойча и Коллинз [30], сравнивавших жителей интегрированных и сегрегированных кварталов, в которых квартиры распределялись в такой период, когда из-за недостатка жилья люди принимали соответствующие предложения более или менее независимо от их расовых установок. При наличии данных только итогового тестирования обнаруженные ими различия между двумя группами могли быть отнесены на счет первоначальных социальных установок. Утверждение, что опыт расовой интеграции приводит к более благожелательному отношению, стало более убедительным, когда ретроспективное исследование показало отсутствие

различий между двумя группами в их воспоминаниях о своих первоначальных установках. В силу аутистических факторов, искажающих, как известно, воспоминания и данные интервью, такие результаты не носят решающего характера.

По-настоящему следовало бы, конечно, провести предварительные интервью и в случайном порядке распределить испытуемых по обеим группам. Такие исследования, несомненно, будут выполнены. Но пока мы не располагаем лучшими данными, результаты Дойча и Коллинз, включая ретроспективное предварительное тестирование, являются ценным добавлением к экспериментальному знанию в этой трудной области.

Читателю следует обратить внимание на то, что вероятная систематическая ошибка памяти состоит в искажении прежних установок и согласовании их с сегодняшними установками или с установками, которые испытуемый считает социально желательными. Таким образом, здесь ошибка памяти скорее затушает значимый эффект  $X$ .

В исследованиях влияния обучения в колледже на студентов путем сравнения установок новичков и старшекурсников желательно использовать методику ретроспективного тестирования, которое будет служить частичным заслоном против конкурентных гипотез об эффектах фоновых воздействий, селективного выбывания и изменений в начальном составе группы. (Это не значит, что мы одобряем дальнейшее повторение таких поэтапных исследований, так как сейчас больше нужны лонгитюдинальные исследования, аналогичные исследованиям Ньюкомба [84], которые предполагали повторные измерения в течение четырех лет в сочетании с повторными обследованиями по методу поперечных срезов, то есть по плану 15. Если же нужно срочно сделать диссертацию, то для этого лучше выбрать другую тематику.)

### Панельные исследования

В определенных однородных исследованиях респондентам нередко предлагается самим определить, были ли они подвержены  $X$ . В таких случаях корреляция между воздействием и результатом итогового тестиро-

вания оказывается смещенной не только в силу наличия общего источника погрешности ( $X$  получают те, кто и без  $X$  дал бы высокий результат  $O$ ), но и из-за искажений памяти по отношению к  $X$ , еще больше усиливающих ложный эффект (Stouffer [120, с. 356]). Хотя эти исследования и обнаруживают каузальную связь, оправдывающую затраты на рекламу (то есть демонстрируют корреляцию между ответами на вопросы типа «Смотрели ли Вы эту программу?» и «Купили ли Вы это изделие?»), они являются тривиальным свидетельством о наличии эффекта. Они привносят новый фактор, угрожающий внутренней валидности результатов, то есть систематическую ошибку в определении того, кто получил  $X$ , который мы не намерены вносить в наши таблицы.

Большим достижением методологии исследования общественного мнения явилось применение панельного метода — повторных интервью с теми же лицами. В лучшем случае, когда  $X$  (например, показ кинофильма или проведение консультации) вклинивается между двумя волнами интервью или анкетирования, панельное исследование представляет собой слабый вариант плана 10 с единственным  $X$ . Следует, однако, иметь в виду, что в социологии это важное методическое нововведение сопровождается ошибочным традиционным анализом. Так, интерпретация «таблицы текучести» (Glock [41]), как отмечают Кэмпбелл и Клэйтон [14], затруднена из-за смешивания каузальных гипотез с регрессионными эффектами. Даже при анализе приращений зависимых переменных в

группах, подвергшихся и не подвергшихся  $X$ , в понятиях предварительного и итогового тестирования остается менее явный источник систематической ошибки. В таких панельных исследованиях факт предъявления респонденту  $X$  (например, показ кинофильма, направленного против предрассудков) устанавливается во второй волне интервью двухэтапного панельного исследования. Схема такого плана выглядит следующим образом:

$$\left( \begin{array}{c} O \\ - \\ O \end{array} \right) - \left( \begin{array}{c} X \\ - \\ O \end{array} \right)$$

Панельное исследование с двумя волнами интервью (неприемлемо).

179

Здесь большими скобками отмечено наличие  $O$  или  $X$  на одном и том же этапе с одним интервью, а вопросительный знак указывает на неоднозначность отнесения респондентов к группам с  $X$  и без  $X$ . В отличие от плана 10 этот план не позволяет однозначно определить, кто принадлежит к экспериментальной, а кто к контрольной группе. Как в худших реализациях плана 10,  $X$  коррелирует с данными предварительного тестирования (лица с минимальными предрассудками более склонны посмотреть фильм). Более того, даже если  $X$  в действительности не влияет на результаты  $O$ , корреляция между  $X$  и данными последующих тестирований будет более высокой, чем между  $X$  и данными предварительных тестирований, просто потому, что они ( $X$  и  $O$ ) осуществляются в одном и том же интервью. Как показывает общий опыт исследований, связанных с тестированием и измерениями, в отношении любых двух пунктов одного и того же вопросника наблюдается сильная тенденция к более высокой корреляции ответов между собой, чем в отношении ответов на те же вопросы, но включенные в разные вопросники. Стокфорд и Бисселл [119] обнаружили, что даже в рамках одной анкеты ответы на смежные (соседние по положению) вопросы имеют более высокую корреляцию, чем ответы на вопросы, отдаленные друг от друга. Корреляция между измерениями, выполненными в один день, как правило, выше, чем между измерениями, сделанными в разные дни. В упомянутом панельном исследовании (Glock [41]) два интервью были проведены с интервалом в восемь месяцев. Увеличению корреляции внутри одного интервью и уменьшению корреляции тех же пунктов между обоими интервью способствовали изменения в составе интервьюеров. Неизбежные ошибки последних и неточности высказываний опрашиваемых при повторной идентификации данных первоначальных респондентов приводят к тому, что некоторые пары интервью на самом деле оказываются принадлежащими разным людям. Получаемая в итоге более высокая корреляция между  $X$  и данными последующего интервью означает, что регрессия от  $X$  к результатам этого интервью меньше, чем к результатам предварительного опроса, и поэтому различия между группами в последующем интервью окажутся большими, чем при предва-

180

рительном обследовании. Это приведет (даже в отсутствие эффекта в популяции) к ложному эффекту прироста показателей для тех, кто отнес себя к числу получивших воздействие, и снижению для тех, кто считал, что не получал воздействие. Подобный результат обычно ошибочно принимается за подтверждение гипотезы об эффекте  $X$  (более подробно см. Campbell, Clayton [14]).

Чтобы обойти этот ложный источник повышенной корреляции, принадлежность респондентов к группе, которой предъявляется  $X$ , можно устанавливать независимо от

интервью или посредством отдельной промежуточной волны опросов. В последнем случае, даже если и имеется ошибка памяти относительно прохождения  $X$ , это не должно искусственно увеличивать корреляцию  $X$  с данными последующей проверки по сравнению с корреляцией между  $X$  и данными предварительного обследования. Вот как выглядит такой план:

$$\begin{pmatrix} O \\ - \\ O \end{pmatrix} - \begin{pmatrix} X \\ - \\ -? \end{pmatrix} - \begin{pmatrix} O \\ - \\ O \end{pmatrix}$$

### Анализ ex post facto

Выражением «эксперимент ex post facto» обозначаются попытки имитировать экспериментальное исследование по плану 3 путем попарного уравнивания исходного состава групп на основании сведений об испытуемых до введения  $X$ . Сам метод и его название впервые ввел Чэпин (Chapin, Queen [20]). Он же [19], а также Гринвуд [46] в дальнейшем широко его использовали. Хотя эти работы относятся скорее к области социологии, чем к педагогике, и хотя мы считаем, что этот анализ подчас приводит к ошибкам, здесь уместно рассмотреть его. Этот метод представляет собой одну из наиболее серьезных попыток приблизиться к квазиэкспериментальному исследованию.

В одном типичном исследовании ex post facto (Chapin [19, с. 99—124]) в роли  $X$  выступало школьное обучение (в частности, окончание средней школы), а в роли  $O$  — индивидуальные интервью, дававшие информацию об успехе в жизни и социальной адаптированности обследуемых 10 лет спустя. Уравнивание

181

групп производилось на основании школьных архивов (в аналогичных, но еще более слабых исследованиях факты о состоянии до воздействия брались из итоговых опросов). Из первоначальных данных следовало, что окончившие школу не только лучше преуспевали, но также имели более высокие отметки в 5—8 классах, лучшее окружение, были моложе, их родители занимали более высокие должности и т. д. Очевидно, эти предпосылки могли способствовать не только окончанию средней школы, но и дальнейшим успехам.

Добавило ли что-либо обучение в школе к лучшей стартовой позиции детей по сравнению с этими факторами? «Решение», предложенное Чэпином, состояло в изучении групп учеников, уравненных по всем этим факторам, но отличавшихся по признаку окончания школы. Введение очередного признака уравнивания сокращало итоговое различие между группами, которые подвергались и не подвергались воздействию, но, когда попарное уравнивание было завершено, различие все еще оставалось значимым. Из этого Чэпин заключил, хотя и осторожно, что среднее образование дало свой эффект. Изначальное число 2127 человек сократилось до 1194 (с ними удалось провести интервью, и о них сохранились нужные архивные записи). Уравнивание сократило число пригодных для использования случаев до 46, по 23 на группу окончивших и не окончивших школу — менее 4% опрошенных. Чэпин правильно отмечает, что 46 сопоставимых случаев лучше, чем 1194 несопоставимых, и подкрепляет это аргументами, аналогичными тем, которые мы приводили в пользу большей важности внутренней валидности по сравнению с внешней. Трагедия в том, что его 46 случаев все еще оставались несопоставимыми и, кроме того, даже если встать на его точку зрения, в сокращении данных не было необходимости.

Его уравнивание было *недостаточным* по двум причинам. Во-первых, попарное уравнивание приводит в действие механизм дифференциальной регрессии, который

создает различия в результатах, имеющие то же направление, что и наблюдавшиеся Чэпином (см. замечания Р. Л. Торндайка [125], а также обсуждение проблемы уравнивания в плане 10). Направление псевдоэффекта, обусловленного регрессией показателей к

182

групповым средним после попарного уравнивания, в этом случае вполне однозначно, поскольку различия по факторам уравнивания для преуспевших и не преуспевших направлены в ту же сторону, что и различия между окончившими и не окончившими школу. Каждый признак, детерминирующий подверженность  $X$ , даже в отсутствие  $X$  детерминирует и  $O$ . Все переменные уравнивания коррелируют с  $X$  и  $O$  в одном и том же направлении. Хотя это и не обязательно должно иметь место по каждой переменной во всех исследованиях *ex post facto*, тем не менее такая картина наблюдается в большинстве, если не во всех опубликованных примерах использования этого метода. Ошибки регрессии и сокращения числа случаев можно избежать, применяя современные статистические методы, которые позволяют избавиться от ошибки уравнивания в плане 10. Переменные уравнивания могли бы быть использованы как сопутствующие переменные в многомерном ковариационном анализе. Насколько мы можем судить, такой анализ показал бы отсутствие эффектов, оказавшихся якобы значимыми в исследовании, представленном Чэпином. Однако применение ковариационного анализа к подобным задачам подвергается критике.

Но имеется и второй, в сущности неустранимый, источник неэквивалентности групп в исследовании Чэпина. Гринвуд [46] называет его самоотбором испытуемых в группы, подвергающиеся и не подвергающиеся воздействию. Наличие  $X$  является закономерным результатом многочисленных предпосылок. Для выбывания из школы до ее окончания, как мы знаем, имеются бесчисленные причины, помимо тех шести факторов, по которым производилось уравнивание. Мы можем с большой уверенностью предположить, что большинство из них повлияло на последующий успех независимо от их действия через  $X$ . Это приведет к еще большей неэквивалентности групп, помимо той, которая вызвала ошибку регрессии. Даже проводя ковариационный анализ исходных данных и результатов  $O$ , мы можем интерпретировать наличие значимого эффекта  $X$  только после учета *всех* совместно действующих переменных уравнивания.

183

### ЗАКЛЮЧИТЕЛЬНЫЕ ЗАМЕЧАНИЯ

Всякое справочное руководство представляет собой концентрированный текст, и дальнейшая его концентрация может ввести в заблуждение. Поэтому мы хотим предостеречь от соблазна использовать табл. 1, 2 и 3 в качестве сжатого эквивалента этого очерка. Их цель — упорядочить изложение и избавить его от излишних повторов.

Конкретные плюсы, минусы и вопросительные знаки постоянно ставились с оговорками и являлись лишь неточными обобщениями итогов соответствующего обсуждения. Для любой конкретной реализации каждого плана картина плюсов и минусов, скорее всего, будет отличаться от соответствующей строчки в таблице. Таблицы лучше всего использовать как канву для добросовестного анализа конкретных деталей эксперимента при его планировании. И конечно же, мы не собираемся подменять догмы одного-двух приемлемых планов эксперимента догмой тринадцати. Напротив, нам хотелось бы способствовать укреплению открытой, поисковой ориентации на разработку новых способов получения данных и выявление слабых сторон традиционных методик.

Подведем теперь основные итоги сказанному. В этой работе мы обсудили возможности, существующие в области построения или планирования эксперимента, и уделили особое

внимание проблемам контроля посторонних переменных и угрозам валидности эксперимента. Мы проводили различие между внутренней и внешней валидностью, или возможностью обобщения результатов. Оценка качества 16 экспериментальных планов и их вариантов велась по восьми классам факторов внутренней и четырем — внешней валидности. Три плана были определены как доэкспериментальные и были использованы в первую очередь для иллюстрации факторов валидности, которые нужно контролировать. Три плана названы «истинными», десять планов — квазиэкспериментальными, то есть такими, в которых отсутствует оптимальный контроль, но которые стоит использовать в тех случаях, когда применение лучшего плана невозможно. При интерпретации результатов квазиэксперимента особенно важна проверка перечисленных

184

12 факторов валидности. На протяжении всей работы обращалось внимание на возможность творческого использования особенностей конкретной ситуации исследования при планировании конкретных испытаний каузальных гипотез.

#### ЛИТЕРАТУРА

1. Allport F. H. The influence of the group upon association and thought. *J. exp. Psychol.*, 1920, 3, 159—182.
2. Anastasi Anne. Differential psychology. N. Y., Macmillan, 1958.
3. Anderson N. H. Test of a model for opinion change. *J. abnorm. soc. Psychol.*, 1959, 59, 371—381.
4. Barch A. M., Trumbo D., Nangle J. Social setting and conformity to a legal requirement. *J. abnorm. soc. Psychol.*, 1957, 55, 396—398.
5. Boring E. G. The nature and the history of experimental control. *Amer. J. Psychol.*, 1954, 67, 573—589.
6. Brim O. G. Family structure and sex role learning by children: A further analysis of Helen Koch's data. *Sociometry*, 1958, 21, 1—16.
7. Brolyer C. R., Thorndike E. L., Woodyard Ella. A second study of mental discipline in high school studies. *J. educ. Psychol.*, 1927, 18, 377—404.
8. Brownlee K. A. Statistical theory and methodology in science and engineering. N. Y., Wiley, 1960.
9. Brunswik E. Perception and the representative design of psychological experiments. Berkeley, Univ. of California Press, 1956.
10. Campbell D. T. Factors relevant to the validity of experiments in social settings. *Psychol. Bull.*, 1957, 54, 297—312.
11. Campbell D. T. Methodological suggestions from a comparative psychology of knowledge processes. *Inquiry*, 1959, 2, 152—182.
12. Campbell D. T. Recommendations for APA test standards regarding construct, trait, or discriminant validity. *Amer. Psychologist*, 1960, 15, 546—553.
13. Campbell D. T., Boruch R. F. Making the Case for Randomised Assignment to treatments by Considering the Alternatives: Six Ways in which Quasi-Experimental evaluations in compensatory Education tend to underestimate effects. — In: Evaluation and experiment. Eds. by C. A. Bennett and A. A. Lumsdaine, Acad. press, 1975.
14. Campbell D. T., Clayton K. N. Avoiding regression effects in panel studies of communication impact. *Stud. pub. Commun.*, 1961, 3, 99—118.
15. Campbell D. T., Fiske D. W. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol. Bull.*, 1959, 56, 81—105.
16. Campbell D. T., McCormack Thelma H. Military experience and attitudes toward authority, *Amer. J. Sociol.*, 1957, 62, 482—490.

185

17. Cane V. R., Heim A. W. The effects of repeated testing: III. Further experiments and general conclusions. *Quart. J. exp. Psychol.*, 1950, 2, 182—195.
18. Cantor G. N. A note on a methodological error commonly committed in medical and psychological research. *Amer. J. ment. Defic.*, 1956, 61, 17—18.
19. Chapin F. S. Experimental designs in sociological research. N. Y., Harper, 1947 (Rev. ed., 1955).
20. Chapin F. S., Queen S. A. Research memorandum on social work in the depression. N. Y. *Social Science Research Council. Bull.* 39, 1937.
21. Chernoff H., Moses L. E. Elementary decision theory. N. Y., Wiley, 1959.
22. Cochran W. G., Cox Gertrude M. Experimental designs. N. Y., Wiley, 1957.
23. Collier R. M. The effect of propaganda upon attitude following a critical examination of the propaganda itself. *J. soc. Psychol.*, 1944, 20, 3—17.
24. Collier R. O., Jr. Three types of randomization in a two-factor experiment. Minneapolis, Author, 1960.
25. Cornfield J., Tukey J. W. Average values of mean squares in factorials. *Ann. math. Statist.*, 1956, 27, 907—949.
26. Cox D. R. Some systematic experimental designs. *Biometrika*, 1951, 38, 312—323.
27. Cox D. R. The use of a concomitant variable in selecting an experimental design. *Biometrika*, 1957, 44, 150—158.
28. Cox D. R. Planning of experiments. N. Y., Wiley, 1958.
29. Crook M. N. The constancy of neuroticism scores and self-judgments of constancy. *J. Psychol.*, 1937, 4, 27—34.
30. Deutsch M., Collins Mary E. Interracial housing: A psychological evaluation of a social experiment. Minneapolis, Univ of Minnesota Press, 1951.
31. Duncan C. P., O'Brien R. B., Murray D. C., Davis L., Gilliland A. R. Some information about a test of psychological misconceptions. *J. gen. Psychol.*, 1957, 56, 257—260.
32. Ebbinghaus H. Memory. N. Y., Teachers Coll., Columbia Univer., 1913 (Original: «Über das Gedächtnis», Leipzig, 1885).
33. Edwards A. L. Experimental design in psychological research. N. Y., Rinehart, 1960.
34. Farmer E., Brooks R. C., Chambers E. G. A comparison of different shift systems in the glass trade. Rep. 24, Medical Research Council, Industrial Fatigue Research Board. London, His Majesty's Stationery Office, 1923.
35. Feldt L. S. A comparison of the precision of three experimental designs employing a concomitant variable. *Psychometrika*, 1958, 23, 335—353.
36. Ferguson G. A. Statistical analysis in psychology and education. N. Y., McGraw-Hill, 1959.
37. Фишер Р. А. Статистические методы для исследователей. М., 1958.
38. Fisher R. A. The design of experiments. London, Oliver & Boyd, 1935.

39. Fisher R. A. The arrangement of field experiments. *J. Min. Agriculture*, 1926, 33, 503—513; also in: R. A. Fisher. Contributions to mathematical statistics. N. Y., Wiley, 1950.
40. Glickman S. E. Perseverative neural processes and consolidation of the memory trace. *Psychol. Bull.*, 1961, 58, 218—233.
41. Glock C. Y. Some applications of the panel method to the study of social change. — In: Lazarsfeld P. F., Rosenberg M. (eds.). the language of social research. Glencoe, Ill., Free Press, 1955, p. 242—249.
42. Glock C. Y. The effects of re-interviewing in panel research. Duplicated research report. Author, 1958.
43. Good C. V., Scates D. E. Methods of research. N. Y., Appleton-Century-Crofts, 1954.
44. Grant D. A. Analysis-of-variance tests in the analysis and comparison of curves. *Psychol. Bull.*, 1956, 53, 141—154.
45. Green B. F., Tukey J. W. Complex analyses of variance: General problems. *Psychometrika*, 1960, 25, 127—152.
46. Greenwood E. Experimental sociology: A study in method. N. Y., King's Crown Press, 1945.

47. Guetzkow H., Kelly E. L., McKeachie W. J. An experimental comparison of recitation, discussion, and tutorial methods in college teaching. *J. educ. Psychol.*, 1954, 45, 193—207.
48. Hammond K. R. Representative vs. systematic design in clinical psychology. *Psychol. Bull.*, 1954, 51, 150—159.
49. Hanson N. R. Patterns of discovery. Cambridge. Eng., Univ. Press, 1958.
50. Hovland C. I., Janis I. L., Kelley H. H. Communication and persuasion. New Haven, Conn., Yale Univ. Press, 1953.
51. Hovland C. I., Lumsdaine A. A., Sheffield F. D. Experiments on mass communication. Princeton, Princeton Univ. Press, 1949.
52. Information and Education Division, U. S. War Department. Opinions about Negro infantry platoons in white companies of seven divisions. — In: Newcomb T. M., Hartley E. L. (eds.). Readings in social psychology. N. Y., Holt, 1947, p. 542—546.
53. Johnson P. O. Statistical methods in research. N. Y., Prentice-Hall, 1949.
54. Johnson P. O., Jackson R. W. Modern statistical methods: Descriptive and inductive. Chicago, Rand McNally, 1959.
55. Jost A. Die Assoziationsfestigkeit in ihrer Abhängigkeit von der Verteilung der Wiederholungen. *Z. Psychol. Physiol. Sinnesorgane*, 1897, 14, 436—472.
56. Kaiser H. F. Directional statistical decisions. *Psychol. Rev.*, 1960, 67, 160—167.
57. Katz D., Maccoby N., Morse Nancy C. Productivity, supervision, and morale in an office situation. Ann Arbor, Survey Research Center, Univ. of Michigan, 1951.
58. Kempthorne O. The design and analysis of experiments. N. Y., Wiley, 1952.
59. Kempthorne O. The randomization theory of statistical inference. *J. Amer. Statist. Ass.*, 1955, 50, 946—967; 1956, 51, 651.
60. Kempthorne O. The design and analysis of experiments, with some reference to educational research, — In: Collier R. O.,

- Elam S. M. (eds.). Research design and analysis: The second annual Phi Delta Kappa symposium on educational research. Bloomington, Ind., Phi Delta Kappa, 1961, p. 97—133.
61. Kendall M. G., Buckland W. R. A dictionary of statistical terms. London. Oliver & Boyd, 1957.
  62. Kennedy J. L., Uphoff H. F. Experiments on the nature of extra-sensory perception. III. The recording error criticisms of extra chance scores. *J. Parapsychol.*, 1939, 3, 226—245.
  63. Kerr W. A. Experiments on the effect of music on factory production. *Appl. Psychol. Monogr.*, 1945, 5.
  64. Lana R. E. Pretest-treatment interaction effects in attitudinal studies. *Psychol. Bull.*, 1959, 56, 293—300 (a).
  65. Lana R. E. A further investigation of the pretest-treatment interaction effect. *J. appl. Psychol.*, 1959, 43, 421—422. (b).
  66. Lana R. E., King D. J. Learning factors as determiners of pretest sensitization. *J. appl. Psychol.*, 1960, 44, 189—191.
  67. Lindquist E. F. Statistical analysis in educational research. Boston, Houghton Mifflin, 1940.
  68. Lindquist E. F. Design and analysis of experiments in psychology and education. Boston, Houghton Mifflin, 1953.
  69. Lipset S. M., Lazarsfeld P. F., Barton A. H., Linz J. The psychology of voting: An analysis of political behavior— In: Lindzey G. (ed.). Handbook of social psychology. Cambridge, Mass., Addison-Wesley, 1954, p. 1124—1175.
  70. Lord F. M. The measurement of growth. *Educ. psychol. Measmt*, 1956, 16, 421—437.
  71. Lord F. M. Further problems in the measurement of growth. *Educ. psychol. Measmt*, 1958, 18, 437—451.
  72. Lubin A. The interpretation of significant interaction. *Educ. psychol. Measmt*, 1961, 21, 807—817.
  73. Maxwell A. E. Experimental design in psychology and the medical sciences. London, Methuen, 1958.
  74. McCall W. A. How to experiment in education. N. Y., Macmillan, 1923.
  75. McNeemar Q. A critical examination of the University of Iowa studies of environmental influences upon the I. Q. *Psychol. Bull.*, 1940, 37, 63—92.

76. McNeMar Q. Psychological statistics. N. Y., Wiley, 1962.
77. McNeMar Q. On growth measurement. *Educ. psychol. Measmt.* 1958, 18, 47—55;
78. Meehl P. E. Clinical versus statistical prediction. Minneapolis, Univ. of Minnesota Press, 1954.
79. Monroe W. S. General methods: Classroom experimentation. — In: G. M. Whipple (ed.). *Yearb. nat. Soc. Stud. Educ.*, 1938, 37, Part II, 319—327.
80. Mood A. F. Introduction to the theory of statistics. N. Y., McGraw-Hill, 1950.
81. Moore H. T. The comparative influence of majority and expert opinion. *Amer. J. Psychol.*, 1921, 32, 16—20.
82. Morse Nancy C., Reimer E. The experimental change of a major organizational variable. *J. abnorm. soc. Psychol.*, 1956, 52, 120—129.
83. Myers J. L. On the interaction of two scaled variables. *Psychol. Bull.*, 1959, 56, 384—391.

84. Newcomb T. M. Personality and social change. N. Y., Dryden, 1943.
85. Neyman J. Indeterminism in science and new demands on statisticians. *J. Amer. Statist. Ass.*, 1960, 55, 625—639.
86. Nunnally J. The place of statistics in psychology. *Educ. psychol Measmt*, 1960, 20, 641—650.
87. Page E. B. Teacher comments and student performance: A seventy-four classroom experiment in school motivation. *J. educ. Psychol.*, 1958, 49, 173—181.
88. Pearson H. C. Experimental studies in the teaching of spelling. *Teachers Coll. Rec.*, 1912, 13, 37—66.
89. Piers Ellen V. Effects of instruction on teacher attitudes: Extended control-group design. Unpublished doctoral dissertation, George Peabody Coll., 1954. Abstract in: *Bull. Maritime Psychol. Ass.*, 1955, 53—56.
90. Popper K. R. The logic of scientific discovery. N. Y., Basic Books, 1959.
91. Rankin R. E., Campbell D. T. Galvanic skin response to Negro and white experimenters. *J. abnorm. soc. Psychol.*, 1955, 51, 30—33.
92. Reed J. C. Some effects of short term training in reading under conditions of controlled motivation. *J. educ. Psychol.*, 1956, 47, 257—264.
93. Rogers C. R., Dymond Rosalind F. Psychotherapy and personality change. Chicago, Univ. of Chicago Press, 1954.
94. Rosenthal R. Research on experimenter bias. Paper read at Amer. Psychol. Ass., Cincinnati, Sept., 1959.
95. Roy S. N., Gnanadesikan R. Some contributions to ANOVA in one or more dimensions: I and II. *Ann. Math. Statist.*, 1959, 30, 304—317, 318—340.
96. Rozeboom W. W. The fallacy of the null-hypothesis significance test. *Psychol. Bull.*, 1960, 57, 416—428.
97. Sanford F. H., Hemphill J. K. An evaluation of a brief course in psychology at the U. S. Naval Academy. *Educ. psychol. Measmt*, 1952, 12, 194—216.
98. Scheffé H. Alternative models for the analysis of variance. *Ann. math. Statist.*, 1956, 27, 251—271.
99. Selltitz Claire, Jahoda Marie, Deutsch M., Cook S. W. Research methods in social relations, N. Y., Holt-Dryden, 1959.
100. Siegel Alberta E., Siegel S. Reference groups, membership groups, and attitude change. *J. abnorm. soc. Psychol.*, 1957, 55, 360—364.
101. Simon H. A. Models of man. N. Y., Wiley, 1957.
102. Smith H. L., Hyman H. The biasing effect of interviewer expectations on survey results. *Publ. opin. Quart.*, 1950, 14, 491—506.
103. Sobol M. G. Panel mortality and panel bias. *J. Amer. Statist. Ass.*, 1959, 54, 52—68.
104. Solomon R. L. An extension of control group design. *Psychol. Bull.*, 1949, 46, 137—150.
105. Sorokin P. A. An experimental study of efficiency of work under various specified conditions. *Amer. J. Sociol.*, 1930. 35, 765—782.

106. Stanley J. C. Statistical analysis of scores from counterbalanced tests. *J. exp. Educ.*, 1955, 23, 187—207.
107. Stanley J. C. Fixed, random, and mixed models in the analysis of variance as special cases of finite model III. *Psychol. Rep.*, 1956, 2, 369.
108. Stanley J. C. Controlled experimentation in the classroom. *J. exp. Educ.*, 1957, 25, 195—201. (a)
109. Stanley J. C. Research methods: Experimental design. *Rev. educ. Res.*, 1957, 27, 449—459. (b)
110. Stanley J. C. Interactions of organisms with experimental variables as a key to the integration of organismic and variable-manipulating research. — In: Huddleston Edith M. (ed.). Yearb. Nat. Counc. Measmt used in Educ., 1960, 7—13.
111. Stanley J. C. Analysis of a doubly nested design. *Educ. psychol. Measmt*, 1961, 21, 831—837. (a)
112. Stanley J. C. Studying status vs. manipulating variables — In: Collier R. O., Elam S. M. (eds.). Research design and analysis: The second Phi Delta Kappa symposium on educational research. Bloomington, Ind.: Phi Delta Kappa, 1961, p. 173—208. (b)
113. Stanley J. C. Analysis of unreplicated three-way classifications, with applications to rater bias and trait independence. *Psychometrika*, 1961, 26, 205—220. (c)
114. Stanley J. C. Analysis-of-variance principles applied to the grading of essay tests. *J. exp. Educ.*, 1962, 30, 279—283.
115. Stanley J. C., Beeman Ellen Y. Interaction of major field of study with kind of test. *Psychol. Rep.*, 1956, 2, 333—336.
116. Stanley J. C., Wiley D. E. Development and analysis of experimental designs for ratings. Madison, Wisconsin, Authors, 1962.
117. Stanton F., Baker K. H. Interviewer-bias and the recall of incompletely learned materials. *Sociometry*, 1942, 5, 123—134.
118. Star Shirley A., Hughes Helen M. Report on an educational campaign: The Cincinnati plan for the United Nations. *Amer. J. Social.*, 1950, 55, 389—400.
119. Stockford L., Bissell H. W. Factors involved in establishing a merit-rating scale. *Personnel*, 1949, 26, 94—116.
120. Stouffer S. A. (Ed.) The American soldier. Princeton, Princeton Univ. Press, 1949, Vols. I, II.
121. Stouffer S. A. Some observations on study design. *Amer. J. Social*, 1950, 55, 355—361.
122. Thistlethwaite D. L., Campbell D. T. Regression-discontinuity analysis: An alternative to the ex post facto experiment. *J. educ. Psychol*, 1960, 51, 309—317.
123. Thorndike E. L., Woodworth R. S. The influence of improvement in one mental function upon the efficiency of other functions. *Psychol. Rev.*, 1901, 8, 247—261, 384—395, 553—564.
124. Thorndike E. L., McCall W. A., Chapman J. C. Ventilation in relation to mental work. *Teach. Coll. Contr. Educ.*, 1916, 78.
125. Thorndike R. L. Regression fallacies in the matched groups experiment. *Psychometrika*, 1942, 7, 85—102.
126. Underwood B. J. Experimental psychology. N. Y., Appleton-Century-Crofts, 1949.

127. Underwood B. J. An analysis of the methodology used to investigate thinking behavior. Paper read at New York Univ. Conf. on Human Problem Solving. April, 1954. (См. также Hovland C. I., Kendler H. H. The New York University Conference on Human Problem Solving. *Amer. Psychologist*, 1955, 10, 64—68.)
128. Underwood B. J. Interference and forgetting. *Psychol. Rev.*, 1957, 64, 49—60. (a)
129. Underwood B. J. Psychological research. N. Y., Appleton-Century-Crofts, 1957. (b)
130. Underwood B. J., Richardson J. Studies of distributed practice. XVIII. The influence of meaningfulness and intralist similarity of serial nonsense lists. *J. exp. Psychol.*, 1958, 56, 213—219.
131. Watson R. I. Psychology of the child. N. Y., Wiley, 1959.
132. Wilk M. B., Kempthorne O. Fixed, mixed, and random models. *J. Amer. Statist. Ass.*, 1955, 50, 1144—1167; Corrigenda, *J. Amer. Statist. Ass.*, 1956, 51, 652.

133. Wilk M. B., Kempthorne O. Some aspects of the analysis of factorial experiments in a completely randomized design. *Ann. Math. Statist.*, 1956, 27, 950—985.
134. Wilk M. B., Kempthorne O. Non-additivities In a Latin square design. *J. Amer. Statist. Ass.*, 1957, 52, 218—236.
135. Windle C. Test-retest effect on personality questionnaires. *Educ. psychol. Measmt*, 1954, 14, 617—633.
136. Winer B. J. *Statistical principles in experimental design*. N. Y., McGraw-Hill, 1962.
137. Wold H. Causal inference from observational data. A review of ends and means. *J. Royal Statist. Soc., Sec. A.*, 1956, 119, 28—61.
138. Wyatt S., Fraser J. A., Stock F. G. L. Fan ventilation in a humid weaving shed. Rept. 37, Medical Research Council, Industrial Fatigue Research Board. London, His Majesty's Stationery Office, 1926.
139. Zeisel H. *Say it with figures*. N. Y., Harper, 1947.

## **Научный вывод, артефакты и контроль<sup>1</sup>**

### **I. Логика вывода**

Если бы мы не отказались от дефиниционного операционализма нашего недавнего прошлого, то не знали бы проблем, которым посвящены работы об артефактах в исследованиях поведения. Наши экспериментальные модели и измерительные процедуры трактовались бы как *дефиниционное* представление наших теоретических понятий. Рассуждая о них как о дефинициях, мы закрывали бы глаза на то, что они подвержены погрешностям, систематически несовершенны, а порой и ошибочны.

Дефиниционный операционализм действительно вызвал у некоторых состояние не критической самоуспокоенности и иллюзию материализации тестовых показателей, но, к счастью, практики научного исследования в большинстве своем либо не часто вступали в контакт с философией науки, либо не слишком верили в нее, чтобы позволить сбить себя с толку. Когда логические позитивисты определяли интеллект по работе Бине 1916 г., Термен уже приступил к пересмотру своего теста, стараясь сделать его менее подверженным погрешностям и более точным средством измерения интеллекта, — эта цель явно свидетельствует о том, что для него этот тест *не был* простой дефиницией. Точно так же физик, работающий с таким измерительным прибором, как, скажем, гальванометр, отдает себе отчет

---

<sup>1</sup> Campbell D. T. Prospective: Artifact and Control. — In: *Artifact in Behavioral Research*. Ed. by R. Rosenthal, R. L. Rosnow. N. Y., Academic Press, 1969, p. 351—382. © Academic Press, Inc., 1969.

в том, что на практике прибор этот не способен дать совершенную картину разностей электрических потенциалов из-за влияний, оказываемых гравитацией, трением, инерцией,

магнитным полем и т. п. (см., например, Wilson [95]). Хотя в скомпенсированном и корректном экспериментальном плане эти источники ошибок могут быть сведены к минимуму, в *теоретическом* плане гальванометр известен как прибор, подверженный систематическим ошибкам, и установление этого явилось в свою очередь историческим итогом накопления научных знаний, а не плодом логического наития.

Коль скоро теперь уже признано, что дефиниционный операционализм и прочие аксессуары логического позитивизма приводят к ошибочному пониманию, как же следует осмыслить наши затруднения на пути познания и при этом еще и извлечь философский смысл из работ, послуживших прообразом для данного подхода? Наиболее приемлемым для меня представляется направление, разрабатываемое Карлом Поппером [73,74], (Campbell [24]), а также то общее, что присуще и его работам, и работам Поляни [72], Тулмина [92, 93], Куна [60] и Куайна [75], хотя они, возможно, менее всего склонны допускать что-либо подобное. Я попытаюсь представить один из аспектов этого направления, хотя и воспользуюсь метафорами, не относящимися к числу общепринятых.

Вслед за Поппером я отдаю должное Юму как логике и отвергаю его как психолога-индуктивиста. Юм привлек внимание к «недоразумению с индукцией» — к тому факту, что научные обобщения являются логически не доказанными или недоказуемыми. Хотя большинство современных философов принимают этот факт как простую техническую деталь, как простое утверждение о неприменимости аналитической логики к условной истине, Поппер с присущей ему пронизательностью увидел в нем фундаментальное ограничение. Научные истины не только не доказаны *логически*, они лишены достоверности также и в любом другом отношении — индуктивном, эмпирическом, научном или имплицативном. Тем не менее, они в каком-то смысле «установлены». Лучшие из теорий если и не «подтверждены», то, по крайней мере «подкреплены».

193

Логика релевантна утверждению о ситуации. «Недоразумение с индукцией» состоит, по существу, в том, что наука пользуется необоснованным (невалидным) логическим доказательством, допуская ошибку «нераспределенной середины» или «ошибку относительно следствия». Но, будучи *невалидным*, это доказательство не *бесполезно*.

Логическое доказательство в науке имеет следующую форму:

Если теория Ньютона *A* истинна, то должно наблюдаться следующее: морские приливы и отливы происходят с периодом *B*, орбита Марса имеет форму *C*, траектория полета пушечного ядра приобретает форму *D*.

Наблюдение подтверждает *B*, *C* и *D*. Следовательно, теория Ньютона *A* истинна.



Р и с. 1. Схема Эйлера.

Мы можем увидеть ошибочность этого доказательства, представив его в виде Эйлеровых кругов (см. рис. 1). Отмеченная невалидность вызвана существованием заштрихованной

области, то есть других возможных объяснений наблюдавшихся фактов  $B$ ,  $C$  и  $D$ . Но этот силлогизм не бесполезен. Если сделаны наблюдения, несовместимые с  $B$ ,  $C$  и  $D$ , то тем самым обоснованно отрицается истинность ньютоновской теории  $A$ . Рассмотренное доказательство очень напоминает, таким образом, процесс просеивания, где предсказания и наблюдения служат для того, чтобы отбрасывать наиболее неадекватные теории. Кроме того, если предсказания подтверждаются, теория остается одним из возможных истинных объяснений. Выявление этой асимметрии между логически обоснованным отрицанием и логически неубедительным подтверждением и является

194

основным моментом в тезисе Поппера о возможности фальсификации.

Теперь этот трюизм приводится в элементарных изложениях индуктивной логики без ссылок на Поппера (см., например, Hempel [54], Salmon [84]). В упомянутом процессе есть и другой критический пункт, на котором сосредоточили свои усилия критики Поппера: в действительности наблюдения все же подтверждают предсказания. Выше предполагалось, что такой подход мог бы иметь и имел место. На данном уровне возможность фальсификации и возможность подтверждения логически более симметричны. И наблюдения, если они выполнены с достаточной точностью, на этом уровне *всегда* фальсифицируют квантифицированное предсказание. На данном уровне пределы точности наблюдений, которые ученые реально признают как допустимые, являются функцией социальной системы, детерминированной уровнем развития данной науки, достигнутым уровнем экспериментального контроля и остротой конкуренции со стороны других теорий. Так, для предсказанного Эйнштейном искривления световых лучей, идущих от звезд, вблизи поверхности Солнца во время затмения 1919 г. предсказанная величина в 1,745" была «подтверждена» полученными при наблюдениях величинами 1,61", 1,98", 1,72", 2,2" и 2,0".

Рассмотрим несколько более подробно круги Эйлера и ту связь, которая существует между подтвержденными предсказаниями и истинностью или правдоподобием теории. Нам никуда не уйти от того факта, что мы не можем доказать теорию. Мы должны работать, не выходя за рамки ограничений, указанных в схеме. Что мы, как ученые, можем сделать — так это попытаться каким-либо *практическим* способом «опустошить» заштрихованную область, уменьшить ее, насколько это возможно. Мы делаем это, максимально увеличивая по возможности число, масштаб и точность подтвержденных предсказаний. Чем больше их число и чем они точнее, тем меньше число возможных альтернативных отдельных объяснений, даже если это число по-прежнему остается в каком-то смысле бесконечным.

Более важно, что мы на деле не уделяем должного внимания самой логической возможности существования альтернативных теорий, простому логическому

195

факту существования некоей заштрихованной области. Тулмин хорошо отметил этот момент: «Философы иногда утверждают, что конечное число эмпирических наблюдений всегда можно объяснить в терминах бесчисленного множества гипотез. В основе этого замечания лежит простое наблюдение, согласно которому через всякое конечное множество точек можно провести бесчисленное множество математических кривых. Если бы «объяснение» заключалось в одном только проведении кривых, эта доктрина имела бы некоторое влияние на практику научного исследования. На самом деле перед ученым стоит совсем другая проблема: его задача состоит, как правило, в том, чтобы в интеллектуальной ситуации, предъявляющей ряд требований, приспособить некоторое новое открытие к унаследованным идеям, не создавая без нужды угроз интеллектуальным достижениям

своих предшественников. Это проблема совсем иного порядка сложности, чем простое проведение кривых. Здесь речь идет отнюдь не о выборе из бесчисленного множества имеющихся возможностей: может потребоваться гениальное озарение, чтобы вообразить хотя бы только одну такую возможность» (Toulmin [92, с. 113—115]).

Только когда имеются по-настоящему разработанные альтернативные объяснения, то есть когда кое-что известно о содержании заштрихованной области, только тогда встают вопросы об обоснованности теорий, предсказания которых подтверждены. Именно потому, что с теорией Ньютона не соперничала ни одна должным образом разработанная альтернативная теория, она в течение 200 лет рассматривалась как бесспорно истинная даже такими критически настроенными эпистемологами, как Кант. Заштрихованная область была пуста в каком угодно практическом смысле. Однако факт последующего отказа от теории Ньютона в пользу теории Эйнштейна заставляет увидеть в Юмовом анализе научной истины уместно поставленную проблему научной индукции.

В действительности положение еще сложнее. Когда такая теория, как теория Ньютона, вообще не находит себе хоть сколько-нибудь достойных конкурентов и когда она с неподражаемым изяществом предсказывает великое множество явлений, мы, как правило, прощаем

196

ей кое-какие неверные предсказания. Так, как подчеркивает Кун [60], во времена Ньютона были известны систематические ошибки в предсказаниях, например, относительно прецессии перигелия Меркурия, которые могли бы быть расценены как доказательство несостоятельности теории Ньютона, если бы в те времена существовала теория Эйнштейна. Более правдоподобной выглядит картина конкуренции между разработанной ранее и более успешно подкрепленной теориями при объяснении сравнения образов (Campbell [23]).

Таким образом, единственный процесс, посредством которого может утверждаться научная теория, — это процесс «ограничения правдоподобных конкурентных гипотез». Поскольку эти гипотезы никогда нельзя перечислить заранее и поскольку они обычно сугубо специфичны и требуют своеобразных способов ограничения, это последнее неизбежно представляет собой довольно-таки несовершенную и неубедительную процедуру. Однако логический анализ затруднений на нашем пути к научному знанию (от Юма до Поппера) убеждает нас в том, что это и есть самое большее, что мы можем сделать, что это и есть наш геркулесов, если не сизифов, труд.

## **II. МЕТОДОЛОГИЯ И ПРАВДОПОДОБНЫЕ КОНКУРЕНТНЫЕ ГИПОТЕЗЫ**

Множество правдоподобных конкурентных гипотез, ставящих под сомнение валидность теории, поставляется другими хорошо разработанными теориями. Они нас здесь не интересуют. Другие правдоподобные конкурентные гипотезы возникают в связи с аппаратом конкретного экспериментального исследования. Эти последние и становятся объектом «методологии». Такие гипотезы часто весьма специфичны, не разработаны и не увязаны с теориями широкого масштаба. Тем не менее, они представляют собой гипотетические эмпирические законы, связывающие между собой две или большее число переменных. В той мере, в какой они становятся методологически существенными, это хорошо установленные и, таким образом, правдоподобные эмпирические закономерности. В этом скорее, чем в логической необходимости, кроется источ-

197

ник их авторитетности. Нельзя отрицать, что в каком-то смысле они еще и логичны (Petri [71]). Существует, однако, столь бесчисленное множество факторов, контроль за которыми представляет логическую необходимость, что их полный охват был бы немислимым делом. С логической точки зрения заштрихованная область беспредельна. И только на практическом уровне мы подходим к ее опустошению.

Этот эмпирический статус методологических требований представляется мне важным моментом, о котором упоминают незаслуженно редко. Мы слишком часто принимаем научный метод так, как будто он получен из логики, априорно по отношению к науке и помимо нее. В одной из ранних своих работ, защищая экспериментальные планы, не предусматривающие предварительного тестирования (Campbell [19]), я писал, как если бы это было логически очевидным, что, если экспериментируют только с популяциями, прошедшими предварительное тестирование, нет никаких оснований для обобщений в отношении популяций, не прошедших такового. Теперь мне представляется, что акценты были расставлены неверно. Источник авторитетности следовало искать не в логике, а в эмпирическом правдоподобии, в вероятном законоподобном характере предварительно выявленных взаимодействий. Следуя одной только логике, было бы одинаково нелогично пытаться распространять сделанные выводы на другие группы людей, другие периоды времени в прошлом и будущем или другие условия, чем-то отличающиеся от тех, которые исследовались в исходном эксперименте. Среди всей обширной массы логических ограничений именно это оказалось особенно убедительным в силу своей эмпирической привлекательности. Гипотетические законы, связанные с этим обстоятельством, как это показал Лана [см. 83], теперь уже совсем не кажутся такими правдоподобными, какими они казались ранее, и в результате мы, без сомнения, увидим, что экспериментальным планам, которые обходятся без предварительного тестирования, будет уделяться меньшее внимание либо будут использоваться иные основания. Более того, когда налицо эффекты взаимодействия, они выступают как нарушающие факторы, приводящие к недооценке другого проверяемого в исследовании за-

кона. Они не относятся к явлениям сенсбилизации, которые были бы способны порождать псевдоэффекты, крайне нетипичные для естественных ситуаций.

В данном случае «правдоподобие» конкурентной гипотезы, как это утверждалось мной ранее, основывалось частично на апелляции к обыденному знанию о психологических процессах. Однако в моей публикации было одно положение, которое представляется теперь ошибочным и которое могло рассматриваться как убедительное. Если я воспроизведу его здесь, то это кое-что добавит к ряду тщательно отобранных Ланой эмпирических результатов. Я сообщал о том, что в исследовании «Организация Объединенных Наций в поле зрения Цинциннати» осведомленность о чрезвычайно интенсивных воздействиях с помощью средств массовой коммуникации или влияние их были обнаружены лишь на одной панельной выборке, прошедшей предварительное обследование. Исследование состояло из двух частей. В основном сравнивались две отдельные, но уравненные в случайном порядке выборки по 1000 человек каждая, взятые одна до, а другая после кампании, и в результате никаких существенных различий обнаружено не было. Эти результаты и были опубликованы в упомянутой мной статье. Из устного сообщения одного из авторов этой статьи мне стало известно о вспомогательном исследовании, проведенном по заказу Бюро прикладных социальных исследований округа Колумбия. Оно заключалось в повторном интервьюировании выборки, проходившей предварительное тестирование. Из упомянутого устного сообщения я узнал (или полагал, что узнал) итоги этого все еще не опубликованного исследования и сообщил о них в своей работе. Они казались как нельзя более уместной иллюстрацией, именно тем, что было необходимо для формулировки моего положения. Спустя несколько лет Клер Зеллиц,

редактируя «Методы исследования в сфере социальных отношений» (Selltiz, Jahoda, Deutsch, Cook [88]), приложила немало усилий, чтобы уточнить это сообщение, но так и не достигла успеха и поэтому упомянула о повторном обследовании как о гипотетической возможности. В конце концов нашелся дубликат отчета, принадлежащий Глоку [46], в котором были представлены совсем иные результаты: между выбор-

кой, прошедшей предварительное тестирование, и выборкой, не прошедшей такого тестирования, последующее тестирование не обнаружило не только никаких значимых различий, но и никаких тенденций в направлении сенсбилизации или усиления влияний коммуникации. Эти факты приводились в более поздней публикации (Campbell, Stanley [26])<sup>1</sup>, но излагались в ней не столь подробно, как это было сделано здесь. Этот случай иллюстрирует, я полагаю, зависимость аргументации от предполагаемых законов, от гипотетических эмпирических закономерностей и, таким образом, соответствие эмпирического факта установлению некоего «необходимого» научного контроля. Одно важное следствие из этой аргументации состоит в том, что у нас вызывают беспокойство не вообще неудачные попытки контроля, а лишь те из них, которые допускают *действительно правдоподобные* конкурентные гипотезы, закономерности, сопоставимые — или даже превосходящие их — по степени научной доказанности с законами, для проверки которых был спланирован наш эксперимент.

Таким образом, наши нынешние стандарты экспериментальных планов представляют собой *научное достижение, эмпирический продукт*, а не дарованную логикой милость. Они представляют, как правило, верифицированные гипотетические законы — как сказал бы философ, условные, дескриптивные, синтетические и вследствие этого доступные коррекции «истины», а не истины логические или аналитические. Мы, психологи, считаем, что «контрольная группа» — это условие, необходимость которого является аксиомой. Не так уж часто обращают внимание на то, что она, равно как и ее аналоги, фактически никогда не фигурировала в физике, химии и физиологии XIX в., из которых мы заимствуем наши методологические модели. По свидетельству Боринга [8], она была изобретена не ранее 1907 г. для проверки одной правдоподобной конкурентной гипотезы, вполне психологической по своему характеру, а именно: предварительные тестирования должны вызывать улучшение результатов даже в отсутствие экспериментальных воздействий (то есть то,

---

<sup>1</sup> См. перевод этой работы в данной книге (с. 34—191).

что мы могли бы обозначить как одно из основных последствий тестирования в отличие от последствий взаимодействия при тестировании, которые описаны выше). Если бы не регулярные подтверждения того факта, что такого рода последствия имеют место на практике, не было бы причин вводить контрольную группу. В экспериментальной психологии дело вряд ли обстоит таким образом, но, как отмечает Лана, оно может принимать подобный оборот в исследованиях процесса убеждения. Наш типичный синхронный план эксперимента с предварительным и последующим тестированием и контрольной группой предполагает также контроль и других конкурентных гипотетических объяснений изменений (Campbell [19]). Но для психолога-экспериментатора, изучающего запоминание бессмысленных слогов, ни одно из них не может служить достаточно правдоподобным объяснением улучшений в работе памяти, так что он обычно обходится без них. Даже в случае самого примитивного плана (эксперимент на одной группе с предшествующим и последующим тестированием) не было бы никакой

нужды в предварительном тестировании, если бы не эмпирические факты, согласно которым значения корреляций между первым и повторным тестированиями больше нуля и индивиды не тождественны друг другу.

Кроме обычной контрольной группы, имеются другие контрольные группы, формирование которых обусловлено нуждами исследования и для которых эмпирическое обоснование более очевидно. Так, при исследованиях изменений, наступающих в результате удаления участков коры, использование контрольной группы, которая подвергается ложной операции, обусловлено эмпирическим фактом хирургического шока. По мере накопления данных, проливающих свет на природу хирургического шока, ссылки на него начинают рассматриваться как крайне неправдоподобное объяснение многих явлений, имеющих место после удаления участков коры и электростимулирования мозга. Использование контрольной группы, получающей плацебо, отражает хорошо установленный закон, касающийся терапевтического эффекта, возникающего при наличии у пациента уверенности, что он подвергается лечебному воздействию [83]. Когда фармацевти-

201

ческое исследование оценивается в связи со столь обобщенной переменной, как «болезнь — здоровье», без этой меры пока не обойтись. Когда же речь идет о более специфических эффектах, в ней часто нет особой необходимости.

В привлечении контрольной группы, получающей плацебо, о чем не знает («двойной слепой эксперимент») тот, кто его вводит, проявляется эмпирический закон, отражающий влияние на результаты эксперимента, оказываемое верой экспериментатора в целебные свойства прописываемого лекарства. Этот пример уводит нас в область фактов, столь хорошо представленных Розенталем [82].

Вклад, который вносят авторы работ об артефактах в исследованиях поведения в науку о психологическом методе, состоит главным образом в установлении необходимости в новых контрольных группах. Орн ввел контрольные группы с варьированием требований [см. 83]. Розенталь вводит двукратное воздействие с варьируемым уровнем ожидания — повышенным и пониженным. Розенберг говорит о контрольных группах с боязнью оценки и рекомендует проводить эксперимент таким образом, чтобы скрыть от испытуемых связь между воздействием и последующей проверкой. Мак-Гайр и Розенталь приводят аргументы в пользу того, что опасности, угрожающие валидности, возникают закономерно, показывая, каким образом такого рода переменные, поначалу порождающие методические проблемы контроля, могут привести к проблемам, оказывающимся центральными.

### **III. ТИПОЛОГИЯ АРТЕФАКТОВ, ПОГРЕШНОСТЕЙ, УГРОЗ ВАЛИДНОМУ ВЫВОДУ**

Хотя угрозы валидному выводу, или артефакты, могут быть обязаны своим происхождением любому аспекту экспериментального процесса и хотя полная типология их невозможна, полезно, быть может, наметить некоторые типы артефактов, появляющихся с известной периодичностью.

#### **A. Смешанные факторы экспериментального воздействия: основные эффекты**

Мы можем рассматривать с этой целью в качестве аспектов экспериментального воздействия *все* характе-

202

ристики, по которым различаются между собой экспериментальная и контрольная группы. Неизбежно, что многие из них не имеют никакого отношения к той теоретической переменной, которой мы манипулируем в эксперименте. Они обусловлены средствами исследования и являются «инструментальными случайностями». Такие характеристики произвольны — в том смысле, что эксперимент мог быть осуществлен как-то иначе, но и неизбежны — в том смысле, что, если бы не эти случайности, вместо них должны были бы фигурировать другие. Никакие «чистые воздействия» невозможны.

Каждая из этих нерелевантных характеристик является потенциальным конкурентным объяснением эффекта воздействия. Большой класс артефактов — именно такого типа. Примером могут служить специальные указания Розенберга о том, что релевантным объектом манипулирования в обсуждавшемся им эксперименте был не когнитивный диссонанс, а боязнь оценки. Еще более обобщенным примером является критика Розенталем большого числа экспериментов, в которых фундаментальной переменной, вызывающей эффект, была не какая-нибудь конкретная характеристика воздействия, а различные ожидания экспериментаторов.

Возможны два способа контролирования смещения данного типа. Во-первых, введение новой контрольной группы или использование *контрольной группы, подвергающейся дополнительному воздействию*. Иными словами, воздействие на контрольную группу модифицируется таким образом, чтобы оно включало максимум того, чему подвергалась ранее одна только экспериментальная группа. Контрольные группы, подвергающиеся ложной операции или получающие плацебо, относятся к данному типу. Таким образом, создается контрольная группа с заданной «боязнью оценки», или «требованиями», или ожиданиями экспериментатора, эквивалентными тем, которые характерны для экспериментальной группы. Общий знаменатель возрастает настолько, что экспериментальная и контрольная группы по этому признаку больше не различаются. Во-вторых, в ряде случаев имеет место установка на поиск новых форм осуществления эксперимента, в которых переменная, связанная с основной теорией, очищается от этой конк-

ретной конкурентной переменной, то есть «уравновешивание» осуществляется за счет известной модификации экспериментальной группы. Эта форма, если рассматривать ее как контрольную меру, а рiогi предполагает экономию усилий, ибо никогда не исключена возможность того, что непротиворечивые с виду результаты объясняются действием двух отдельных нерелевантных факторов, различных в каждой экспериментальной ситуации. Нам следует оставить без внимания и эту возможность, пока не будут разработаны какие-нибудь конкретные правдоподобные версии.

#### **Б. Смешанные аспекты воздействия: эффекты взаимодействия**

Может случиться, что какой-то подлинный эффект действия основной теоретической переменной приобретает специфическую форму (или подавляется) в результате использования конкретных технических компонентов. В этом случае также возникает так много различных возможностей, что мы уделяем внимание только вполне разработанным и правдоподобным гипотезам такого рода. Более того, мы скорее даже не примем во внимание или сочтем неправдоподобным какой-то из эффектов взаимодействия, чем какой-то из основных эффектов. По-видимому, основанием для этого отчасти может служить то обстоятельство, что основные эффекты в большей мере поддаются контролю. Возможно, более важным следует признать полученное путем вывода весьма широкое обобщение, согласно которому основные эффекты более вероятны, чем эффекты взаимодействия. Это

положение могло бы служить аналогом или, может быть, даже частью индуктивной предпосылки Милля, утверждавшего, что природа упорядочена. Возможно, что с дескриптивной точки зрения такое обобщение действительно и, на мой взгляд, имело бы смысл скрупулезное исследование приемов дисперсионного анализа в докторских диссертациях (составляющих менее смещенную выборку, чем опубликованные исследования). Но даже если оно не является дескриптивным по отношению к природе, какой мы ее находим, оно дескриптивно по отношению к тем аспектам природы, которые доступны знанию — той несовершенной

204

выборке, на которой по необходимости сосредоточиваются наука и более простые познавательные процессы. Под знанием мы здесь понимаем, в частности, некие пригодные к употреблению, поддающиеся повторному воспроизведению общие закономерности, наблюдаемые в ситуациях и условиях, которые не являются идентичными. Если взаимодействия высших порядков, обладающие пространственной, временной спецификой, всегда оказываются значимыми, то обобщение невозможно, а следовательно, невозможно ни знание, ни наука. Успешно установленный основной эффект представляет собой куда более широкое обобщение, чем эффект взаимодействия. Основанием для переградуировки параметров измерения по большей части является поиск возможностей квантификации, которая переводит большую часть регулярностей в основные эффекты.

Конкурентные гипотезы этого типа *не доступны* контролю путем использования контрольной группы, подвергающейся дополнительному воздействию; контроль за ними может быть осуществлен путем изменения экспериментального воздействия. Отметим, что этот последний подход, будучи здесь контрольной мерой, также а priori предполагает экономию усилий и содержит презумпцию предпочтения основных эффектов. Ибо всегда есть вероятность того, что очевидное общее подтверждение определенного закона в двух различных ситуациях было не чем иным, как случайным совпадением двух отдельных частных эффектов взаимодействия.

#### **В. Фоновые взаимодействия**

Под фоном (background) здесь имеются в виду характеристики, общие для экспериментальной и контрольной групп. К нему неизбежно относятся многие характеристики, которые не отражены в теории, и, по всей вероятности, они более многочисленны в социальных науках, чем в физических. Все они являются потенциальными источниками эффектов взаимодействия — либо с теми аспектами воздействия, которые имеют прямое отношение к проверяемой теории, либо с каким-то нерелевантным фактором, примешанным к воздействию.

205

И опять-таки их так много, что мы можем уделить внимание только правдоподобным и хорошо разработанным конкурентным гипотезам.

Гипотезы взаимодействия, относящиеся к этой, к предыдущей, а также к ряду нижеперечисленных категорий (например, составу групп), представляют в каком-то смысле не столь серьезную угрозу валидности, как гипотезы, принадлежащие к самой первой категории. Они представляют собой всего лишь потенциальные *ограничения* в отношении общности какого-нибудь закона, уже установленного для одной совокупности условий. Только в том случае, когда эта одна совокупность условий является «искусственной» и когда нас интересует, прежде всего, возможность распространения наших обобщений на иные совокупности условий, нежели эта искусственная лабораторная

ситуация, такие угрозы валидности вызывают у нас беспокойство. Такова по своей природе сенсбилизация, возникающая, возможно, в ходе предварительных тестирований при исследованиях процесса убеждения. Следует признать, что изящные научные исследования процессов убеждения, которые ограничиваются изучением аудиторий, прошедших предварительное тестирование, были бы значительным научным достижением, даже если бы они не имели большого практического значения. Ведь физические науки сплошь рядом занимаются предсказаниями только в лабораторных условиях, хотя, конечно, если иметь в виду их действительно впечатляющие достижения, они на основе одной искусственной лабораторной ситуации делают предсказания относительно других ситуаций, совершенно иных по своей структуре.

И в этом случае контроль с помощью контрольной группы, подвергающейся дополнительному воздействию, невозможен и общий контроль обеспечивается лишь путем изменения экспериментального воздействия.

### **Г. Взаимодействие с характеристиками популяции**

Эта статья написана в контексте изучения «истинных экспериментов», а не «квазиэкспериментов» (Campbell, Stanley [26]), и, следовательно, отличительные признаки популяций выступают здесь не как побоч-

206

ные источники основных эффектов, то есть не как различия между экспериментальной и контрольной группами, а, напротив, как потенциальные ограничения степени общности законов, действие которых наблюдалось при исследовании какой-нибудь конкретной популяции. Мы, социальные психологи, рискуем унаследовать от социологии ложный идеал *super-ego*, предписывающий искать решение проблемы в нахождении репрезентативной выборки из некоей генеральной совокупности, взятой в соответствии с теоретическими, соображениями, может быть из всего человечества., (Наше требование рандомизации как средства достижения эквивалентности между экспериментальной и контрольной группами, выбранными в сугубо ограниченных частных условиях, не следует смешивать с тем значением, которое придает рандомизации социолог, видя в ней средство достижения репрезентативности какой-то конкретной популяции [26].)

Это, конечно, идеал, не достигнутый на практике. Но он настолько несовместим с нашим представлением о науке, что от него следует отказаться даже в нашей философии науки. Одно соображение по поводу временного параметра поможет показать крайнюю несостоятельность этого идеала. В физических науках была доказана разумность презумпции, согласно которой не происходит взаимодействия с временем (не считая суточных, лунных, сезонных и прочих циклов). Что же касается социальных наук, то рассмотрение потенциально релевантных характеристик популяций показывает, что изменения во времени (например, при сравнительном изучении студентов колледжа на протяжении 30 лет) ведут к различиям столь же значительным, что и синхронные различия между социальными классами и субкультурами. Для репрезентативности выборки из генеральной совокупности, на которую мы намерены распространить полученные данные, необходимо было бы произвести репрезентативный отбор во времени, а это, очевидно, вещь невозможная.

Для науки более типична история, которая произошла с Николсоном и Карлайслом. Взяв в мае 1800 г. образец воды в районе Сохо (образец очень частный, характерный для местных условий и данного периода) и опустив в него весьма специфичный кусок медной

207

проволоки, по которой пропускаться электрический ток (локализованный именно в этом куске), они получили водород на одном электроде, кислород — на другом, и, недолго думая, распространили этот результат на всю воду в мире на веки веков. Это было, разумеется, гипотетическое обобщение, а не доказанный факт. К настоящему времени проведено немало исследований, выясняющих влияние «примесей» на гидролиз, но и они проводились на крайне специфичных образцах. Идея репрезентативного отбора всех вод мира или хотя бы Англии никогда не фигурировала даже в качестве идеала. Само понятие «примесей», разделения состава воды на «чистое» вещество и чужеродные компоненты никогда не появилось бы на свет, если бы в случае с водой имел место репрезентативный отбор. В успешно развивающихся науках обобщения никогда еще не были «индуктивными» в смысле суммирования фактов, наблюдавшихся в пределах, в которых проводятся обобщения, напротив, они всегда были гипотетическими, даже если и руководствовались ранее открытыми законами. Ограничения в отношении обобщения возникают из-за проверки первоначального дерзкого обобщения, осуществляемой нерепрезентативными способами. Ученые считали, что гидролиз осуществляется поистине повсюду и всегда, пока не было показано, что бывает и иначе.

При таком подходе раз уж что-то было установлено, то не было бы нужды обосновывать правомерность успешно разрабатываемой «психологии второкурсников» американских колледжей, или студенток Северо-Западного университета, или белых мышей. В этом случае предполагалось бы, что для всех людей или для всех позвоночных во все времена действуют одни и те же могущественные законы, пока конкретные приложения этого предположения не окажутся ошибочными. Мы уже сейчас находимся на этой последней стадии, но даже здесь ответ заключается не в репрезентативной выборке особей или популяций школьников. Типичный путь науки — это путь исследований, руководствующихся теорией и опирающихся на количественные измерения, как это имеет место при сравнительном изучении приматов, значительно различающихся в эволюционном плане.

208

Таким образом, было бы прекрасным достижением, если бы мы располагали законосообразной психологией испытуемых-добровольцев (см., например, Rosenthal, Rosnow [83]), даже если бы она и не стала еще наукой, универсальность которой доказана. Однако и в этом случае, как только была бы разработана какая-нибудь конкретная правдоподобная гипотеза, устанавливающая ограничения для обобщений, которые нам очень хотелось бы сделать в отношении популяций, состоящих не из добровольцев, мы попытались бы установить над ней контроль. Нам хотелось бы не только распространить наши выводы на нелабораторные популяции ради прикладной науки, но, работая в области экспериментальной социальной психологии, мы надеемся на чисто научных основаниях перенести обобщения в сферу неэкспериментальных социальных наук. В связи с проблемой испытуемого-добровольца следует отметить еще и тот факт, что правдоподобная гипотеза о взаимодействии затрагивает не просто какую-то одну переменную воздействия, но очень большой класс переменных, так как испытуемые-добровольцы в *любом* эксперименте будут показывать результаты, которые, как они полагают, хочет получить экспериментатор. Подобная гипотеза действительно довольно опасна, так что, получив эмпирическое подтверждение, она может вызвать у нас желание сменить популяцию при проведении основных исследований.

#### **Д. Смешанные аспекты измерения: основные эффекты**

Всякое измерительное устройство, как и всякое воздействие, представляет собой сложный комплекс со многими посторонними с теоретической точки зрения техническими

компонентами. Измеренный эффект воздействия может быть обязан своим происхождением одной из этих посторонних деталей.

Такие артефакты в измерениях вызвали к жизни обширную литературу. Имеются, несомненно, сотни исследований, посвященных установкам на ответы при предъявлении вопросников, тестов по измерению социальных установок, тестов личности и т. д. (см., например, Cronbach [33, 34], Roger [79], Campbell, Siegman, Rees [27]). Обширная литература посвящена

209

также проблеме социальной желательности (Edwards [39], Block [12]). По поводу оценок существуют исследования галоэффектов и теории личности, имплицитно рассматривающие личность как оценщика. Имеются артефакты в показателях разногласий в диадах, межличностного восприятия и сходства образов (Cronbach [35], Corsini [29], Silverman [89]). Без аналогичных проблем не обходится и использование иного рода мер, включая наблюдения, выявление следов, регистрационные акты (Webb et al. [94]).

До сих пор подобные артефакты упоминались скорее в литературе, критически освещающей интерпретацию корреляций и межгрупповых различий, а не в связи с критическим анализом экспериментов. Так, критики указывают, что источник различий в показателях авторитарности по  $F$ -шкале, существующих между социальными классами, следует искать не в собственно авторитарном, а в интеллектуальном (Christie [28]) компоненте шкалы или в имплицитной установке на ответ, которую она формирует, однако такого рода критических замечаний в связи с исследованиями изменения установки, применяющими ту же  $F$ -шкалу, так и не появилось.

Для этих проблем контроль достигается путем применения многочисленных различных по своим техническим или методическим компонентам мер (Campbell, Fiske [25], Campbell [21]). В большинстве лабораторных экспериментов в этом плане может быть сделано гораздо больше, чем делается обычно, и притом при меньшем увеличении затрат на исследование, чем это имело бы место при повторении воздействий в методологических целях. Вероятно, не все, что делается, предается огласке, потому что использование многочисленных мер создает опасность расхождения результатов, о чем довольно-таки неприятно сообщать.

#### **Е. Смешанные аспекты измерения: взаимодействия с процессами воздействия**

Если даже установленное при измерении изменение обусловлено спецификой данного теоретически релевантного аспекта измерения, реакция может быть тем не менее обусловлена побочными компонентами метода: на-

210

блюдавшаяся реакция может быть, таким образом, характерна именно для данного конкретного способа измерения. Чтобы проконтролировать такого рода правдоподобную конкурентную гипотезу, следует опять-таки использовать альтернативные измерительные средства. (Отметим, что вопрос о взаимодействии *релевантных*, компонентов предварительного измерения с процессом воздействия обсуждался выше.)

#### **IV. КОНТРОЛИРОВАНИЕ АРТЕФАКТОВ**

В предыдущем разделе было представлено несколько различных форм контроля. Это: 1) контрольные группы, подвергаемые дополнительному воздействию в духе традиций

ложных операций и опытов с плацебо, 2) повторение воздействия с использованием иных методов и 3) применение многочисленных методов измерения. В настоящем разделе к этому обсуждению добавляются еще три пункта, более общих по своей природе.

#### **А. Контролирование правдоподобных конкурентных гипотез посредством дополнительного варьирования**

Этим заголовком обозначена очень общая методика частичного контроля или контроля за выводом, которая может быть использована во многих ситуациях, где прямой или полный контроль невозможен. Хотя она применяется главным образом в квазиэкспериментальных условиях, к ней можно прибегать и в экспериментальных ситуациях. С ней связано одно достойное упоминания следствие, согласно которому можно иной раз добиться большей ясности вывода, намеренно понизив качество какой-то части собранных данных. Начнем с примера.

Одно из правдоподобных конкурентных объяснений результатов, полученных нами при изучении культурных различий в отношении подверженности оптической иллюзии (Segall et al. [87]), состояло в том, что антропологи, участвовавшие в исследовании, предъявляли визуальные задачи по-разному. Чтобы проконтролировать этот момент, мы совершили шаг, который в настоящее время представляется мне на удивление смелым: мы решили намеренно снизить качество по-

211

ловины наших данных, собранных в условиях самого строгого контроля. Согласно инструкции, листочки с материалом теста следовало предъявлять в вертикальном положении на расстоянии примерно 1,25 м от глаз испытуемого. Эту инструкцию не так-то легко выполнить, если тот же самый человек должен еще и регистрировать результаты. В Эванстоне наша выборка включала 200 человек, опрошенных подряд при обходе домов. Половина данных была получена по всем правилам, тогда как другой половине испытуемых материал предъявлялся так, как если бы он лежал на поверхности стола, то есть на расстоянии примерно 45 см от глаз и в горизонтальном положении. Предполагалось, что в этом последнем случае результаты проиграны в чистоте, но укажут на вероятную тенденцию отклонения. Эти две формы предъявления материала действительно порождали различия в результатах, но не большие и ни в коей мере не достаточные для того, чтобы объяснить культурные различия, которые были в 5—7 раз больше. В итоге наши выводы были более убедительны, чем в том случае, если бы все эванстонские данные были высшего качества.

Приведем еще один пример из того же исследования. После того как была собрана основная масса данных, появилась публикация, указывавшая на то, что одним из возможных конкурентных объяснений были различия во времени предъявления. Мы «проконтролировали» эти различия, набрав в Эванстоне еще одну выборку и использовав два времени экспозиции: одно — очень короткое, а другое — заведомо превышающее любое время экспозиции, какое только могло иметь место для какой бы то ни было выборки. Здесь, как и в предыдущем случае, полученные различия были слишком малы, чтобы объяснить большие культурные различия.

В своей книге, посвященной контролю качества данных, Нейролл [69] относит этнографические исследования, использовавшиеся при количественных межкультурных сравнениях, по качеству к двум или большему числу уровней. Если, к примеру, этнограф прожил в данном ареале свыше двух лет и изучил при этом местный язык, этноописание попадает в разряд высококачественных. Такие переменные, как вера в колдовство

212

и нетипичные для европейцев положения при рождении детей, считаются скоррелированными с качеством данных: их в большей мере способны подмечать те, кто лучше знаком с местной культурой. В итоге любые две такие переменные обнаруживают между собой ложную корреляцию. Введя показатель вариативности данных по качеству, Нейролл получил возможность исключать или подтверждать наличие таких ложных корреляций.

Занимаясь совершенно иными сравнительно-психологическими лабораторными исследованиями, Биттермен [10] поставил более общий методологический вопрос, по отношению к которому эти качественные вариации являются одной из возможных его иллюстраций: «Мне, конечно, неизвестно, как организовать для рыб такую совокупность условий, которая предъявляла бы точно такие же требования к их сенсорике и моторике, какие предъявляются крысам в некоторой заданной экспериментальной ситуации. Не знаю я и того, как уравнивать у этих двух видов уровень побуждения или силу подкрепления. К счастью, однако, возможность значимых сравнений тем не менее остается, так как *контроль путем уравнивания* мы можем заменить тем, что я называю *контролем путем систематического варьирования*. Рассмотрим, к примеру, гипотезу о том, что различия кривых, которые вы здесь видите, объясняются не различиями в научении, а в уровне голода. Эта гипотеза предполагает, что существует некий уровень голода, при котором рыбы будут действовать все более успешно. Такую гипотезу нетрудно проверить. Нам остается теперь только варьировать в широком диапазоне уровень голода для различных групп рыб, а мы хорошо знаем, как это делается. Если, несмотря на максимально возможный диапазон варьирования голода, не обнаруживается прогрессивного улучшения в поведении рыб, мы можем отвергнуть гипотезу, связанную с голодом. Путем систематического варьирования могут быть проверены также гипотезы относительно других переменных. В связи с вопросом о научении в измененных условиях я могу здесь лишь сказать, что у крыс прогрессивные улучшения наступают в самых разнообразных экспериментальных условиях. В самом деле, трудно найти такую совокупность условий, в которых в поведении крыс улучшений не наблюдается.

Для рыб, напротив, не получено надежных свидетельств улучшения при варьировании условий в достаточно широком диапазоне» (Bitterman [10, с. 396—410].

В этом плане можно выделить два типа контроля артефактов. Это, во-первых, *интерполирующее*, или *протекающее в определенных рамках, варьирование*, при котором дополнительное варьирование охватывает весь диапазон вероятных вариаций или выходит за его пределы. Таковы по своей природе два приведенных примера из области изучения оптических иллюзий. Взятая в качестве средства контроля, такая мера предполагает, что законы по своему характеру линейны или монотонны, то есть что промежуточные величины дадут промежуточный эффект. Обычно это предположение дает вполне разумные основания для того, чтобы счесть опасность неправдоподобной, коль скоро об этом свидетельствуют крайние величины.

Во-вторых, существует *экстраполирующее варьирование*, при котором нам недоступны все значения данного параметра, и, чтобы осуществить контроль, надо произвести экстраполяцию за пределы исследованного диапазона.

Такова, должно быть, проблема испытуемых (респондентов)-добровольцев. Экстраполяция на популяцию, тех, кто не вызвался быть добровольным участником обследования, — вот что нам хотелось бы осуществить, но даже в самом лучшем случае какой-то процент добровольцев необходим. Известная степень контроля будет достигнута в том случае, если мы дополнительно используем в исследовании какую-нибудь ситуацию, более экстремальную по степени добровольности участия, чем та, с которой мы обычно

имеем дело. Если в обеих ситуациях проявляется действие одних и тех же законов, мы, основываясь на предположении о монотонности закона, экстраполируем этот результат на ситуацию, где вообще отсутствует добровольность. Предположения, которых мы придерживаемся, представляются здесь интуитивно менее правдоподобными, чем в случае интерполирующего (ограничительного) варьирования, но степень их правдоподобия тем не менее достаточна для того, чтобы считать такой контроль оправданным. Мы и в этом случае приобщаем к делу

214

известную сумму данных, качество которых намеренно снижено.

Использование дополнительного варьирования в качестве средства контроля получает широкое распространение в общепринятой практике классификации респондентов на основе интервью, проведенных после обследования и рассчитанных на выявление степени осознания ими целей экспериментатора, для проверки действия тех же законов в выделенных при этом подгруппах. Мак-Гайр [см. 83] сообщает о своих исследованиях, в которых он пошел еще дальше, намеренно вводя на экспериментальной основе более экстремальные степени осознания.

### **Б. Повторение с использованием разных методов**

В исследованиях артефактов всегда повторяется одна и та же история: весьма обнадеживающий первоначальный эксперимент, построенный по схеме «использование одного воздействия и одного замера», подвергается затем критике с точки зрения конкретных правдоподобных конкурентных гипотез и сопровождается рядом других экспериментов, предполагающих использование либо контрольных групп, подвергаемых дополнительному воздействию, либо измененного метода воздействия, либо измененного метода измерения, пока, наконец, первоначальные данные не получат двойного подтверждения или не будут отвергнуты в пользу какой-нибудь конкурентной интерпретации. Даже там, где поле исследования доступно, а само исследование обходится недорого, этот цикл занимает добрый десяток лет, в чем можно убедиться на примере работы Розенберга [83] с экспериментальным исследованием диссонанса. Поэтому можно лишь приветствовать любую стратегию, которая могла бы сократить эту расточительную процедуру.

Познакомившись с описанием средств контроля, представленным в предыдущем разделе о типах артефактов, нетрудно подметить их обобщенный характер и применимость к разнообразным экспериментальным процедурам. Многочисленные методы измерения имеют аналогичное значение. Это наводит на мысль взять за правило программирование гетерометодического повторения

215

уже на начальной фазе исследования. Так, от всякого диссертационного исследования требовалось бы вводить экспериментальную переменную двумя методически независимыми способами и в каждом случае измерять эффект воздействия двумя независимыми методами. Если какой-нибудь гипотетический закон был первоначально подтвержден в каждом из четырех произведенных таким образом гетерометодических повторений, большая часть потенциально правдоподобных конкурентных гипотез будет сразу же исключена (еще до того, как они будут сформулированы в явном виде). Если все четыре повторения не дали согласованных между собой результатов, а наблюдался ряд значительных эффектов, то соискатель столкнется с интригующей эмпирической загадкой, над которой стоит поразмыслить, но у него не будет искушения высказывать

сверхкатегорические теоретические утверждения, что могло бы иметь место, если бы его взору открылась какая-то одна часть этой загадки. Однако выполнять это методическое предписание не рекомендуется. Если я правильно оцениваю наши сегодняшние теоретические успехи и экспериментальные умения, полного подтверждения все равно почти никогда нельзя получить. Описанный процесс в целом был бы более обескураживающим, чем принятая ныне практика, притом настолько, что у многих вовсе пропала бы охота проводить исследования. При нынешних стандартах, не поощряющих публикацию негативных результатов, не подтверждающих новые гипотезы, редакторы журналов почти наверняка отказались бы опубликовать честный отчет, и были бы, вероятно, правы, если смотреть на дело с точки зрения оптимальной коллективной переработки информации (Campbell [20, с. 168—170]). Науке как социальной системе требуется мотивация, достаточная для того, чтобы продуцировать научные исследования в избыточном количестве. Эта мотивация гораздо выше, когда каждый исследователь считает, что он пользуется оптимальным методом, а когда его ожидания подтвердились, верит, что доказал истинную теорию. Какая-то степень сверхоптимизма, возможно, необходима как при предвосхищении еще не сделанного, так и при ретроспективном взгляде на выполненное исследование. Так же обстоит дело и с проблемой артефактов вообще: с точки зрения мотива-

216

ции самое лучшее, по-видимому, не думать о них заранее как о неотвратимом закономерном аспекте любого исследования, а, напротив, закрыть глаза на их возможность и рассматривать всякого рода возмущение, коль скоро оно возникло, как специфическую аномалию в целеустремленном во всех других отношениях научном поиске.

Если избранное нами поприще и в самом деле чрезвычайно трудно, то систематическое появление на свет поражающих воображение результатов, в которых позднее станут подозревать артефакты, в мотивационном плане даже полезно. Они служат подкреплениями, которые побуждают потенциальных первооткрывателей и знающих свое дело критиков (чем поразительнее результат, тем выше репутация беспочвенных заявлений). Это подкрепления и мотивация, побуждающие к экспериментальной работе и эмпирическим исследованиям. И то и другое было бы утрачено, если обратиться к процедуре, успешно страхующей от чрезмерно оптимистических псевдоподтверждений будоражащих умы теорий.

### **В. Замаскированные эксперименты в естественных условиях**

Формальная схема, намеченная в предыдущем разделе, дает полезное общее представление о возможных артефактах и в то же время выделяет один из основных классов правдоподобных конкурентных гипотез. Мы можем обозначить его как *осведомленность об экспериментировании*, или, как я когда-то назвал его, *реактивные настройки*. В 1957 г. я писал: «При использовании любого экспериментального плана респонденты могут установить, что они участвуют в эксперименте, и такая осведомленность может привести к эффекту взаимодействия, вызывая реакции на  $X$  (экспериментальное воздействие), которых не было бы, если бы  $X$  было встречено без установки «Я подопытный кролик». Лазарсфельд [63], Керр [58], Розенталь и Франк [81] выступили с ценными обсуждениями этой проблемы. Такого рода эффекты ограничивают возможности обобщения. Полученные результаты могут быть отнесены только к тем респондентам, у кого есть

217

такая осведомленность, и не могут быть распространены на популяцию, у которой в отношении  $X$  нет экспериментальных установок.

Этот эффект может давать сдвиг в сторону негативизма вроде нежелания внять какому бы то ни было убеждению или пойти на какое бы то ни было изменение. Это можно было бы сравнить с отсутствием, какого бы то ни было непосредственного эффекта в случае, когда коммуникатор не пользуется доверием (Hovland et al. [55]). Чаще же, вероятно, наблюдается своего рода результат взаимодействия, когда респондент принимает ожидания экспериментатора и снабжает его псевдоподтверждениями. Отчасти имеет место положительная реакция, вероятная в ситуациях, где респондентами являются добровольцы, заинтересованные возможностями, связанными с  $X$ . Так, хоуторнские исследования (Roethlisberger, Diskson [77]) иллюстрируют изменения, вызванные осведомленностью об экспериментировании, а не специфической природой  $X$ .

Проблема реактивных настроек касается всех особенностей эксперимента, которые могут привлечь внимание респондента к факту экспериментирования и его целям. Очевидное предварительное тестирование представляет собой особенно уязвимое звено, поскольку оно сигнализирует о предмете и целях экспериментатора. В случае сообщений, явно направленных на убеждение респондента, о содержании намерений экспериментатора сигнализирует само воздействие  $X$ , если только это сообщение не выглядит как часть естественной обстановки. Даже группы, подвергающиеся одному только итоговому тестированию, не составляют исключения: факт итогового тестирования может вызвать тот или иной реактивный эффект. Респондент может сказать себе: «Ага, теперь понятно, почему нам показывали этот фильм». Это соображение оправдывает практику маскировки связи между  $O$  (наблюдение или измерение) и  $X$ ... как, например, путем смены персонала, участвующего в эксперименте, использования различных способов предъявления, включения фрагментов содержания  $O$ , которые имеют отношение к  $X$ , в некое маскирующее многообразие других тем» (Campbell [19, с. 308—309]).

Многие, хотя и не все, артефакты, отмеченные в работах, посвященных артефактам, можно охватить гипотезой, согласно которой результаты таковы, каковы они есть, только потому, что испытуемые знали, что с ними экспериментируют, а также потому, что степень осведомленности участников экспериментальной группы может быть различной. Таковы по своей природе характеристики, отмеченные Орном, в то время как на эффекты плацебо, рассматриваемые в его обзоре наряду с первыми, это не распространяется, поскольку они, несомненно, присущи также неэкспериментальному медицинскому применению лекарств. Сенсibiliзирующее влияние предварительного тестирования, будь оно эмпирически установлено, тоже вошло бы в этот класс, и, возможно, предписывающее влияние предварительного тестирования также проистекает из сознания, что данные предварительного и итогового тестирований подлежат экспериментальному сравнению. Добровольное участие в эксперименте предполагает знание об эксперименте, в котором испытуемые намерены участвовать. Влияние со стороны экспериментатора в целом имеет иную природу (в самом деле, для эффектов Пигмалиона нет надобности в том, чтобы учитель или ученик сознавали, что они участвуют в эксперименте), но эффекты, связанные с влиянием экспериментатора и обязанные своим происхождением сотрудничеству с ним респондента в достижении целей эксперимента, как он их себе представляет, относятся сюда же. Подозрительность, вызываемая намерением экспериментатора, — это чуть ли не синоним используемой здесь главной категории, и меры, вводимые Мак-Гайром, равно как и его экспериментальные процедуры, могут служить конкретной иллюстрацией этого. Боязнь оценки полностью зависит от осведомленности об экспериментировании, судя по крайней мере по тем примерам, которые приводит Розенберг. Эффекты взаимодействия

при измерении, которые отмечались при косвенном измерении социальных установок (Campbell [17], Kidder, Campbell [59]), и ненавязчивые измерения (Webb et al. [94]) имеют ту же природу. Вопрос об «экспериментальном реализме» и «примитивном реализме» (Aronson, Carlsmith [5]) во многом связан с проблемой осведомленности. Многие из существующих исследований свидетель-

219

ствуют об этом, хотя другие — нет. Симуляция (см., например, Broun [14], Kelman [57], Orne [83]) может служить здесь полезной дополнительной альтернативой, но она не может быть заменой, так как осведомленность об эксперименте доводится при ней до крайности.

Очевидным средством избежать все эти артефакты является замаскированный эксперимент, в котором респонденты (а то и сами экспериментаторы) не знают, что они принимают участие в эксперименте, не знают, что «с ними экспериментируют». Такие эксперименты лучше всего проводить в естественных, а не в лабораторных условиях не потому, что естественные условия в большей мере репрезентируют объект дальнейших обобщений, а, скорее, потому, что в естественных условиях респонденты не подозревают, что с ними экспериментируют. Лаборатория, как правило, воспринимается именно как совокупность условий, созданных специально для экспериментов.

Этот аргумент может стать еще более убедительным, если мы обратим внимание на то, что большинство лабораторных исследований, характеризующихся большим «экспериментальным реализмом», добиваются его, отвлекая респондента каким-нибудь правдоподобным прикрытием или легендой, в то время как воздействие вводится как неожиданное или случайное событие. Так, Френч [45] собирал группы для дискуссии, а затем, применяя в качестве экспериментального воздействия дым, впускал его в комнату сквозь дверные щели. Орн пользовался «случайной неудачей», а Дарли и Латане [37] — эпилептическим припадком. В некоторых случаях реальный эксперимент проводится среди респондентов, ожидающих своей очереди участвовать в эксперименте. В бесчисленном множестве экспериментов с использованием подставных испытуемых воздействие состоит в выполнении задания кем-нибудь из партнеров. Использовался и такой случайный факт, что один экспериментатор был по происхождению негр, а другой — кавказец (Rankin, Campbell [76]). Респондента часто заставляют поверить, что он является экспериментатором (см., например, Festinger, Carlsmith [42], Milgram [66] и т. д.). Все это попытки использовать естественные факторы ситуации, чтобы избежать эффек-

220

тов осведомленности об экспериментировании. Становясь достоянием гласности, эти меры утрачивают свою полезность, однако они могут обрести ее вновь, если будут совсем вынесены из стен лаборатории.

К настоящему времени социальная психология накопила уже достаточный опыт замаскированных экспериментов в естественных условиях, чтобы обеспечить базу для их зрелой таксономии и методологии. Работы Уэбба и его соавторов [94] могут служить весьма полезным началом, хотя внимание этих авторов сосредоточено скорее на измерении, чем на экспериментальных воздействиях. Аронсон и Карлсмит [5] разрабатывают другой аспект проблемы. Следует упомянуть работу Гросса, Коллинса и Байрен [50], а также Розенבלата и Миллера [80]. Однако задача все еще не решена, не будет решена она и в настоящей работе. Тем не менее не повредит делу несколько соображений и иллюстраций. Прежде чем дать иллюстрации, мы остановимся на двух вопросах, чтобы сформулировать точки зрения, нужные для дальнейших оценок.

1. *Ограничения, обусловленные содержанием.* Все, что нужно для вполне замаскированного эксперимента в естественных условиях, — это какая-нибудь естествен-

ная форма контакта с людьми (или с социальными единицами, достаточно малыми и достаточно многочисленными, чтобы, подвергнув их воздействию в случайном порядке, достичь эффективного уравнивания) при требуемом личном характере этого контакта, не вызывающем подозрения, что другие единицы подвергаются воздействиям иного рода, и при возможности для респондентов в тех же условиях давать на воздействие какую-нибудь естественную реакцию, которой можно было бы воспользоваться как релевантным средством измерения эффекта. Коль скоро такие условия должны сохранять естественность, они не могут создаваться произвольно для всех возможных воздействий и с использованием всех возможных мер. Наоборот, их следует своевременно найти. Каждая данная совокупность условий неизбежно будет накладывать серьезные ограничения на те классы проблем, которые могут в ней изучаться. Эти ограничения будут касаться разновидностей экспериментальных переменных, которые могут быть включены в исследование, и доступных форм измерения.

221

2. *Обман, демаскировка (debriefing) и другие этические проблемы.* Очевидно, что замаскированные эксперименты в какой-то мере предполагают обман, и, как ясно показали Мак-Гайр [83] и Келмен [57], это нежелательный момент, оправданный разве что соображениями более высокого порядка. Одно из этих соображений — моральная ценность созидания нетривиальной социальной науки. При всяком таком сравнительном взвешивании соперничающих ценностей становятся важными количественные характеристики каждой из них, например масштаб обмана. Если рассматривать страдания того, кто обманывает (экспериментатора), то неявная «белая ложь» не столь болезненна, как злонамеренная «черная» (например, «активный обман» Мак-Гайра), и, хотя и та и другая могут одинаково травмировать реципиента, он также будет склонен признать первую менее аморальной в силу наших языковых привычек. Ложь того и другого типа менее болезненна и не столь аморальна, когда ее ждут и когда она конвенционально оправданна. В связи с неточностью языка и нашей общей способностью опираться на словесные отчеты других людей (Asch [6], Campbell [22]) этот эффект тем выше, чем заметнее выявляется ложь в ситуациях, связанных с высоким престижем. Во всех этих случаях уровень адаптации, обязанный своим происхождением другим сферам общественной практики, служит своего рода релятивистским основанием для сравнения. Какая-нибудь сладкая ложь не столь аморальна в сообществе, где подобные случаи часты, нежели там, где она прозвучала бы впервые.

В этом плане замаскированные естественные эксперименты в значительной мере отличаются друг от друга и в принципе не ставят, вероятно, перед исследователем более сложных проблем, чем лабораторные эксперименты. Они, по-видимому, больше опираются на невербальную, или «белую», ложь и меньше — на прямой обман. Они обычно находят себе применение в таких ситуациях словесного общения, которые уже разъедены обманом в большей мере, чем ситуация обучения. Если ложь обнаружена, то смоделированное влияние будет, вероятно, ниже, чем это имеет место в отношениях между профессором и студентом. Однако в естественных условиях обычно отсутствует молча-

222

ливое соглашение по поводу приемлемой лжи, которое, возможно, достижимо в психологической лаборатории<sup>1</sup>.

Особняком стоит проблема получения от респондента согласия на участие в эксперименте, — проблема, которая приобрела большую практическую важность теперь, когда стала актуальной для части наших мероприятий по поддержке исследований. В случае тех экспериментов, которые мы собираемся здесь описать, такое согласие — вещь,

очевидно, невозможная, коль скоро речь идет о том, чтобы добиться их маскировки и неосведомленности испытуемых. С другой стороны, в условиях, где используются каналы коммуникации, по средствам и размаху представляющие общественный интерес, и эти каналы без всякого разрешения на то свободно используются лицами, не являющимися экспериментаторами, это требование становится просто неразумным.

Другая этическая проблема — это вторжение в личную жизнь. Оно не обязательно в замаскированных естественных экспериментах, а иной раз и невозможно. Одну из сторон этой проблемы составляет анонимность регистрационных записей. Однако, когда сбор сведений доверительного и .сугубо личного характера осуществляется таким способом, который дает возможность связать их с именем данного человека, маскировка и отсутствие согласия в любом случае усугубляют опасность вторже-

---

<sup>1</sup> Действительно, на практике, чтобы избежать этических проблем в университетской аудитории, можно было бы, например, в начале семестра собрать всех испытуемых и обратиться к ним с таким заявлением: «Примерно в половине всех экспериментов, в которых вы будете принимать участие в течение этого семестра, потребуется во имя валидности эксперимента полностью или частично держать вас в неведении относительно его истинных целей. Не будет у нас и возможности информировать вас о том, что это были за эксперименты и в чем заключалась их действительная цель, до тех пор, пока не будут собраны все экспериментальные данные. Мы гарантируем вам, что какая бы то ни было опасность или вторжение в личную жизнь исключены и ваши ответы будут анонимны и будут содержаться в полной тайне. А теперь мы просим вас скрепить своей подписью соглашение, составленное по надлежащей форме, которое будет свидетельствовать о вашем согласии участвовать в экспериментах на этих условиях». Тем самым мы попросту сделали бы явным то, что понимается сегодня всеми, не усугубляя, по-видимому, существующую сейчас проблему осведомленности и подозрительности.

ния в личную жизнь. В случае естественного эксперимента проблема разного рода травм, включая унижения и оскорбления личности, стоит не более остро, чем в случае лабораторного эксперимента.

Демаскирование, разъяснение респонденту истинной природы эксперимента, извинения за обман и, если это возможно, обеспечение обратной связи в отношении результатов суть процедуры, типичные для открытой лаборатории в университетском городке, но полностью исключенные из замаскированного полевого эксперимента. Хотя такое демаскирование и стало обычным компонентом вводящих в заблуждение лабораторных экспериментов, в этическом плане ему присущи многочисленные изъяны. Оно приносит во много раз большее облегчение экспериментатору, освобождая его от болезненных переживаний, связанных с обманом, чем респонденту, который может узнать из признаний экспериментатора о своей собственной доверчивости, конформности, жестокости или предубежденности. Обеспечивая модель обмана и разглашая его, оно тем самым лишает основательности словесное общение и для респондента, и для экспериментатора. Оно подрывает доверие к лаборатории и снижает пользу, которую мог бы иметь обман в будущих экспериментах<sup>1</sup>. Аргайл [4], Милтон Рокич (личные сообщения), Столлак [91], Мак-Гайр [83] и Аронсон и Карлсмит [5] обратили внимание на эти изъяны, и они достаточно значительны, чтобы считать оправданным отказ от демаскировки в тех случаях, когда экспериментальное воздействие не выходит за рамки обыденного опыта респондента, представляя собой всего-навсего экспериментальное использование коммуникаций на обыденном уровне. Этот нормальный диапазон, безусловно, превзойден в исследованиях Аша [7], в которых предъявленный материал вступает в резкое противоречие с обыденным опытом испытуемых — простым перцептивным суждением восьми свартморских студентов, или в исследованиях Милгрэма [66], в которых респондент должен был воздействовать сильным разрядом электрического тока на своего товарища по колледжу. Он превзойден, по-видимому, в убеждающих

<sup>1</sup> В свете этого представляется тем более примечательным безмятежное повествование о вводящих в заблуждение экспериментах во вступительных курсах психологии.

224

сообщениях, содержащих вымышленные данные ни по поводу весьма серьезных проблем, но не превзойден, вероятно, в большинстве исследований убеждающей коммуникации. В экспериментальной социальной психологии мы исчерпали возможности наших лабораторий. По этой причине мы уже покидаем стены колледжа, отдавая предпочтение средней школе и улице. Гласность погубит в конце концов и эти лаборатории, но это произойдет намного быстрее, и гнев общественности, вызванный обманом, не уменьшится, если мы в ходе замаскированных естественных экспериментов будем заниматься саморазоблачениями.

3. *Примеры классических исследований.* Госнелл [48] рассылал зарегистрированным избирателям убеждающие сообщения, имевшие целью побудить их к голосованию, и по протоколам избирательных участков определял в дальнейшем, проголосовал ли член той или иной экспериментальной или контрольной группы; он провел, таким образом, никем не замеченный эксперимент, пользуясь рядом сообщений, вполне укладывающихся в пределы нормы. Хотя сегодня не очень-то доверяют протоколам, составленным в избирательных участках Чикаго, к нашим услугам избирательные участки других городов. Это лаборатория, которая к настоящему времени могла бы быть использована сотни раз, однако, насколько мне известно, ею не воспользовались с тех пор ни разу. Хотя тематика здесь крайне узкая, убеждающие сообщения можно было бы варьировать в широком диапазоне экспериментальных параметров, к которым прибегают при исследовании убеждающей коммуникации в лаборатории. Ценность этой лаборатории значительно возросла бы, если бы наряду с информацией о том, проголосовал ли данный избиратель, можно было бы получить сведения о том, как он проголосовал. Хотя применительно к отдельным индивидам эта информация не подлежит огласке, в отношении избирательных участков в целом она является гласной, благодаря чему эти последние и использовались Гартманном [52] в качестве единицы выборки в его классическом исследовании влияния, оказываемого политическими листовками с рациональными или эмоциональными доводами. Пользуясь этим приемом в таком штате, как Калифорния, где избиратели участвуют в

225

голосовании не только по поводу тех или иных кандидатур, но и тех или иных вопросов, можно проверить обширный класс теорий убеждающей коммуникации. Однако и лаборатория Гартманна больше не использовалась. В подобных исследованиях согласие и демаскирование были бы совершенно неуместными, разве что в сообщениях содержались бы клевета и прямой обман; но в этом случае демаскирование после выборов наверняка вызвало бы бурю вполне законного негодования. Таким образом, диапазон экспериментальных воздействий, конечно, ограничен, но он мог бы тем не менее охватывать и односторонние, и двухсторонние сообщения, экстремальность защищаемых в них позиций или их насыщенность льстивыми эпитетами. И Госнелл, и Гартманн защищали ту позицию, в которую они действительно верили. (Гартманн сам баллотировался в мэры от социалистической партии.) Эта искренность и связанная с ней правдивость были бы утрачены в исследованиях первичности и новизны, в которых защищаются обе противоположные альтернативные позиции, разве что в эксперименте сотрудничали бы экспериментаторы с противоположными убеждениями или экспериментатор получил бы от противоборствующих сторон одобрение на распространение нужных ему сообщений. Это — смещение в сторону «белой лжи». Но, с другой стороны, в эксперименте необходимо манипулировать лишь тем, когда и кому посылать сообщения, чтобы они соответствовали частичному и случайному распределению. (В подобных исследованиях мы часто склонны предавать непропорционально широкой гласности второстепенные вопросы просто

потому, что другие сообщения на ту же тему почти отсутствуют.) Пользуясь этой лабораторией при изучении проблемы конформности (Campbell [18]), следовало бы воздержаться от использования в функции обратных связей вымышленных результатов опроса общественного мнения, столь охотно применяемых в лаборатории колледжа, и ограничиться сравнением случаев наличия и отсутствия обратной связи, а также соответствующих источников (избирательный участок, штат или страна), от которых эта обратная связь поступила. В некоторых случаях это ограничение означало бы весьма реальную жертву в отношении ясности и четкости экспериментального вывода, однако предъявление фаль-

226

сифицированных результатов опроса было бы недопустимым манипулированием голосованием, а это уже совсем иное дело, чем, если бы речь шла не о столь важной акции, как выборы.

Замаскированные полевые эксперименты по большей части не имели столь обширной лаборатории и служили от случая к случаю весьма конкретным целям. Так, в одном из исследований конформности Лефкович, Блейк и Мутон [64] моделировали поведение пешеходов при переходе перекрестка на красный свет. «Пешеходы» были в одежде, которая символизировала либо высокий, либо низкий их социальный статус. Отмечалось влияние этого фактора на тенденцию наблюдателя пренебрегать сигналами. Шварц и Сколник [85] варьировали содержание посланий, адресуемых нанимателям с просьбой о предоставлении временной подсобной работы на летнем курорте, и изучали при этом, как влияет на найм сообщение о данных уголовного порядка. Шварц и Орлеане [86] пользовались налоговыми декларациями, чтобы измерить страх, возникавший перед юридическими санкциями. Брайан и Тест [16] предоставляли испытуемому возможность совершить альтруистический поступок, оказав помощь женщине, у которой села крышка; в одних случаях этому предшествовало предъявление модели поведения по оказанию помощи, в других — нет. Пейдж [70] в случайном порядке делал на ученических работах замечания, содержащие мотивационные суждения, и с помощью последующих проверок измерял впечатление, которое они производили на учащихся. Дуб и Гросс [38] сравнивали реакцию водителей высокого и низкого статуса на стоящий впереди автомобиль, который (в чем и состояло экспериментальное воздействие) не мог сдвинуться с места, когда зажигался зеленый свет.

Некоторые из этих лабораторий настолько узки, что трудно представить себе еще какую-нибудь проблему, для решения которой они могли бы подойти; однако некоторые из них могут найти и более широкое применение. Так, ситуацией, созданной Шварцем и Сколником, можно было бы воспользоваться для исследования широкого круга проблем, связанных с презентацией личности и предполагающих обращение к особым, количественным измерениям эффекта. Даже в такой на

227

первый взгляд узкой методике, как методика утерянных писем, предназначенной для изучения честности (Merrit, Fowler [65]), можно добавить много других переменных. Написывая на конвертах адреса лиц, принадлежащих к группам, релевантным определенным социальным установкам, Милгрэм (Milgram, Mann, Harter [67]) получил весьма валидные на вид поведенческие количественные меры социальных установок. Гросс [49], оставляя конверты незапечатанными и используя различные варианты в содержании писем, имел возможность оперировать множеством переменных.

4. *Использование рабочей силы.* Важное значение в экспериментировании имеет специальный контроль за тем или иным отрезком личного времени индивида. Именно эта

мера разрывает причинные связи, существующие между прошлыми условиями и экспериментальными воздействиями, и позволяет случайным образом подвергать эквивалентные выборки различным воздействиям. Чем больше этот контроль, тем выше полезность эксперимента во многих отношениях. Одной из таких совокупностей условий является ситуация использования рабочей силы. Я не буду говорить здесь о разработке этой области в прикладных экспериментальных исследованиях, сосредоточенных на проблемах нанимателя и праве выбора, имеющемся у администрации (см., например, Feldman [40], Kerr [58]), и остановлюсь вместо этого на использовании производственной ситуации для исследований в области теоретической социальной психологии. Особенно ярким примером здесь могут служить исследования Адамса [2], посвященные несправедливой оплате за выполненную работу. В них использовались работники, нанятые на короткий срок и занятые в течение неполного рабочего дня, при этом им выплачивалось денежное вознаграждение того же порядка, как испытуемым в обычной лаборатории. Стюарт Кук [30] воспользовался этой ситуацией в своем (до сих пор не опубликованном) классическом исследовании влияния контакта лиц с равным статусом на расовые установки.

Сходной ситуацией (бюро по найму) воспользовались Рокич и Мецей [78], чтобы воспроизвести в своем прекрасном исследовании результат, уже наблюдавшийся до этого в более искусственных лабораторных

228

условиях. (При ближайшем рассмотрении, однако, их результаты указывают на то, что люди впадают в другую крайность, тяготея при межрасовых контактах к честной игре. Эта тенденция симптоматична, возможно, для экспериментальных ситуаций, вызывающих реактивную настройку.)

В приведенных в качестве иллюстраций исследованиях испытуемые не подвергались никаким чрезвычайным или опасным воздействиям: просто заранее планировался жизненный опыт, который все равно так или иначе был бы — или мог бы быть — приобретен некоторыми из них. В этих исследованиях демаскирование представляется не только не необходимым, но и, пожалуй, неразумным, а этические проблемы, связанные с обманом, сводятся к минимуму. Но весь вопрос упирается, конечно, в природу самих воздействий. Достаточно вспомнить о таком противоположном примере, как использование ситуации воинской службы для создания реальной угрозы неминуемой гибели (Berkun, Bialek, Kern, Jagi [9]; Dailu Palo Alto Times[36]; Argule [3]), красноречиво свидетельствующем о безнравственном избытке ученого рвения.

5. *Контакты в общественных местах.* Целый ряд экспериментальных воздействий может быть использован при изучении случайных контактов незнакомых людей. Брайан и Тест [16] ставили на тротуар кружку Армии спасения, а в качестве представителей, собирающих пожертвования, выступали лица различных этнических групп, учитывалось также наличие или отсутствие модели поведения. Фелдман [41] просил местных жителей и приезжих указать правильную дорогу, помочь отправить письмо, а также спрашивал, не ронял ли респондент только что найденную долларовую бумажку. Кук, Бин, Кьяльдини, Кровец и Рей [31] провели интересное исследование, в котором какой-то непривычно развязный, но не опасный на вид незнакомец отпускал комплименты идущим по университетскому городку женщинам; эффект измерялся с помощью нищего, стоящего с протянутой рукой на обочине (а также интервьюеров, которые говорили, что они проводят обследование, и задавали женщинам вопросы относительно их реакций на разного рода комплименты). Милгрэм с сотрудниками собирали для приманки экспериментальные

229

толпы разного размера и регистрировали число прохожих, которые были ими привлечены. В исследованиях межличностного пространства по методике Соммера [90] использовались ситуации, в которых экспериментатор задает вопросы незнакомым людям, подсаживается к ним в транспорте, в кафетерии и т. д. В этом случае возникает множество возможностей, в том числе и неприемлемые в этическом плане. Ходят слухи об эпилептических припадках, разыгрываемых на улице, о водителях такси — экспериментаторах, устраивающих в пути необъяснимые задержки, чтобы фрустрировать тревожных пассажиров, и т. п.

6. *Формирование выборки путем обращения к испытуемым.* В исследованиях процесса убеждения чрезвычайно гибкой замаскированной лабораторией служат все те ситуации, в которых к незнакомым лицам обращаются с предложениями, узаконенными обычаем. Естественная реакция на обращение становится, по существу, мерой ответа на воздействие. Предложение о покупке, сбор средств, сбор подписей под воззваниями — вот примеры узаконенных целей, а непосредственная почтовая и телефонная связь или посещение на дому — примеры узаконенных средств. Институты общественного мнения обеспечивают весьма мобильную технологию и персонал для выборочного обследования (и разве существует перевоплощение более артистичное, чем сотрудник такого института, переодетый коммивояжером). Пользуясь в качестве вариантов убеждающего обращения посещениями на дому или почтовой перепиской, мы получаем прекрасную возможность создавать эквивалентность случайным образом, не давая понять респондентам, что с ними проводится эксперимент. (Чтобы уберечь респондентов от подобных подозрений, часто желательно пространственно обособить сравниваемые группы, к которым обращаются с разными предложениями, и тем самым лишить их возможности обмениваться впечатлениями.) В порядке комментария к этике «белой лжи» отметим, что и у экспериментатора, и у визитеров обман вызывал бы значительно меньший дискомфорт, если бы они были действительно заинтересованы в сборе средств или в сбыте продукта, как это часто бывает в соответствующих случаях.

В случае сбора средств возможность соответствующую-

щей квантификации обеспечивается суммой пожертвования, и даже по комментариям, сопровождающим отказ от пожертвования, можно судить о степени благосклонного отношения респондента к мероприятию. В случае сбыта товаров дихотомическое деление на «проданное» и «непроданное» можно обогатить серией градаций по шкале Гутмана, предоставляя почтовые карточки, которыми респондент может воспользоваться, приняв запоздалое решение о покупке, а также буклеты с дополнительной информацией и т. д. В случае сбора подписей под воззваниями естественная количественная мера дихотомична, но и ею можно воспользоваться для построения градуальной шкалы, да и комментарии респондентов могут быть при этом проанализированы, как при опросах общественного мнения (хотя регистрация комментариев при личном контакте была бы исключена). Иногда можно, не уменьшая правдоподобия, предлагать воззвание в двух вариантах — умеренном и резком.

Блейк и его сотрудники (Blake, Mouton, Hain [11]; Nelson, Blake, Mouton [53]), а также Гор и Роттер [47] первыми применили сбор подписей под воззваниями для внелабораторных исследований в университетском городке. Широкое применение в рекламной промышленности находят эксперименты с использованием непосредственного рекламирования по почте. Кук и Инско [32] в качестве экспериментального воздействия использовали письма различного содержания, отправленные по почте. У Брока [13] разнообразные экспериментальные воздействия вводились продавцами в магазине. Вероятно, компании, сбывающие товары при посредничестве коммивояжеров, сознательно

экспериментировали со своими методами. Однако в целом это широкое поле возможностей не было освоено в научных целях.

Хотя опросы общественного мнения, по-видимому, способствуют возникновению эффекта «подопытного кролика», они в достаточной мере стали частью общественной жизни, и это позволило целому ряду экспериментаторов, ориентированных на теорию, прибегнуть к ним в целях предъявления замаскированного воздействия в условиях полевого эксперимента (см., например, Abelson, Miller [1]; Freedman, Fraser [44]; Miller, Levy [68]). В целом они замаскированы в меньшей

231

степени, чем обращения с предложениями услуг. Введение при их помощи информации убеждающего содержания или другого воздействующего фактора требует обращения к искусственным средствам, тогда как предложение услуг само по себе открывает возможность убеждающего воздействия. В то же время оно заключает в себе благоприятную возможность обоснования количественных измерений вербальных социальных установок. Клиент (потребитель) может выполнять функцию экспериментатора с тем же успехом, что и продавец. Традиционный подход к изучению гражданских прав с помощью тестов мог бы служить подобной экспериментальной парадигмой при полевым экспериментальном исследовании влияния расовой принадлежности на предоставление жилья, ночлега (La Piere [62]), обслуживания в ресторане (Kutner, Wilkins, Yarrow [61]) и т.п. Фрэнзен [43] экспериментально варьировал поведение клиента (потребителя), шкалируя готовность аптекарей дать врачебный совет. Юнг [56] экспериментировал, сводя торговцев автомобилями с покупателями, которые отличались разным уровнем доверчивости. Фелдман [41] привлекал клиентов различной этнической принадлежности и изучал, как воспринимается нечаянная переплата лавочниками и водителями такси. Шапс исследовал готовность оказать помощь у продавцов обуви, рассматривая ее как функцию степени зависимости клиента от данного продавца (например, помощь женщине, у которой в этот момент каблук либо сломан, либо цел), издержек или уровня, на котором происходит сравнение альтернатив (дожидаются ли обслуживания другие клиенты), и иных факторов (например, приходит ли клиентка с приятелем или одна). Основной зависимой переменной является число пар обуви, которые продавец предлагает клиенту, хотя могут регистрироваться и вербальные реакции. При таком плане исследования этические проблемы заключаются в том, нужно ли оповещать об эксперименте продавца и компенсировать ему потраченное время. Возражения против этих процедур состоят в следующем. Такое воздействие по своему фрустрирующему эффекту не выходит за рамки, обычные для женщин-покупательниц, около 30% которых после каждого посещения магазина покидают его, не сделав покупки. При наличии достаточных ассигнований и ис-

232

пользовании большого числа клиентов-экспериментаторов, что, между прочим, никогда не лишне (Hammond [51], Brunswik [15]), эта процедура без убытка для общества может завершаться какой-нибудь покупкой. Рассматриваемая ситуация относится к числу тех, где социальные нормы, запрещающие обман, уже утратили свое значение не только потому, что обман практикуется продавцами, но еще и из-за того, что с помощью псевдоклиентов наниматели проверяют обходительность служащих, их эффективность и честность, создавая ловушки, которые тоже нарушают суверенитет личной жизни, связывая те или иные действия с именем продавца. В отличие от этого использование клиента, занятого в исследовании, гарантирует полную конфиденциальность и анонимность. Демаскирование, по-видимому, не привело бы к снижению фрустрации у продавца, но просто-напросто

направило бы ее на другой объект. Поскольку искушенность и цинизм составляют в каком-то смысле предмет его профессиональной гордости, сознание, что он попался, было бы ему неприятно. Польза, которую естественные эксперименты могли бы принести в будущем, уменьшилась бы уже в результате одних лишь личных бесед с продавцом, не говоря уже о значительно возросшей возможности огласки через прессу. Природа экспериментального воздействия является решающим фактором, и те воздействия, которые нуждаются в демаскировании, не должны, вероятно, осуществляться без разрешения испытуемого. В то же время существуют этические ценности релевантной социальной науки и сказывается безнадежная нехватка у нас соответствующих лабораторий.

7. *Артефакты.* Предыдущему изложению свойствен дух апологетики, но мы не можем позволить ему скрывать от нас тот факт, что замаскированным полевым экспериментам сопутствует та же эпистемологическая трудность, которая столь пессимистически описывалась в более ранних разделах настоящей работы. В них контролируется только одна группа артефактов, связанная с осознанием людьми своего участия в эксперименте.

Ясно, в частности, что влияние экспериментатора вероятно во всех описанных здесь естественных экспериментальных ситуациях и усугубляется в некоторых из них тем обстоятельством, что экспериментатор вынуж-

233

ден регистрировать вербальные реакции уже после того, как он расстался с респондентом. В каждом отдельном случае воздействующая переменная и количественная мера реакции могут оказаться концептуально сложными, и ответственность за результаты, будь то основные эффекты или модифицирующие взаимодействия, ложится на факторы экспериментальной процедуры, не относящиеся к теоретической переменной. Что касается используемых здесь количественных мер естественных реакций, то работа Уэбба и его соавторов [94], несмотря на ее оптимистический в целом тон, дала все основания для пессимизма. В конечном счете, здесь потребуются контрольные группы, подвергаемые дополнительному воздействию, или повторные исследования с использованием иных воздействий и количественных мер, как это имеет место в случае лабораторных исследований.

## V. РЕЗЮМЕ

Логика научного вывода указывает на то, что эксперименты не могут подтверждать теории, они лишь подвергают их проверке. Для любого экспериментального результата, соотносимого с теорией, существует бесконечное множество потенциальных конкурентных объяснений, часть которых требует нашего внимания, потому что они, с одной стороны, открыто защищаются, а с другой — сопоставимы с подтверждаемой теорией по степени правдоподобия. Основную группу таких правдоподобных конкурентных гипотез составляют методологические артефакты, обязанные своим происхождением посторонним техническим аспектам экспериментального воздействия или использованным средствам измерения и принимающие вид либо основных эффектов, либо эффектов взаимодействий.

Контроль никогда не может быть настолько полным, чтобы заранее исключить все правдоподобные конкурентные гипотезы. Как правило, исследователь должен искать способы контролирования каждого артефакта по ходу развития самого исследования, пользуясь средствами, которые специфичны для каждого отдельного сочетания гипотезы по поводу артефакта и теоретической переменной. Однако для повторяющихся классов

234

артефактов найдены такие способы контроля, которые служат общим целям, и они приобретают статус эмпирически разработанных методологических требований данной области. В число общих стратегий контроля входит применение дополнительных воздействий в контрольной группе, варьирование технических нерелевантных характеристик воздействующей переменной, изменение метода измерения и внесение дополнительных вариаций в качество данных.

Так как в лабораторной социальной психологии большинство фундаментальных гипотез по поводу артефактов становится возможным благодаря осведомленности респондента о том, что он принимает участие в эксперименте, внимание сосредоточивается на методах и этике замаскированных экспериментов в естественных, нелабораторных условиях. Таким экспериментам удается обойти проблему артефактов не в целом, а в связи с одним только данным их типом.

## Литература

1. A b e l s o n R. P., M i l l e r J. C. Negative persuasion via personal insult. *J. of Exp. and Soc. Psychol.*, 1967, 3, 321—333.
2. A d a m s J. S. Toward an understanding of inequity. *J. of Abn. and Soc. Psychol.*, 1963, 67, 422—436.
3. A r g y l e M. Report to the Council of the British Psychological Society on my dealings with the APA Committee on Scientific and Professional Ethics and Conduct. June 24, 1960 (Mimeo).
4. A r g y l e M. Experimental studies of small social groups. — In: A. T. Welford, M. Argyle, O. V. Glass, J. N. Morris, (eds.), *Society: Problems and methods of study*. London, Routledge and Kegan Paul, 1962, 77—89.
5. A r o n s o n E., C a r l s m i t h J. M. Experimentation in social psychology. *Handbook of Social Psychology*, Vol. 2, Reading, Massachusetts, 1968, 1—79.
6. A s c h S. E. *Social Psychology*, Englewood Cliffs, Prentice Hall, 1952.
7. A s c h S. E. Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychol. Mon.*, 1956, 70, № 9 (whole № 416).
8. B o r i n g E. G. The nature and history of experimental control. *Amer. J. of Psychol.*, 1954, 67, 573—589.
9. B e r k u n M., B i a l e k H. M., K e r n R. P., Y a g i K. Experimental studies of psychological stress in man. *Psychol. Mon.* 1962, 76 (15, whole № 534), 39.
10. B i t t e r m a n M. E. Phyletic differences in learning. *Amer. Psychologist*, 1965, 20, 396—410.

235

11. B l a k e R. R., M o u t o n J. S., H a i n J. D. Social forces in petition signing. *Southwest Social Science Quarterly*, 1956, 36, 385—390.
12. B l o c k J. *The challenge of response sets*. N. Y., Appleton-Century-Crofts, 1965.
13. B r o c k T. C. Communicator-recipient similarity and decision change. *J. of Pers. and Soc. Psychol.*, 1965, 1, 650—654.
14. B r o w n R. Models of attitude change. — In: R. Brown, E. Galanter, E. H. Hess, G. Mandler. *New directions in psychology*. N. Y., Holt, Rinehart & Winston, 1962.
15. B r u n s w i k E. *Perception and the representative design of psychological experiments*. Berkeley, Univ. of California Press, 1956.
16. B r y a n J., T e s t M. A. Models and helping: naturalistic studies of aiding behavior. *J. of Pers. and Soc. Psychol.*, 1967, 6, 400—407.
17. C a m p b e l l D. T. The indirect assessment of social attitudes. *Psychol. Bull.*, 1950, 47 (1), 15—38.
18. C a m p b e l l D. T. On the possibility of experimenting with the «bandwagon» effect. *Intern. J. of Opin. and Att. Res.* 1951, 5 (2), 251—260. Reprinted in: H. Hyman, E. Singer (eds.). *Readings in Reference Group Theory and Research*. N. Y., The Free Press, 1968, 452—460.
19. C a m p b e l l D. T. Factors relevant to the validity of experiments in social settings. *Psychol. Bull.*, 1957, 54 (4), 297—312.

20. C a m p b e l l D. T. Methodological suggestions from a comparative psychology of knowledge processes. *Inquiry* (Univ. of Oslo Press), 1959, 2, 152—182.
21. C a m p b e l l D. T. Recommendations for APA test standards regarding construct, trait, or discriminant validity. *Amer. Psychologist*, 1960, 15, 546—553.
22. C a m p b e l l D. T. Variation and selective retention in socio-cultural evolution. — In: H. R. B a r r i n g e r, G. I. B l a n k s t e n, R. W. M a c k (eds.). *Social change in developing areas: a reinterpretation of evolutionary theory*. Cambridge, Mass., Schenkman, 1965, 19—49.
23. C a m p b e l l D. T. Pattern matching as an essential in distal knowing. — In: K. R. H a m m o n d (ed.). *Egon Brunswik's Psychology*. N. Y., Holt, Rinehart and Winston, 1966, 81—106.
24. C a m p b e l l D. T. Evolutionary Epistemology. — In: P. A. S c h i l p p (ed.). *The philosophy of Karl R. Popper*. — In: *The library of living philosophers*. La Salle, Illinois, The Open Court Publishing Co. 1974, vol. 14—1; 413—463.
25. C a m p b e l l D. T., F i s k e D. W. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol. Bull.*, 1959, 56 (2), 81—105.
26. C a m p b e l l D. T., S t a n l e y J. C. *Experimental and quasiexperimental designs for research*. Chicago, Rand McNally, 1966. (См. перевод в данной книге.)
27. C a m p b e l l D. T., S i e g m a n C. R., R e e s M. B. Direction-of-wording effects in the relationships between scales. *Psychol. Bull.*, 1967, 68, 293—303.
28. C h r i s t i e R. Authoritarianism reexamined. — In: R. C h r i s t i e, M. J a h o d a (eds.). *Studies in the scope and the method*

- of the authoritarian personality. N. Y., The Free Press, 1954, 123—196.
29. C o r s i n i R. J. Understanding and similarity in marriage. *J. of Abn. and Soc. PsychoL*, 1956, 52, 327—332.
  30. C o o k S. W. Desegregation and attitude change. Address to the Southeastern Psychological Association, 1964 (Mimeo).
  31. C o o k T. D., B e a n J. R., C i a l d i n i R. B., K r o v e t z M. L., R a y A. A. Three contexts of ingratiation, and their effects on attributions, affect, and donating to charity: Two field experiments, 1969.
  32. C o o k T. D., I n s k o C. A. Persistence of attitude change as a function of conclusion reexposure: A laboratory-field experiment. *J. of Pers. and Soc. PsychoL*, 1968, 9, 322—328.
  33. C r o n b a c h L. J. Response sets and test validity. *Educ. and PsychoL Measurement*, 1946, 6, 475—494.
  34. C r o n b a c h L. J. Further evidence on response sets and test design. *Educ. and Psychol Measurement*, 1950, 10, 3—31.
  35. C r o n b a c h L. J. Proposals leading to analytic treatment of social perception scores. — In: R. T a g i u r i, L. P e t r u l l o (eds.). *Person perception and interpersonal behavior*. Stanford Univ. Press, 1958, 353—379.
  36. *Daily Palo Alto Times*. Psychologists protest tests by Army to see recruits reaction to danger. Thursday, August 13, 1959, p. 5.
  37. D a r l e y J. M., L a t a n é B. Bystander intervention in emergencies: Diffusion of responsibility. *J. of Pers. and Soc. PsychoL*, 1968, 8, 377—383.
  38. D o o b A. N., G r o s s A. E. Status of frustrator as an inhibitor of horn-honking responses. *J. of Soc. PsychoL*, 1968, 76, 213—218.
  39. E d w a r d s A. L. *The social desirability variable in personality assessment and research*. N. Y., Dryden, 1957.
  40. F e l d m a n H. *Problems in labor relations*. N. Y., Macmillan, 1937.
  41. F e l d m a n R. E. Response to compatriot and foreigner who seek assistance. *J. of Pers. and Soc. PsychoL*, 1968, 10, 202—214.
  42. F e s t i n g e r L., C a r l s m i t h J. M. Cognitive consequences of forced compliance. *J. of Abn. and Soc. PsychoL*, 1959, 58, 203, 210.
  43. F r a n z e n R. Scaling responses to graded opportunities. *Public Opinion Quarterly*, 1950, 14, 484—490.
  44. F r e e d m a n J. L., F r a s e r S. C. Compliance without pressure: The foot-in-the-door technique. *J. of Pers. and Soc. PsychoL*, 1966, 4, 195—202.

45. French J. R. P. Organized and unorganized groups under fear and frustration. *Univ. of Iowa Studies in Child Welfare*, 1944, 20, 229—309.
46. Clock C. Y. The effects of reinterviewing in panel research, 1958; Duplicated research report, 1958.
47. Gore P. M., Rotter J. B. A personality correlate of social action. *J. of Personality*, 1963, 31, 58—64.
48. Gosnell H. F. Getting out the vote: an experiment in the stimulation of voting. Chicago, Univ. of Chicago Press, 1927.

237

49. Gross A. E. Some determinants of honesty in a naturalistic situation. Talk presented at the Western Psychological Association, San Diego, California, March 1968.
50. Gross A. E., Collins B., Byran J. Experiments in social psychology. Mimeo.
51. Hammond K. R. Representative vs. systematic design in clinical psychology. *Psychol. Bull.*, 1954, 51, 150—159.
52. Hartmann G. W. A field experiment on the comparative effectiveness of «emotional» and «rational» political leaflets in determining election results. *J. of Abn. and Soc. Psychol.*, 1936, 31, 99—114.
53. Helson H., Blake R. P., Mouton J. S. Petition-signing as adjustment to situational and personal factors. *J. of Soc. Psychol.*, 1958, 48, 3—10.
54. Hempel C. G. Philosophy of natural science. Englewood Cliffs. Prentice Hall, 1966.
55. Hovland C. E., Janis I. L., Kelley H. H. Communication and persuasion. New Haven, Yale Univ. Press, 1953.
56. Jung A. F. Price variations among automobile dealers in Chicago, Illinois. *J. of Business*, 1959, 32, 315—326.
57. Kelman H. C. The human use of human subjects. *Psychol. Bull.*, 1967, 67, 1 — 11.
58. Kerr W. A. Experiments on the effect of music on factory production. *Applied Psychol. Mono.*, 1945, 5.
59. Kiddler L., Campbell D. T. The indirect testing of social attitudes. — In: G. Summers (ed.). *Attitude Measurement* Chicago, Rand McNally, 1970, 333—385.
60. Кун Т. Структура научных революций. М., 1977.
61. Kutner B., Wilkins C., Yarrow P. R. Verbal attitudes and overt behavior involving racial prejudice. *J. of Abn. And Soc. Psychol.*, 1952, 47, 649—652.
62. LaPiere R. T. Actions versus actions. *Social Forces*, 1934, 13, 230—237.
63. Lazarsfeld P. F. Training guide on the controlled experiment in social research. Columbia University, 1948 (Mimeo).
64. Lefkowitz M., Blake R. R., Mouton J. S. Status factors in pedestrian violation of traffic signals. *J. of Abn. and Soc. Psychol.*, 1955, 51, 704—706.
65. Merritt C. B., Fowler R. G. The pecuniary honesty of the public at large. *J. of Abn. and Soc. Psychol.*, 1948, 43, 90—93.
66. Milgram S. Behavioral study of obedience. *J. of Abn. and Soc. Psychol.*, 1963, 67, 371—378.
67. Milgram S., Mann L, Harter S. The lost-letter technique: A tool of social research. *Public Opinion Quarterly*, 1965, 29, 437—438.
68. Miller N., Levy B. H. Defaming and agreeing with the communication as a function of emotional arousal, communication extremity, and evaluative set. *Sociometry*, 1967, 30, 158—175.
69. Naroll R. Data quality control. N. Y., The Free Press, 1962.
70. Page E. B. Teacher comments and student performance: A seventy-four classroom experiment in school motivation. *J. of Educ. Psychol.*, 1958, 49, 173—181.
71. Petrie H. G. The strategy sense of «methodology». *Philosophy of Science*, 1968, 35, 248—257.

238

72. Polany M. Personal knowledge: Toward a post-critical philosophy. London, Routledge and Kegan Paul, 1958.
73. Popper K. R. The Logic of Scientific Discovery. London, Hutchinson. N. Y., Basic Books, 1959.

74. Popper K. R. *Conjectures and Refutations*. London, Routledge and Kegan Paul, N. Y., Basic Books, 1963.
75. Quine W. V. *From a logical point of view*. Cambridge, Mass. Harvard Univ. Press, 1953.
76. Rankin R. E., Campbell D. T. Galvanic skin response to Negro and white experimenters. *J. of Abn. and Soc. Psychol.*, 1955, 51 (1), 30—33.
77. Roethlisberger F. J., Dickson W. J. *Management and the worker*. Cambridge, Mass., Harvard Univ. Press, 1939.
78. Rokeach M., Mezei L. Race and shared belief as factors in social choice. *Science*, 1966, 151, 167—172.
79. Rorer L. G. The great response-style myth. *Psychological Bulletin*, 1965, 65, 129—156.
80. Rosenblatt P. C., Miller N. Experimental method. — In: C. G. McClintock (ed.). *Experimental Social Psychology*. N. Y., Holt, Rinehart, and Winston, 1972.
81. Rosenthal D., Frank J. O. Psychotherapy and the placebo effect. *Psychol Bull.*, 1956, 53, 294—302.
82. Rosenthal R. *Experimenter effect in behavioral research*. N. Y., Appleton-Century-Crofts, 1966.
83. Rosenthal R., Rosnow R. L. (eds.). *Artifact in Behavioral Research*. N. Y. Acad. Press, 1969.
84. Salmon W. *Logic*. Englewood Cliffs, Prentice Hall, 1963.
85. Schwartz R. D., Skolnick J. H. Two studies of legal stigma. *Social Problems*, 1962, 10, 133—142.
86. Schwartz R. D., Orleans S. On legal sanctions. *Univ. of Chicago Law Review*, 1967, 34 (2), 274—300.
87. Segall M. H., Campbell D. T., Herskovits M. J. *The influence of culture on visual perception*. Indianapolis, Bobbs-Merrill, 1966.
88. Sellitz C., Jahoda M., Deutsch M., Cook S. W. *Research methods in social relations*. N. Y., Holt-Dryden, 1959.
89. Silverman L. H. A. Q-sort study of the validity of evaluations made from projective techniques. *Psych. Mono.*, 1959, 73 (7, Whole № 477).
90. Sommer R. Studies in personal space. *Sociometry*, 1959, 22, 247—260.
91. Stollak G. E. Obedience and deception research. *Amer. Psychologist*, 1967, 22, 678.
92. Toulmin S. *The philosophy of science*. London, Hutchinson, 1953.
93. Toulmin S. *Foresight and understanding: An inquiry into the aims of science*. Bloomington, Indiana Univ. Press, 1961.
94. Webb E. J., Campbell D. T., Schwartz R. D., Sechrest L. B. *Unobtrusive measures: nonreactive research in the social sciences*. Chicago, Rand McNally, 1966.
95. Wilson E. B. *An introduction to scientific research*. N. Y., McGraw-Hill, 1952.

### **КАЧЕСТВЕННОЕ ЗНАНИЕ В ИССЛЕДОВАНИЯХ ДЕЙСТВИЯ<sup>1</sup>**

Теперь, когда со дня первого вручения настоящей премии прошло 26 лет, от ее лауреатов со все усиливающимся интересом ждут рассказов о любых непосредственных контактах с Куртом Левином. В 30-е годы в Университете Беркли (я поступил туда в 1937 г.) мы знали о Левине и по-своему его чтили. Толмен широко ссылался на него в своем классическом труде 1932 г. [89] и пытался привлечь его к работе в Беркли. (Ходили слухи, что администрация заставляла Толмена выбирать среди ученых-эмигрантов между Левином и Брунsvиком.) Левин написал одну из глав «Учебника по детской психологии» Мэрчисона [66], и я ознакомился с ней, когда работал с Гаролдом Джонсом. В курсе лекций

«Современные европейские теории личности», прочитанном Эгоном Брунsvиком в 1938 г., Левин был центральной фигурой. Я, таким образом, был хорошо подготовлен к визиту Левина в Беркли летом 1939 г.

В те времена летние факультативные занятия были весьма напряженными. Левин вел два курса для старших студентов — по детской психологии и по психологии личности. Я посещал и тот и другой, и у меня до сих пор сохранились конспекты лекций. Это был удивительный преподаватель. К тому же я имел счастье быть допущенным в узкий круг участников дискуссионного семинара по теории Левина, который вели Левин, Толмен и Брунsvик, а аудиторию составляли шесть или

---

<sup>1</sup> Выступление по случаю вручения премии Курта Левина на совместном заседании Общества психологических исследований социальных проблем и Американской психологической ассоциации (Нью-Орлеан, 1 сентября 1974 г.). Campbell D. T. Kurt Lewin Award Address. Society for the Psychological Study of Social Issues, Meeting with the American Psychological Association. New Orleans, Sept. 1, 1974.

семь человек, в большинстве своем аспиранты. (Мне запомнилось, в частности, что там участвовал Ричард Крачфилд.) Так что мое пребывание в роли ученика Левина было хоть и кратковременным, но исключительно насыщенным. С тех пор я считаю себя отчасти левиновцем и уделяю особое внимание выявлению изоморфизма между его и другими поведенческими теориями (Campbell [13, 18]).

В те дни мы произносили его фамилию «Лювин» (Luhveén), когда же его стали звать «Левин» — в военные и первые послевоенные годы, в период его работ по групповой динамике, — у меня не было с ним никаких непосредственных контактов. Тем не менее благодаря визитам Джона Хардинга я был постоянно осведомлен о работе основанного Левином центра по исследованию действий, Комитета общественных взаимоотношений, который работал в Нью-Йорке с 1944 по 1951 г., имея у себя в штате Стюарта Кука, Мортон Дойча, Айсидора Чэйна, Джона Хардинга, Клэр Зеллиц, Мариан Радке, Гаролда Прошански, Альберта Пепитоне, Дороти Диннерстайн, Сьюзен Дери и др. При подготовке материалов к данному докладу Стюарту Куку, Айсидору Чэйну и Джону Хардингу пришлось изрядно потрудиться, чтобы раздобыть для меня неопубликованные отчеты Комитета об «исследованиях действия», «обследованиях опыта» («experience survey») и общественных самообследованиях («community self-surveys»).

Мои собственные интересы в последние годы сосредоточены на методологии социального экспериментирования, на оценивании экспериментальных программ, на трактовке социальных реформ как экспериментов (Campbell [17]). Поскольку мои занятия несут на себе печать левиновских исследований действия, я допускаю, что они оправдывают получение мной премии Курта Левина. Поэтому я и выбрал для своей лекции именно данную сферу и использовал в заголовке словосочетание «исследования действия», а не «оценивание программ»<sup>1</sup>.

---

<sup>1</sup> Как отмечает сам автор, данная статья не столь типична для его работы в этой области, как ему хотелось бы. Наиболее важной работой в этом направлении автор считает свой неопубликованный доклад «Методы для экспериментирующего общества» [20]. Некоторые идеи этой работы нашли отражение и в статьях, представленных в данной книге. — *Прим. ред.*

Прежде чем начать, я хотел бы поделиться с вами теми чувствами, которые я испытываю в связи с предстоящим выступлением. Когда я около десяти месяцев назад остановился на данной теме, я полагал, что она таит в себе известное потенциальное продуктивное на-

пряжение. С одной стороны, я снискал репутацию ревностного сторонника количественного экспериментального подхода к оцениванию программ (см., например, Salasin [79]). С другой стороны, в своих экскурсах в эпистемологию я отдавал предпочтение обыденному, качественному, зависящему от контекста знанию. Случалось мне заниматься и антропологией, к которой я питал определенные симпатии.

К несчастью, настоящий случай вызвал во мне напряжение, которое сохраняется до сих пор, а творческое прозрение, интегративный подход по-прежнему от меня ускользают. Проблема эта представляется мне теперь еще более важной и значительной, чем вначале. Ощущение, что я вполне подготовлен к ней, исчезло. Во многих местах я наткнулся на горы самым непосредственным образом относящейся к ней литературы, в которой я несведущ, и на давно поставленные проблемы, которые я еще не продумывал. Я могу лишь надеяться, что это предварительное сообщение поможет моим коллегам, работающим над данной проблемой.

\* \* \*

В настоящее время методология оценивания программ занята энергичными поисками альтернатив количественно-экспериментального подхода. В академической социальной науке вновь делается упор на методы гуманитарных наук и вновь растут сомнения в приемлемости модели естественных наук для социальных наук. В этой связи я и говорю о полярности качественного и количественного подхода, подразумеваемой в заглавии. Эти термины представляют собой сокращенные обозначения некоего общего содержания целого ряда частично перекрывающих друг друга понятий. Под *количественным* имеется также в виду научный, сциентистский, естественнонаучный, *naturwissenschaftlich*. Под *каче-*

242

*ственным* понимается также гуманистический, гуманитарный, *geisteswissenschaftlich*, опытный, феноменологический, клинический, исследование случаев (прецедентов), полевая работа, «включенное» наблюдение, оценивание и обыденное знание.

Хотя поначалу я буду обсуждать эти вопросы в плане эпистемологии и социальных наук в целом, моим главным объектом будет, в конечном счете, оценка последствий намеренно осуществляемых социальных инноваций. Я употребляю термин «*знание*» («*knowing*»), а не «оценка», чтобы указать на стремление подвести под эти вопросы эпистемологическое основание. Словосочетание *исследования действий* особенно подходит к исследованиям, сосредоточенным на сознательных попытках осуществить социальное действие. В том виде, в каком этот тип исследований разработали Левин и его сотрудники по героическому Комитету общественных взаимоотношений (Lewin [61, 62]; Luppitt, Radke [63]; Chein, Cook, Harding [31, 32]; Selltitz, Cook [84]; Cook: [33]; Harding [51]; Chein [29, 30]; Marrow [64, с. 191—218]; Selltitz [83]), он предполагал мудрое сочетание качественного и количественного подходов. Он обладал к тому же одной важной для данного контекста особенностью, которая не получила отражения в заглавии: разработкой процедур, посредством которых группы действия могут измерять свои собственные достижения (самооценка программы). По всем этим причинам название *исследование действия* лучше, чем оценивание программ, не говоря уже о том, что оно символизирует преемственность в отношении работ Курта Левина.

Следующий далее анализ начинается с эпистемологических соображений, призванных наметить какую-то единую перспективу для количественного и качественного знания, согласующуюся с критическим анализом, предпринимаемым современной философией науки. Он приводит к заключению, что, выходя за пределы качественного знания, количественное знание находится в зависимости от него. Эта зависимость слабо представ-

лена в количественной социальной науке. Затем обсуждаются качественные основы количественных социальных наук и роль качественного знания в оценивании программ.

243

### Некоторые проблемы дескриптивной эпистемологии

Научный статус нелабораторной социальной науки в лучшем случае сомнителен. Но даже в самых строгих науках теории, которые принято считать истинными, обоснованы совершенно недостаточно и могут рассчитывать самое большее на статус теории, которая «лучше, чем», но не теории, которая «доказана». Всякие обыденные и научные знания гипотетичны. В любых условиях, когда нами, как мы полагаем, добыты новые знания, они получены ценой многих непроверяемых предположений, не говоря уже об их неподтвержденности в данной ситуации. Хотя некоторые предпосылки могут быть проверены по одной или в небольших сочетаниях, это возможно лишь за счет допущения правильности огромного множества других предпосылок. Единичные предпосылки или их небольшие сочетания могут подвергаться апробации, но валидность всей совокупности предпосылок не может быть доказана, она явно страдает недостаточной обоснованностью. Таковы пессимистические заключения большинства современных работ в области философии науки (Quine [75, 76]; Goodman [45]; Hanson [50]; Polanyi [71]; Popper [72, 73, 74]; Toulmin [91]; Kuhn [57, 58]; Feyerabend [39]; Lakatos [59]).

Но все же наука гораздо лучше невежества и во многих вопросах лучше традиционной мудрости. Наша задача как методологов — найти свой путь между крайностями пассивного скептицизма и наивной доверчивости. Может оказаться полезным опыт эволюционной эпистемологии, поскольку она представляет обыденное знание как основанное на предпосылках, встроенных в сенсорную или нервную систему и в повседневный язык, а эти предпосылки хорошо проверяются и тщательно обрабатываются и таким образом получают косвенное подтверждение посредством естественного отбора в ходе биологической и социальной эволюции<sup>1</sup>. В случае на-

---

<sup>1</sup> См. об этом: К э м п б е л л Д. Т. Слепые вариации и селективный отбор как главная стратегия процессов познания. — В кн.: Самоорганизующиеся системы. М., «Мир», 1964; е г о ж е: Социальные диспозиции индивида и их групповая функциональность. Эволюционный аспект. — В кн.: Психологические механизмы регуляции социального поведения. М., «Наука», 1979. — *Прим. ред.*

244

учных теорий, еще более гипотетичных, процесс естественного отбора получает свое продолжение в выборе между соперничающими теориями из числа существующих (Popper [72, 73, 74]; Campbell [10, 11, 16, 19]; Toulmin [90, 91]; Quine [76]; Caws [28]; Shimony [87]).

При таком подходе, чтобы внести некоторое успокоение в души представителей социальных наук, где для того, чтобы прийти к каким-нибудь выводам, нужно нагромоздить массу предпосылок, конкретные действия «методологии» можно квалифицировать следующим образом. *Когда ученый утверждает, что имеющаяся совокупность данных свидетельствует в пользу некоей теории, признание этого утверждения недействительным исходит на практике только от столь же или более правдоподобных объяснений тех же данных.* Само по себе существование бесчисленного множества логически возможных альтернатив не является на практике опровержением. Как мы знаем со времен Юма, научные индукции недоказуемы ни дедуктивно, ни индуктивно. Мы не должны позволять нашим критикам приводить конкретные примеры этого общего

«недоразумения с индукцией», как будто это является исключительной особенностью нашего исследования или нашей области исследований.

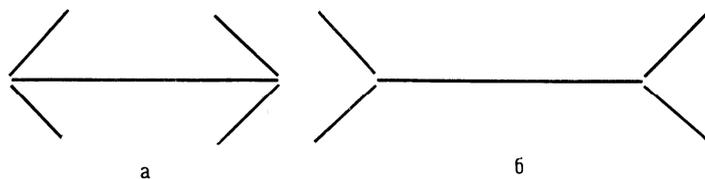
### Зависимость науки от обыденного знания

Мы не должны полагать, что научное знание заменяет обыденное знание. Наука, скорее, находится в зависимости от здравого смысла, даже если она в лучшем случае выходит за его пределы. Наука в конечном счете вступает в противоречие с некоторыми положениями здравого смысла, но она делает это, лишь доверяя большей части остального обыденного знания. Подобная ревизия здравого смысла наукой аналогична ревизии здравого смысла другим здравым смыслом, что, как это ни парадоксально, может быть осуществлено лишь путем доверия здравому смыслу в целом.

Рассмотрим в качестве примера иллюзию Мюллер-Лайера (рис. 1). Если мы спросим обычного представителя «прямоугольной» культуры (Segall et al. [82]), какая линия длиннее,  $a$  или  $b$ , он ответит —  $b$ . Если бы мы снабдили его линейкой или позволили воспользо-

245

ваться вместо линейки кромкой другого листа бумаги, он, в конце концов, убедился бы, что был неправ и что  $b$  не длиннее  $a$ . Принимая такое решение, он отвергает как неточный один продукт зрительного восприятия, доверяя более обширной совокупности других зрительных перцептивных данных. Он, кроме того, вынужден будет сделать ряд предположений, большей частью неявных, в том числе предположения о том, что размеры линий в процессе измерения оставались относительно постоянными, что линейка была твердой, а не эластичной, что температура его руки и ее влажность не влияли на длину линейки таким образом, чтобы в результате случайного совпадения она давала разные результаты, уве-



Р и с. 1. Иллюзия Мюллер-Лайера.

личиваясь при приближении к линии  $a$  и уменьшаясь при приближении к линии  $b$ .

Возьмем в качестве еще одного примера научную статью, которая излагает теорию и экспериментальные результаты, демонстрирующие квантовую природу света и резко контрастирующие с представлениями здравого смысла. Если бы авторы подобных статей отказывались от обыденного языка, ограничивались математическими символами и чисто научными терминами, им не удалось бы вступить в контакт с другими учеными, чтобы те смогли повторить эксперименты и верифицировать наблюдения. Напротив, немногочисленные научные термины вкраплены в тексты, написанные на донаучном обыденном языке, который читатель, как предполагалось (и предполагается им самим), понимает. Этот язык явно несовершенен, эллиптичен, метафоричен и неоднозначен. В ходе лабораторной работы при постановке и повторении экспериментов для выработки заключений, в которых пересматривалось обыденное понимание объектов, тел и света, были использованы и приняты с

246

доверием и обыденный донаучный язык, и обыденное восприятие этих явлений.

Чтобы бросить вызов здравому смыслу и скорректировать обыденное понимание в какой-то частности, следует доверять обыденному пониманию в целом.

### Соотношение сомнения и доверия

Эти два примера служат прологом к разговору о преемственности между обыденным и научным знанием и о соотношении сомнения и доверия. Если мы выбираем полный скептицизм или солипсизм, мы отказываемся от «знания», или науки. Если мы выбираем полную доверчивость (что, по всей вероятности, имеет место в некоторых версиях феноменологии и «непосредственного реализма»; Campbell [18, 22]), мы отказываемся от последовательности, простоты и от расширения или совершенствования «знаний». Повседневное знание и наука располагаются между этими двумя крайностями и каким-то образом сочетают в себе способность концентрированного недоверия и пересмотра с верованием в некую совокупность провозглашенных знаний. Одним из аспектов этого процесса, который делает возможной кумулятивную ревизию науки, является практика доверия (по крайней мере предположительного) к подавляющей массе имеющихся в настоящее время научных и обыденных верований и представлений («знаний») и их использование для дискредитации и пересмотра какого-то одного аспекта научных убеждений. Отношение подвергаемого сомнению к пользующемуся доверием всегда очень маленькая дробь. Это выражено в метафоре, заимствованной Куайном [75] у фон Нейрата: мы напоминаем моряков, которым необходимо починить в открытом море корабль с гниющей обшивкой. Мы полагаемся на большую часть шпангоутов, хотя заменяем какую-то особенно ненадежную доску. Очередь на замену может подойти для любого из брусев, которым мы в данный момент доверяем. Пропорция досок, которые мы заменяем, сравнительно с теми, которые мы полагаем надежными, всегда должна быть мала. Или, по словам самого Куайна: «Целокупность нашего так называемого знания или верований от самых что ни на есть казуальных, с которыми имеет дело география и

247

история, до самых глубинных законов атомной физики или даже чистой математики и логики — это творение человека, которое подвержено столкновению с опытом только вдоль внешних границ... Конфликт с опытом на периферии обуславливает перестройку во внутреннем поле... Однако поле в целом настолько не детерминировано своими пограничными условиями, опытом, что существует значительная широта выбора утверждений, подлежащих переоценке в свете какого бы то ни было единичного противоречащего им опыта. Жесткий опыт может... быть освоен с помощью любой из самых различных альтернативных переоценок в самых разных участках общей системы... однако... наша естественная тенденция состоит в том, чтобы вносить как можно меньше нарушений в систему в целом» (Quine [75, с. 42, 44]).

По поводу гипотетического характера даже самых обычных и базовых явлений восприятия и обыденного знания существует широко известный трюизм, согласно которому нет упрямых фактов, которые говорят сами за себя и на которых может быть испытана истинность теории. Напротив, так называемые факты сами «обременены теорией» (Hanson [50], Kuhn [57, 58], Feysabend [39]). На этом моменте стоит остановиться. Иллюзия Мюллер-Лайера, использованная выше, вполне может быть обусловлена имплицитной «теорией» отношений в окружающей среде, встроенной в нервную систему научением или генетическим наследованием, по которой тупые и острые углы образуются в поле зрения прямоугольными фигурами (Segall et al. [82], Stewart [88]). Иллюзия точки и рамки Дункера [36] может служить более убедительным примером. В

совершенно темной комнате находится большая светящаяся рама со светящейся точкой внутри нее. Рама смещается на несколько дюймов вправо. Наблюдатель, напротив, видит, что точка сместилась на несколько дюймов влево. Перцептивная система содержит в себе встроенные предсознательные «деревя решений», которые преобразуют данные об относительном смещении в вывод об абсолютном смещении. Используемая при этом «предпосылка», или «теория», состоит в следующем: в сомнительном случае смещается небольшой фрагмент зрительного поля, а на месте остается большая его часть. Это прекрасное общее правило, но в экологически атипичных

248

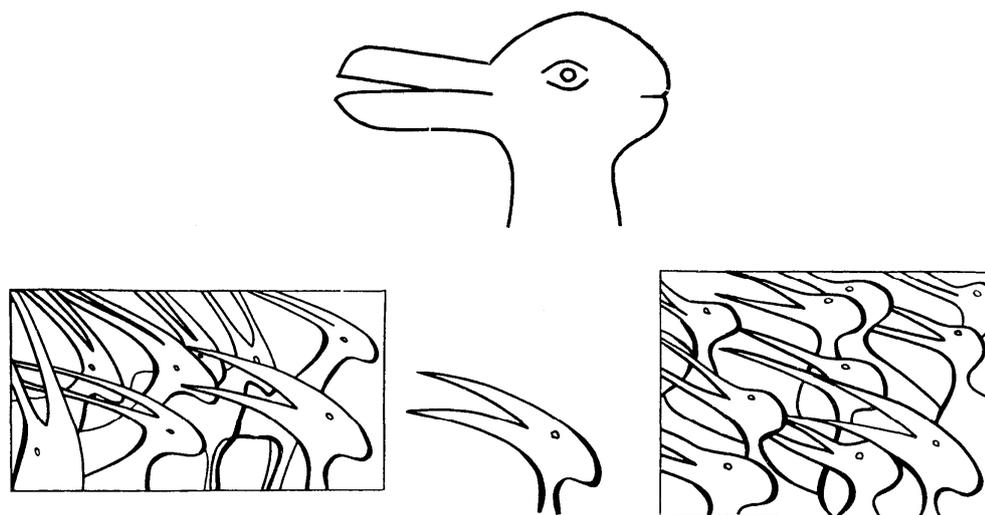
условиях Дункеровой лаборатории оно неприменимо. Это «обремененность теорией» на уровне естественного знания. Когда нам приходится читать показания гальванометра или сцинтилляционного счетчика, обремененность полученных фактов «теорией» носит еще более очевидный «теоретический» характер, так как для того, чтобы интерпретировать показания приборов, следует доверять эксплицитным научным теориям.

Подчеркивание этого может вести к неправильному пониманию. Когда Кун [57] говорит о научных революциях, он, видимо, предполагает, что все физическое знание связано в одну-единственную целостную теорию, в одно-единственное уравнение и что «факты», на которых испытывается истинность этой теории, обременены именно этой, и только этой, всеобъемлющей теорией. Предполагается, что когда эта ведущая теория меняется, то меняются одновременно все «факты». То же самое утверждают Хэнсон [50] и Фейерабенд [39].

Это предположение о единственной всеохватывающей теории ошибочно. Реальное положение науки таково, что она на самом деле «интегрирована» или «унифицирована» в гораздо меньшей степени. Обремененность «фактов» теорией предполагает многочисленные глубоко укоренившиеся предпосылки, не ниспровергаемые какой-то конкретной научной революцией. Напротив, существует разработанная иерархия верований — предпосылок — теорий, в которой более простые уровни здравого смысла остаются в значительной части неизменными для основывающихся на них уровней до и после революции. Точно так же в ходе научной революции пользуются доверием 90% зависящих от инструментария «фактов» лабораторного уровня. Конечно, некоторые из этих фактов, полученных в предреволюционные периоды, вызывают аномалии, которые служат главным рычагом для ниспровержения старой парадигмы.

Не подлежит сомнению наглядная убедительность двойственной фигуры «утка/заяц» (см. рис. 2), приобретшей известность благодаря Джастроу [52], Витгенштейну [96], Хэнсону [50] и Куну [57, 58]. Факты *действительно* воспринимаются в ином ракурсе или под иным углом зрения после смены «теории» *утки* «теорией» *зайца*. Однако наблюдается постоянство рисунка на бумаге, по отношению к которой переход от

249

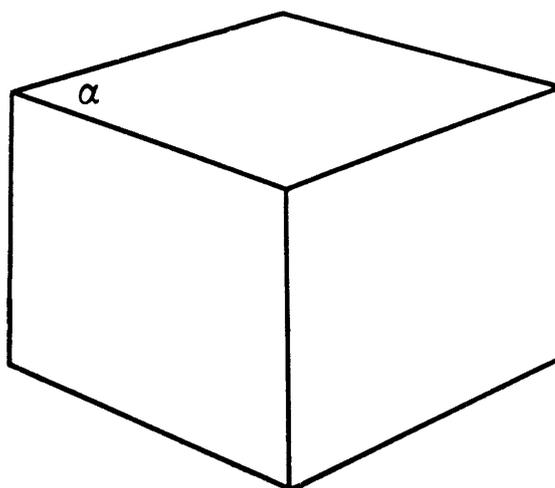


Р и с. 2. Что вы видите: утку или зайца? (По: Jastrow [52, с. 295].) Джастроу ссылается на издание «Харперс Уикли», которое в свою очередь опирается на другое — «Флигенде Блаттер». Витгенштейн [96, с. 194] ссылается на Джастроу, но использует свой собственный небольшой сделанный от руки рисунок, вызывая удивление тем, что приводит в своей работе единственный этот рисунок. Хэнсон [50] и Кун [57] ссылаются на Витгенштейна, но не приводят рисунка. Прекрасной иллюстрацией является рисунок Хэнсона [50, с. 13, 14] «антилопа — аист».

«утки» к «зайцу» есть измененное решение (или истолкование). Рисунок представляет собой постоянно существующее общее ограничение возможных интерпретаций. «Утка» или «заяц» — это «непримиримые» *интерпретации* общей загадки. Существует очень мало таких приемлемых, практически доступных интерпретаций, их число ограничено, вероятно, этими двумя. Более того, рисунок, который лежит, по-видимому, в основе «факта», составляющего фундамент этих двух «теорий», сам является зрительным образом, обремененным теорией. Если изучить его вплоть до микроскопических деталей, его фактурная линейная организация превращается в дискретные частички краски, а те в свою очередь — в структуры, к которым категории цвета и краски неприменимы. Но эта обремененная теорией гипотетичность низшего уровня, порождающая линейную конфигурацию, при переходе от «утки» к «зайцу» не изменяется, хотя происходит, возможно, изменение восприятия некоторых отдельных углов или других линейных отношений. Так, угол  $a$  на рис. 3 воспринимается как значительно бо-

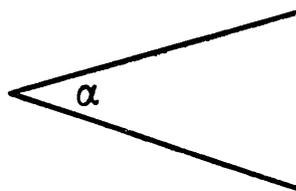
лее близкий к  $90^\circ$ , когда он интерпретируется как часть прямоугольника, чем когда он предъявлен в виде отдельного фрагмента (рис. 4); во втором случае соответствующая «теория» менее убедительна. (Можно по-

Р и с. 3.



интересоваться, нет ли здесь просто-напросто перехода от одного понимания вопроса к другому. Действительно, о каком угле идет речь — изображенном на бумаге или репрезентированном? Но даже если всем ясно, что оценке подлежит угол на бумаге, в составе прямоугольника он выглядит менее острым, чем отдельно от него.) Подобные же эффекты могут наблюдаться и при пере-

Р и с. 4.



ходе от «утки» к «зайцу». Я, таким образом, не собираюсь оспаривать приводимые Хэнсоном, Куном и Фейерабендом конкретные примеры ряда «фактов» более низкого уровня, которые претерпевают резкие изменения в ходе научной революции. Что я оспариваю, так это то, что все факты или даже большая их часть претерпевают подобные изменения. Я не согласен с тем, что обремененность теорией, присущая всем фактам из соответствующей области, определена (будь то до или после научной революции) какой-то отдельно взятой интегрированной всеобъемлющей теорией. То же самое утверждают Блахович [7] и Кордиг [56].

**Первичность распознавания целостного образа  
по отношению к знанию деталей**

Еще один аспект такой дескриптивной эпистемологии — указание на гностическую (epistemic) первичность целостных образов, а не фрагментов (Campbell [16]). Возьмем задачу распознавания «одной и той же» точки на двух газетных оттисках одной и той же фотографии. Пусть каждая картинка будет накрыта большим листом бумаги с маленьким отверстием в нем. Когда такой редуцированный экран экспонирует по одной точке, задача не может быть решена: взятый в отдельности, фрагмент почти всегда допускает множество толкований. Когда экспонируется квадратный сантиметр или дюйм, задача уже не столь непосильная, потому что появляется возможность распознавания целостного образа — чем больше образ, тем определеннее распознавание. Коль скоро достигается распознавание

соответствующих точек, это происходит в силу первичного распознавания целостных образов. Это очевидный трюизм, тем не менее все еще существуют эпистемологические направления, продолжающие искать определенность во фрагментах, будь то дефиниции терминов или фрагментарные сенсорные данные. Все наше знание атомов или мельчайших деталей по крайней мере столь же зависит от контекста и столь же гипотетично, как и знание больших и более сложных целостных образов. Атомарные частицы или сенсорные данные весьма отличны от надежных строительных блоков. Качественное, обыденное знание целокупного и целостных образов служит общим контекстом, необходимым для интерпретации фрагментарных количественных данных.

### **Перекрестная валидизация науки здравым смыслом**

Наука не только зависит от обыденного знания, которое служит надежным основанием для разработки ею сложных эзотерических приемов и квантификации; многие результаты и достижения этой эзотерической науки подвергаются к тому же перекрестной валидизации с позиций здравого смысла. Тайны тригонометрии получают зрительное подтверждение, когда встречаются концы туннеля или моста. Фарадеевы невидимые электрические токи вызывали видимые вспышки и видимое магне-

252

тическое движение железных пластин. Стало быть, даже люди, далекие от науки, обладают способностью воспринимать свидетельства валидности эзотерической науки в терминах здравого смысла. Так, видимые невооруженным глазом инженерные достижения значимым образом валидируют науку в общественном сознании.

Среди самих лабораторных ученых эта перепроверка с позиций здравого смысла является постоянным делом и составляет один из фундаментальных компонентов той амальгамы распознавания и ожидания, которой оправдывается отклонение ими многих количественных показаний их приборов как ошибочных (из-за неверной калибровки, неправильного соединения и т. п.). Такая перекрестная валидизация количественного качественным в настоящее время в социальных науках либо отсутствует, либо проявляется как враждебный критицизм, о котором мы будем говорить ниже. В хорошо оснащенных научных лабораториях появляется иного рода сплав качественного с количественным, где механические квантифицирующие инструменты становятся настолько привычными средствами, что включаются в качественное знание как его органичная часть, как посох слепца. Можно спорить, достигнем ли мы когда-нибудь подобного положения в социальных науках; бесспорно, однако, что оно ни в коем случае не сложится без интенсивного предварительного взаимодействия здравого смысла и научного знания в изучении одних и тех же проблем.

### **Качественное знание в количественном оценивании программ**

Качественные опоры, на которых покоятся количественные данные, могут быть обнаружены путем прослеживания исходных данных для любого прокола на перфокарте или какой угодно численной величины на выходе ЭВМ. За тестовыми показателями какой-нибудь программы из серии программ компенсаторного начального обучения (Head Starts Programs)<sup>1</sup> лежат вербаль-

---

<sup>1</sup> Об оценке программ (и в частности, программ компенсаторного начального обучения) см. также статью «Измерение воздействий, оказываемых запланированным социальным изменением» в данной книге. — *Прим. ред.*

ные или наглядные инструкции, полученные сотрудником, проводившим тестирование, вербальные и наглядные разъяснения, которые этот сотрудник давал детям, качественное понимание вопросов детьми и т. д. В основе опознания детей, подвергшихся тестированию, как тех же самых, которые принимали участие в предшествующих занятиях по данной учебной программе, тоже лежит чье-то качественное знание. Учебные инструкции в рамках этой программы излагались обыденным языком, и если добросовестность их выполнения подвергалась проверке, то это делалось посредством качественных наблюдений. Если практиковались интервью с родителями, то беседа держалась на вере в общность языка и мотивов. Регистрация реакций и кодирование ответов на открытые вопросы получают количественную форму лишь как конечный продукт какого-нибудь процесса качественных суждений. Эти опоры обыденного знания в большинстве своем столь распространены и надежны, что мы не обращаем на них внимания. Однако я уверен, что на таком основании покоится все количественное знание в социальных науках, равно как и в физических.

Это основание столь надежно, что мы оставляем его без внимания, и тем не менее зависимость от качественного знания налицо даже тогда, когда она ничем не оправдана. Критика качественной основы необходима, и она, как выясняется, зависит от качественного знания, полученного из более полного контекста. В настоящее время у нас наблюдается в этом плане ненормальное разделение труда: с одной стороны, ряд ученых в сфере социальных наук придерживаются количественного подхода, они пользуются данными переписей, сведениями о доходах и количественными показателями преступности, забывая, что эти данные основываются на качественных суждениях. С другой стороны, есть социологи, придерживающиеся качественного подхода, они занимаются критикой способов, которыми получены эти цифры, анализируя, например, социальную динамику интервью, проводимого при переписи, различное понимание явлений разными участниками программ, страхи и давление, скрывающиеся за ответами, и, поскольку при этом вырабатывается скептицизм, отказываются принимать участие в проведении количественных анализов.

Особенно решительная критика содержится в литературе, появляющейся в области квантифицированных административных отчетов. Ее суть можно уловить уже по следующим заголовкам: «„Хорошие" организационные основания для „плохих" клинических отчетов» (Garfinkel [42]); «Заметки о применениях официальной статистики» (Kitsuse, Cicourel [55]); «Фальсификация полицейских отчетов о преступности и политическое давление» [Seidman, Couzens [85]; «План Никсона по борьбе с преступностью подрывает статистику преступлений» (Morrissey [65]); «Занижение квалификации преступлений, обнаруженное в Балтиморе» (Twigg [92]). (См. также Becker [4, 5]; Douglas [35]; Beck [3]; Blau [8]; Etzioni, Lehman [37]; Ridgeway [78].)

Подобные же исследования необходимы в связи с тестированием успехов. Если верить множеству различных анекдотов, вычислительные машины слишком часто с бесстрашной серьезностью обрабатывают не представляющие ценности данные, полученные от детей, либо проводивших массовые бойкоты обследований, либо саботировавших процесс тестирования, либо обращавших свои ответы в шутку. Получили распространение анекдоты о подобных же происшествиях, связанных с вопросниками, шкалами по исследованию установок и интервью (Vidich, Bensman [93]). Мы нуждаемся в том, чтобы такая критика была увязана с количественными исследованиями. В настоящее время некоторая связь между этими двумя подходами хотя и намечается, но недостаточна. Известная интеграция этих данных нужна, по крайней мере, в методологии полевого экспериментирования, куда она входит под рубриками «угрозы валидности» или

«правдоподобные конкурентные гипотезы» (см., например, Campbell, Stanley [26], Cook, Campbell [34]<sup>1</sup>). Качественное рассмотрение различных контекстов измерения часто приводит к гипотезам, согласно которым различия между данными предварительного и итогового тестирования или между экспериментальной и контрольной группами проистекают из

---

<sup>1</sup> См. перевод этих работ в данной книге, а также: Webb E. J., Campbell D. T., Schwartz R. D., Sechrest L. Unobtrusive measures: Nonreactive Research in the Social Sciences. Chicago, Rand McNally, 1968. — *Прим. ред.*

различного понимания процесса измерения респондентом и исследователем, а не из изменений свойства, на отображение которого рассчитана данная мера. Некоторые из этих конкурентных гипотез могут быть исключены путем использования формальных свойств плана эксперимента, однако многие из них должны оставаться качественными даже в самом доскональном отчете — в том, что касается способа создания конкурентной гипотезы, выявления других свидетельств в пользу валидности этой гипотезы и общего суждения относительно ее правдоподобия. Там, где дело касается измерений социальных взаимодействий, это составляет проблему, по крайней мере, столь же значительную для рандомизированных «истинных» экспериментов, как и для квазиэкспериментов.

Для многих работ по оценке эффективности экспериментальных (особенно квазиэкспериментальных) программ угрозу валидности содержат многочисленные события, не относящиеся к экспериментальной программе, которые могли вызвать наблюдаемые изменения. Выявление этих конкретных альтернативных объяснений и оценка их правдоподобия являются делом обыденного знания и требуют детального знакомства с конкретными локальными условиями.

Хотя эти правила относительно использования качественного знания ясно сформулированы в пособиях по количественным методам, они редко применяются. Исследователь склонен полагать, что открыто говорить о таких вещах — значит ставить под угрозу видимость научной достоверности или что слабость в этих вопросах свидетельствует о его собственной некомпетентности. Оборонительная позиция полевого экспериментатора в отношении своих коллег — лабораторных исследователей ведет к дальнейшему пренебрежению этой стороной проблемы, тем более, если такой исследователь зарабатывает себе на жизнь, работая по контракту в сфере оценивания программ.

В текущих работах по количественной оценке программ качественное знание обычно представлено в описательной истории программы, а также в описании ее содержания, в количественных мерах и средствах сбора данных и в вербальном обобщении ее итогов. Иначе числовые данные не поддавались бы интерпре-

тации. И все же такое качественное содержание часто представляет собой запоздалое незапланированное размышление.

В этом докладе о роли качественного знания в количественном оценивании программ действия уже содержались указания на настоятельную необходимость эксплицитного признания этой роли. Ниже эта тема получит дальнейшее развитие. Нам придется также — хотим мы этого или нет — учитывать и соображения, возникающие при качественном оценивании программ в отсутствие количественных мер. Прежде чем мы приступим к этому, неплохо было бы, по-видимому, остановиться на типах качественного знания, которые уже существуют или могли бы быть предложены.

Племена, правительства и добровольные объединения на протяжении веков учреждали новые программы, которые затем приостанавливались, поддерживались или распространялись. С точки зрения социальной эволюции (Campbell [15]) такие судьбы программ являются оцениваниями и процессы принятия решений, определяющие судьбу программы, составляют механизмы оценивания. Некоторые из этих механизмов связаны с выживанием организации — своего рода социально-эволюционным естественным отбором. Но гораздо чаще эти решения — и, конечно, заимствование решений — включают в себя людские суждения, которые представляют ценности, необходимые для потенциального выживания. Эти процессы замещающих суждений, вероятно, имеют ценность, приближающуюся к адаптивной, так как они сами являются продуктами биологического и социального естественного отбора.

Сырьем для таких оценок служит накопленный опыт *участников*, преобразованный в институциональное решение посредством некоторого формального или неформального политического процесса. Опыт, накопленный участником, следует трактовать как наиболее примитивное и распространенное качественное знание, имеющее отношение к оцениванию программ. Поскольку память о прошлом непрерывно модифицируется под влиянием наличного контекста и более позднего опыта, особое

значение приобретает *зарегистрированный опыт участников* — например, посредством записей и учета голосов. В случае исследования действия имеется ряд других релевантных групп непрофессиональных наблюдателей, лиц, находящихся в удобной для наблюдения позиции, что не характерно для иных социальных ролей. Сюда входит руководящий персонал программы и те группы граждан, которые находятся в постоянном контакте с лицами, испытывающими на себе действие программы.

Качественная оценка, производимая экспертами по оцениванию (специалистами в области социальных наук), может принимать многочисленные формы. Она может содержать в себе собственный качественный опыт социального исследователя, полученный, например, на какой-то единовременной выборке *при обследовании на месте*, или весь опыт целостного социального действия, включая предпрограммный период, как при *включенном наблюдении*, так и при *наблюдении со стороны*. Но еще более общим для качественного подхода в социальной науке является участие эксперта в регистрации и сопоставлении опыта участников путем *интервью с информантом*, *обследования мнений* или *интервью по поводу опыта*, проводившихся Комитетом по изучению общественных взаимоотношений.

В политических процессах, в ходе которых группы трансформируют сохранившийся в памяти участников опыт в оценочные решения, может применяться голосование — один из способов квантификации качественного опыта в целях объединения субъективных суждений. Точно так же роль социального исследователя в сопоставлении опыта участников может включать и квантификацию, как это имеет место при кодировании свободных интервью, при подсчете числа участников, в целом оценивающих программу негативно, или в случае предъявления оценочных шкал и структурированных вопросов, по которым участники могут квантифицировать свой собственный субъективный опыт. На самом деле многое из того, что мы считаем экспериментальными и количественными показателями, зафиксированными при предварительном и итоговом тестировании, как в экспериментальной, так и в контрольной группе, является фактически квантификацией субъективных

суждений. Социальный исследователь, сторонник качественного подхода, может, кроме того, прибегнуть к количественной обработке при регистрации или суммировании своего собственного непосредственного опыта. Ниже я попытаюсь провести различие между качественным и количественным сравнением качественного опыта, однако и то и другое будет рассматриваться в рамках общей категории качественного.

*Оценивание процесса* предлагалось в качестве дополнения (Freeman, Sherwood [40]) или альтернативы к количественному оцениванию экспериментальных программ (см., например, Weiss, Rein [94], Guttentag [48, 49]). Этот подход в резкой форме представлен в одном фундаментальном положении: «Причина, по которой мы требуем использования «контрольных групп» в экспериментальной науке, состоит в том, что изучаемые процессы протекают, согласно предположению, в знаменитом «черном ящике». Так что мы не можем наблюдать крайне важное внутреннее промежуточное состояние. Мы можем лишь контролировать вход и измерять выход. Но там, где возможно наблюдать промежуточные звенья — сам процесс, — можно обойтись без грубой экспериментальной модели» (Bennis [6, с. 231]).

На практике оценивание процесса предполагает качественное рассмотрение конкретных событий, имеющих место при осуществлении программы, включая типичный опыт участников. Оценивание процесса может производиться самим персоналом, занятым в осуществлении программы, или специалистами по оцениванию процесса, занятыми только этой работой, или, возможно, путем привлечения участников для дневниковой регистрации событий. Типичный продукт такой работы может иметь форму литературно-исторических описаний опыта, связанного с программой. (Каплан [54] говорил о них как об экстенсивных данных, составляющих основу его суммарной оценки моделей города. См. также Gilbert, Specht [43].) Хотя до сих пор нет классических примеров и критических реакций, мы можем быть уверены, что предсказанное Беннисом непосредственное восприятие каузальных связей отсутствует. Однако хорошо спланированное и добросовестно выполненное описание процесса представляется желательным компонентом в оценивании любой программы как ее пере-

крестная валидизация и своего рода критика процесса измерения и экспериментальных планов. (Задание, получаемое обычно оценщиком процесса, должно быть явно сформулировано шире, чтобы оно охватывало эти последние виды деятельности.)

Если бы нам предстояло стать свидетелями широко распространенной практики оценивания процессов и если бы оценщики процессов должны были критиковать друг друга и спорить по поводу измерений эффектов программы, их обоюдный критицизм почти непременно заставил бы их заново изобретать некоторые черты экспериментального плана. Иными словами, ощущалась бы потребность в регистрации сходных процессов, предшествовавших инновации, и в регистрации процессов, которые протекали в сопоставимых социальных образованиях, переживавших те же исторические события, что и изучаемые, за исключением предусмотренных программой. Экспериментальный план может быть отделен от квантификации. Неопределенность вывода, на снижение которой рассчитан экспериментальный план, существует для качественной оценки в той же мере, что и для количественной.

От *суммативного оценивания*, представляющего собой количественное измерение воздействия, оказываемого программой, которое является целью типичного количественного оценивания экспериментальной программы, отличается *формативное оценивание*. Формативное оценивание связано с постоянным контролем за новой программой, спланированным с целью обеспечивать безотлагательные пересмотры тех аспектов программы, которые оказываются «неработающими» или явно неэффективными.

Хотя и не исключено, что какая-то часть этих оценок могла бы быть осуществлена с помощью количественных экспериментов, связанные с ними усложненность, необходимость немедленного реагирования и тот факт, что многие ошибки планирования можно заметить невооруженным глазом, обуславливают то, что формативное оценивание становится по преимуществу делом обыденного знания. Это, несомненно, желательное введение к количественному экспериментированию или к полному оцениванию процесса, поскольку эти дорогостоящие процедуры не следует выполнять в отношении программ, которые те, кто осуществляют эти програм-

260

мы, не желали бы видеть вновь воспроизведенными и распространенными в том же виде. Там, где это возможно, такая формативная «доводка» программы должна предшествовать ее оцениванию. Необходимость формативного оценивания требует внимания и учета. В противном случае возможна реализация программы, спроектированной только на бумаге, и при этом можно не увидеть вполне доступные качественные данные, свидетельствующие о неудаче реализации.

*Системный анализ* представляет собой широко рекомендуемую альтернативу и дополнение к количественному экспериментальному подходу (Baker [1, 2]; Schulberg, Baker [80]; Schulberg, Sheldon, Baker [81]; Weiss, Rein [94]). В то время как в биологии и инженерном деле (см., например, Jones [53]) он является в высшей мере квалифицированным использованием экспериментальных результатов, в оценивании социальных программ он представляет собой пример качественного подхода, для которого характерны внимание к организационному функционированию, связь с другими системами и пристальное внимание к процессам, включенным в программу. Тем самым ему удается избежать узости и близорукости многих количественных оценок. В ожидании дня, когда в нашем распоряжении окажется какой-то исчерпывающий системный анализ новых программ и их воздействий, я не стану пытаться отграничить этот подход от оценивания процессов или методов полевой работы, которые будут описаны ниже, и повторю, что он представляет ценным приложением к количественным оценкам.

Использование *антропологов* в роли оценщиков программ было предпринято в программе национальных институтов образования «Экспериментальные школы». Хотя это и не приводит к полному отказу от количественного измерения, антропологическая полевая работа представляет собой одну из главных исследовательских затрат в довольно дорого обходящихся оцениваниях программ. Антропологи, которых воспринимали как постоянных жителей, работали также по оценке эффективности некоторых «моделей городов».

До Малиновского большинство антропологических исследований представляло собой собирание фактов, отражающих опыт информантов, и часто ставило своей

261

целью реконструировать картину местной культуры до контакта с европейцами. Со времен Малиновского обычной целью стало описание сегодняшней живой культуры какого-то определенного поселения, безотносительно к ее историческим истокам или культурной типичности<sup>1</sup>. В связи с этим антрополога стали рассматривать как непосредственного наблюдателя, участвующего в коллективной жизни данного сообщества настолько полно, насколько это возможно, основательно изучающего местный язык и живущего жизнью сообщества 24 часа в сутки в течение ряда лет. Именно эта последняя антропологическая модель преобладала до сих пор в оценивании программ. В частности, большое влияние имело обследование управления школой с использованием метода внешнего наблюдения (Wolcott [97]), хотя это исследование и не было само по себе сосредоточено на оценивании

какой-либо программы. В качестве моделей использовались также методы полевой работы, применяемые социологами, придерживающимися качественного подхода (Nelson, Giannotta [68]).

В настоящее время целый ряд примеров этого подхода содержится в программе «Экспериментальные школы» (Nelson, Reynolds, French, Giannotta [69]). Существует также методологическая работа, связанная с практикой подобного рода (Everhart [38]). Через несколько лет мы освоим этот опыт и будем в состоянии более точно оценить этот подход. Уже сейчас я взялся бы утверждать, что эти материалы создают основу, которая поможет интерпретировать количественные оценки, стоит только этим последним появиться на свет. В них мы найдем богатое деталями повествование о том, как протекали события и что «было не так» и потому могло иногда привести к игнорированию количественного оценивания как не заслуживающего внимания, поскольку, в конечном счете, не наступило никаких ре-

---

<sup>1</sup> Ряд специальных методологических проблем антропологического полевого исследования и взаимной связи антропологии и психологии рассматривается: Campbell D. T. The Mutual Methodological Relevance of Anthropology and Psychology. — In: Psychological Anthropology. Ed. by F. K. L. Hsu. Homewood, Ill., The Dorsey press 1961; Levine R. A., Campbell D. T. Ethnocentrism. Theories of Conflict, Ethnic Attitudes and Group Behavior. N. Y., Wiley, 1972. — *Прим. ред.*

альных изменений, или из-за того, что одновременно вмешивалось такое множество более могущественных сил, что воздействие экспериментальной программы не имело никаких шансов проявиться. Но мне, тем не менее, кажется, что эти этнографические описания не могут быть приняты сами по себе в качестве свидетельств в пользу эффективности программы и заменить собой адекватную количественную экспериментальную оценку.

Антропологи, о которых идет речь, не имели никакого опыта изучения каких бы то ни было школьных систем. Они были привлечены уже после того (или в тот самый момент), как программа была запущена, и им неизбежно приходится изучать смесь старого и нового в условиях, где легко по ошибке приписать программе результаты, которые имели бы место в любом случае. Делу можно было бы помочь, если бы наши антропологи уделяли половину своего времени изучению какой-нибудь другой школы, во всем похожей на исследуемую, с тем лишь исключением, что в ней отсутствует новая экспериментальная программа. Это, очевидно, не было предусмотрено. Польза была бы и в том случае, если бы антрополог стал изучать данную школу за год или два до начала программы. (Тут трудно что-либо планировать, но мы могли бы рассматривать ныне ведущиеся этнографические описания школ как исследования, предшествующие будущим инновациям.)

Всякое знание является сравнительным, каким бы абсолютным оно ни казалось феноменально, и антрополог занимает обычно очень слабые позиции для осуществления валидных сравнений, поскольку его собственный исследовательский опыт и знания о школах, полученные им из вторых рук, предполагают столь различные подходы и точки зрения, что остаются лишь незначительные возможности для сравнений. Позиция непрофессиональных участников программы фактически более выгодна, чем позиция антропологов, как мы покажем ниже. В исследованиях в Миннеаполисе (Nelson et al. [68], [69]) несколько антропологов, работавшие с разными школами, имели преимущества ввиду возможности сравнивать свои записи.

Методы антропологического и социологического полевого наблюдения в начале их применения не были

сосредоточены на каузальном воздействии какого-то единичного фактора или нового института. Напротив, их цель заключалась в том, чтобы описывать вещи, как они есть, без обращения к причинам. В самом деле, методологическая ориентация часто предполагает отрицание релевантности или возможности установления каузальных связей. Принимая заказ на оценивание влияний и эффективности программы, антрополог или социолог утрачивает в значительной мере свою свободу и возвращается в более традиционную научную сферу каузального вывода. Уже приступив к работе, он осознает неопределенность этой задачи и в конце концов часто может действительно оправдывать свои усилия той ценностью, какую могут иметь для антропологии и педагогики подробные этнографические описания школ, оставив без внимания центральный вопрос о воздействии программы. Каким бы испытанным средством в достижении обычных для них целей ни являлись методы полевого наблюдения, они пока не прошли проверки в оценивании программ.

Исследования надежности или валидности полевого наблюдения, насколько я знаю, отсутствуют. Я склонен думать, что если бы двух этнографов послали исследовать одну и ту же культуру или двух социологов отправили бы в качестве включенных наблюдателей на одну и ту же фабрику, то их результаты в значительной мере совпадали бы. Но когда вы предлагаете такие исследования, то оказывается, что «качественного» социального исследователя больше интересуют различия, чем совпадения, и он не стал бы непременно ожидать сходных результатов, потому что за каждым наблюдателем признается уникальный угол зрения или потому, что считается, что не существует никакой социальной реальности, кроме той, которая сконструирована наблюдателем. С таким взглядом на многообразие результатов вряд ли сможет примириться какое-нибудь правительственное учреждение или добровольное объединение, пытающееся принять решение о расширении или завершении новой программы. Степень, в которой два независимо друг от друга работающих антрополога приходят к согласию относительно описания той или иной институциональной программы, в скором времени будет, по всей вероятности, подвергаться проверке. Что

264

касается меня, то я полагаю, что существует вполне достоверная социальная реальность, которая и должна быть описана, и поэтому будет обнаружено довольно значительное совпадение результатов. Но, разумеется, дадут о себе знать и индивидуальные особенности антропологов (Campbell [14]), и я не уверен, что согласие будет простирается до мало контрастных деталей, связанных с эффективностью программы.

Однако такая возможность повторения нуждается в демонстрации не только в случае качественных оценок. Количественные экспериментальные исследования предполагают такое множество решений в отношении способа осуществления программы, выбора мер и их формулировок, объединения данных в процессе анализа и т. п., что они тоже должны дублировать друг друга. Наши обширные программы оценивания должны осуществляться не менее чем в два приема и независимо друг от друга. Когда результаты совпадают, ясно, какое решение следует принять. Когда же они не совпадают, мы тем самым получаем предупреждение об ограниченной общности полученных данных. Если качественная и количественная оценки были бы осуществлены в отношении одних и тех же программ, я склонен ожидать, что они совпадут. Если же этого не произошло, нам, как мне кажется, следует признать возможным, что имела место ошибка в количественных расчетах. (Я думаю, что именно так обстояло дело при первом основном оценивании программ компенсаторного начального обучения, хотя, насколько мне известно, качественные оценки в этом случае никогда систематически не собирались и не суммировались (см. Campbell, Erlebacher [24].))

К качественной оценке программ следовало бы привлекать и историков. Специалисты в области политических наук уже привлечены к этой работе (см., например, Greenstone, Peterson [47]; впрочем, они избегают оценивать воздействие). Если качественной оценке программ суждено стать обычным делом, методологические основы гуманитарных исследований должны быть перестроены и пересмотрены. Самым распространенным методологическим комплексом приемов является «изучение отдельных случаев», объединяющее все доступные данные, поступающие от информантов, из периодики, документов, архивов и непосредственных на-

265

блюдений. Существуют такие стандарты регистрации данных, до которых доросли пока что не все исследования по оцениванию процессов и этнографические описания программ. В условиях полевой антропологической и социологической работы существует требование ежедневного ведения полевых записей, которые затем используются для проверки попыток обобщения. На факультете антропологии в Беркли Лови и Крёбер требовали обычно, чтобы все мифы и истории были записаны отдельно по меньшей мере от двух информантов. В отчете, где Гринстоун и Петерсон бойко перечисляют имена и события, восстанавливая незаписанную историю программы, они считают возможным утверждать следующее: «Во всех случаях отчеты о поведении участника основываются... на свидетельствах, полученных от нескольких действующих лиц, которые высказываются о нем с разных точек зрения» [47, с. 7]. Подобное требование способствует, конечно, повышению воспроизводимости исследования.

В прежних своих работах [12, 13, 19] я непочтительно отзывался о единичном случае, единичной совокупности условий, одномоментном исследовании случаев не из-за их качественной природы, а вследствие того, что такая ограниченность точек наблюдения сочеталась в них с избытком возможных каузальных концепций, что обманчивое совершенство их вида было почти очевидным. В недавней донкихотской и двойственной статье «„Степени свободы“ и изучение отдельных случаев»<sup>1</sup> я покаяться в своих заблуждениях, напомнив себе, что подобные исследования регулярно вступают в противоречие с первоначальными ожиданиями их авторов и для таких скептиков, как я, являются убедительными и информативными до такой степени, которую мой прямолинейный негативизм не допускает. Мое предположительное решение состоит в том, что в пространстве многосмысловых атрибутов должно существовать множество «степеней свободы», что придает подобному исследованию силу тестируемого и проверяемого, чего я раньше не допускал. Возможно, что результаты конкретных исследований случаев при оценке программ смогут, в конце концов, убедить меня в том, что такие исследо-

---

<sup>1</sup> См. перевод этой статьи в данной книге (с. 278).

266

вания в состоянии оценивать программы. В частности, они, возможно, могут валидно обнаруживать неожиданные последствия, упущенные при более структурированных подходах. Я жду, однако, конкретных примеров. (Другие соображения по поводу антропологического метода, затрагивающие иногда квантификацию, но, как правило, с уважением относящиеся к качественному подходу, изложены в работах: Campbell [10, 12, 14, 21]; Campbell, LeVine [25]; Werner, Campbell [95].)

Все увеличивающееся со временем знакомство с событиями и людьми дает даже стороннику количественного подхода качественное знание, которое помогает ему обнаруживать в своих данных неверные истолкования, ошибки и подтасовки. В случае, когда один и тот же человек запускает программу воздействий, собирает данные и

анализирует результаты, качественные наблюдения могут в значительной мере подкреплять количественные результаты. В современном крупномасштабном количественном оценивании программ эта опора на знакомство с материалом может быть утрачена — в значительной мере или полностью. В худшем случае сбор данных перед началом воздействия, руководство программой, сбор итоговых данных и анализ результатов производят четыре разные группы по четырем разным контрактам. Разделение труда в рамках проекта вызывает подобное разъединение даже в том случае, если исследования ведутся по одному контракту. Убеждение в том, что внешние оценщики более объективны, служит еще одним препятствием для заинтересованности в качественном опыте. Несомненно, что цифры, выдаваемые вычислительной машиной, часто создают псевдонаучный фасад, для которого использованы принятые на веру количественные данные, не имеющие на деле валидного качественного основания. Этот пробел могли бы заполнить занятые в работе над проектом антрополог, социолог или историк, если бы перед ними была поставлена задача обыденного ознакомления с общим контекстом, включающим социальные взаимодействия, которые приводят к таким количественным результатам.

### Оценивание программы ее участниками

Участники, как мы уже отмечали, будут, как правило, занимать более благоприятные для наблюдения по-

267

зиции, чем антропологи или другие внешние наблюдатели новой программы. Они будут воспринимать условия, предшествующие запуску программы, с той же точки зрения, с какой они подходят к данной программе. Поэтому их впечатления от программы будут более релевантными, прямыми, обоснованными и непосредственными. Вместе взятые, они позволят благодаря своей большей многочисленности усреднить индивидуальные особенности наблюдателей, которые могут преобладать в отчете любого отдельного этнографа.

Хотя в процессе оценивания программы к участникам обращаются за изрядным количеством сведений, им редко прямо предлагают оценить программу, высказать суждение о ее адекватности, дать совет по поводу ее продолжения, прекращения, распространения или модификации. Участников просят дать не оценку программ, а, скорее, сведения о них самих и их собственной адекватности. Мы теряем, таким образом, множество хорошо обоснованных мнений. Оценки учебных курсов составляют здесь единственное исключение, и их лучшие образцы подтверждают компетентность участников как наблюдателей (см. Frey [41]).

Если говорить о программе компенсаторного начального обучения, то многие матери располагают опытом относительно своих старших детей, которых они могут сравнивать с младшими, охваченными программой. Участники эксперимента «Негативный подоходный налог» в Нью-Джерси<sup>1</sup> имеют за плечами три года жизни в условиях весьма специфической системы денежных субсидий по доходу. Они располагают в качестве основания для сравнения, как своим собственным прошлым опытом, так и одновременными впечатлениями своих столь же неимущих соседей, не охваченных экспериментом. Они знают, что участвуют в эксперименте, и должны иметь множество впечатлений по поводу влияний, оказываемых этим экспериментом на их жизнь.

Нам следовало бы попробовать собрать суждения этих участников. В согласии с духом качественного обобщения качественных суждений нам нужно было бы,

<sup>1</sup> Подробнее об этой программе см. в статье «Измерение воздействий, оказываемых запланированным социальным изменением» (с. 345). — *Прим. ред.*

вероятно, дать возможность участникам суммировать свои суждения непосредственно в личном общении, вначале, по-видимому, в малых группах, имеющих в своем составе протоколиста, а затем на оживленных совещаниях групп уполномоченных представителей. Собrania групп могли бы предлагать решения и ставить их на голосование, однако обычным итогом были бы словесные формулировки, отражающие некватифицированное общее согласие.

Следует ли контрольным группам участников встречаться с экспериментальными, чтобы дать дополнительное основание для сравнения? Если мы хотим использовать такие методы оценки, разве в этом случае не желательно иметь одновременно смежные альтернативные программы, с тем чтобы их участники на протяжении всего исследования имели возможность в процессе общей беседы сравнивать свои впечатления об этих альтернативах? Во всяком случае, собрания участников различных альтернативных программ для общего обмена мнениями можно, по-видимому, считать плодотворным способом получения компетентных качественных сравнений.

В программах компенсаторного начального обучения имеются еще и другие компетентные наблюдатели, чьи оценки остаются незарегистрированными, например учителя, работающие по этим программам, учителя первых классов, сами дети, охваченные программой. В экспериментах с «негативным подходным налогом» на роль компетентных качественных наблюдателей, вполне достойных того, чтобы их выслушали, претендуют и постоянные социальные работники данной общины, и интервьюеры, и соседи по дому.

Роберт Вулф [98] предложил пользоваться для выработки единого решения процедурами судопроизводства — слушанием сторон с привлечением свидетелей, таких, как причастные к программе родители, учителя и воспитанники, с адвокатами, ведущими перекрестный допрос, и состоящим из экспертов «судом». Он приводит один весьма убедительный случай из судебной практики, свидетельствующий о высокой валидности, обеспечиваемой процедурами судопроизводства, особенно перекрестным допросом, ссылаясь среди прочих и на Левайна [60].

Однако процедуры обобщения качественных данных не представляются мне чем-то особенно существенным. Групповое мнение, если существует полное согласие, — это упрямый факт, который может быть обнаружен при помощи целого ряда методов. Я думаю, что основные важные данные об оценках программы ее участниками могут быть получены с помощью интервью, состоящего из прямых вопросов об эффекте, который имела программа, и т. п. Ответы на эти вопросы могли бы быть в дальнейшем статистически обобщены. Правда, при этом иногда могут быть упущены из виду некоторые тонкости. Участники могли бы обнаружить их в ходе групповой дискуссии и прийти по их поводу к единому мнению. Столь же часто, однако, во время групповой дискуссии возможно доминирование какого-то одного участника, и тем самым будет утрачено преимущество усреднения, достигаемое при статистическом обобщении.

Для институтов и групп, которые будут служить постоянными полигонами изменений, вводимых с помощью программ, таких, как, например, система социального обеспечения или школы, Гордон и я [46] предложили, чтобы все группы участников (социальные работники и их клиенты, воспитанники, родители и учителя) заполняли «годовой отчет по оцениванию программы», который мог бы играть роль предварительного тестирования и построения временных рядов для квазиэкспериментальных оценок изменений.

Мы нуждаемся в опробовании целого ряда таких процедур. Как указано выше, они представляли бы собой методологически независимые перекрестные валидации

количественных результатов. С их помощью можно было бы выявить такие последствия применения программы, которые не предполагались при обследовании программы с помощью формальных измерительных средств. Они, по всей вероятности, подтвердили бы основные результаты, полученные по предусмотренным параметрам. В противном случае, как я уже говорил, нам следовало бы принять во внимание возможную ошибочность количественных процедур. Если я допускаю все это, то почему мне все же не хотелось бы, чтобы качественные процедуры применялись без количественных? Потому, что я склонен признавать за количест-

270

венными методами, когда они основываются на твердом и проверенном качественном знании, возможность выйти за пределы, допускаемые качественным подходом, и выявить тонкости, которые могли бы ускользнуть от качественного подхода.

В данный вопрос можно внести большую ясность, выяснив, какую еще полезную роль могли бы играть участники или наблюдатели из числа сторонников качественного подхода в социальных науках, то есть как критики количественного подхода.

Если бы этнографам или историкам, занятым в программе, а также участникам дискуссионных групп были предъявлены основные результаты и было разрешено их критиковать, мы получили бы ценный способ валидизации. С некоторыми результатами они были бы согласны. С другими не согласились бы и объясняли бы их как артефакты, обязанные своим происхождением тем аспектам ситуации, которые упущены из виду при количественных измерениях. Может оказаться, что некоторые результаты вызовут у них удивление, но, если они выдержат повторную проверку, они будут признаны валидными и этими наблюдателями. В этих последних случаях количественное знание действительно вышло бы за границы качественного знания.

Прежде чем будет завершен этот анализ, предстоит еще извлечь пользу из опыта наших многолетних исследований суждений человека и их погрешностей. Следует пересмотреть перечни опасностей, угрожающих валидности экспериментов и измерений, сравнивая качественный и количественный подходы по степени подверженности каждой из них. Следует рассмотреть особые опасности, связанные с политикой оценивания — политикой в рамках проекта и политикой, вызванной внешними условиями. Каким образом эти давления влияют на количественное знание, с одной стороны, и качественное — с другой?

\* \* \*

Подведем итоги. Проблема полярности количественного и качественного подходов к исследованию социального действия остается нерешенной, если считать ее

271

решением преимущественное утверждение одного по сравнению с другим. Социальное познание — это в еще большей мере, чем познание физическое, рискованный и гипотетический процесс. При любом подходе всегда остается место для обоснованной критики. Каждый полус сильнее всего в своей критике другого, но не в непогрешимости своих собственных притязаний на дескриптивное знание.

Постоянно выступая сам как сторонник количественного экспериментального подхода к исследованию действия, я не могу рекомендовать ни ориентированную на качественный подход социальную науку, ни групповой процесс выработки группового согласия в качестве замещения количественного подхода. Но я настоятельно рекомендую оба подхода как необходимые, взаимно дополняющие средства перекрестной валидизации.

В еще большей мере я стремился к тому, чтобы напомнить моим коллегам — сторонникам количественного подхода, что в успешных лабораторных науках квантификация строится на всепроникающих качественных знаниях ученого и перепроверяется ими. Условия массового производства, характерные для количественной социальной науки в области оценивания программ, таковы, что это качественное основание может быть в значительной мере утрачено. Если мы хотим оставаться в рамках истинной науки, мы должны восстановить в исследованиях действия эту качественную основу количественного знания.

## ЛИТЕРАТУРА

1. Baker F. The changing hospital organizational system. A model for evaluation. Read in General Systems Session of AAAS Annual Meeting. Dallas, Texas, 1968.
2. Baker F. General systems theory, research and medical care. — In: Sheldon A., Baker F., McLaughlin C. P. Systems and Medical Care. Cambridge, The MIT Press, 1970, 1—26.
3. Beck B. Cooking the welfare stew. — In: R. W. Habenstein (ed.). Pathways to data: Field methods for studying ongoing social organizations. Chicago, Aldine, 1970.
4. Becker H. M., Geer B., Hughes E. C. Making the grade. N. Y., Wiley, 1968.
5. Becker H. W. Sociological work: Method and substance. Chicago, Aldine, 1970.
6. Bennis W. G. The case study. *J. of Appl. Behav. Science*, 1968, 4, 2, 227—231.

272

7. Blachowicz J. A. Systems theory and evolutionary models of the development of science. *Philosophy of Science*, 1971, 38, 178—199.
8. Blaau P. M. The dynamics of bureaucracy. Chicago, Univ. of Chicago Press, 1963.
9. Campbell D. T. The informant in quantitative research. *Amer. J. of Soc.*, 1955, 60, 339—342.
10. Campbell D. T. Methodological suggestions from a comparative psychology of knowledge processes. *Inquiry*, 1959, 152—182.
11. Campbell D. T. Blind variation and selective retention in creative thought as in other knowledge processes. *Psychol Rev.*, 1960, 67, 380—400.
12. Campbell D. T. The mutual methodological relevance of anthropology and psychology. — In: F. L. K. Hsu (ed.). Psychological anthropology: Approaches to culture and personality. Homewood, Ill. Dorsey, 1961, 333—352.
13. Campbell D. T. Social attitudes and other acquired behavioral dispositions. — In: S. Koch (ed.). Psychology: A study of a science. Vol. VI. N. Y., McGraw-Hill, 1963, 94—172.
14. Campbell D. T. Distinguishing differences of perception from failures of communication in cross-cultural studies. — In: F. S. C. Northrop, H. H. Livingston (eds.). Cross-cultural understanding: Epistemology in anthropology. N. Y., Harper & Row, 1964, 308—336.
15. Campbell D. T. Variation and selective retention in socio-cultural evolution. — In: Barringer H. R., Blanksten G. I., Mack R. W. (eds.). Social change in developing areas. Cambridge, Mass., Schenkman, 1965.
16. Campbell D. T. Pattern matching as an essential in distal knowing. — In: K. R. Hammond (ed.). The psychology of Egon Brunswik. N. Y., Holt, Rinehart & Winston, 1966, 81 — 106.
17. Campbell D. T. Reforms as experiments. *Amer. Psychologist*, 1969a, 24, 409—429.
18. Campbell D. T. A phenomenology of the other one: Corrigible, hypothetical and critical. — In: T. Mischel (ed.). Human action: Conceptual and empirical issues. N. Y., Academic Press, 1969b, 41—69.
19. Campbell D. T. Natural selection as an epistemological model. — In: Naroll R., Cohen R. (eds.). A handbook of method in cultural anthropology. Garden City, N. Y., Natural History Press, 1970, 51—85.
20. Campbell D. T. Methods for the experimenting society. Paper presented to the meeting of the American Psychological Association, Washington, D. C., Sept. 1971.

21. Campbell D. T. Herskovits, cultural relativism, and meta-science. — In: M. J. H e r s k o v i t s. Cultural relativism. N. Y., Random House, 1972, V—XXIII. (Introduction).
22. C a m p b e l l D. T. Evolutionary epistemology. — In: P. A. S c h i l p p (ed.). The philosophy of Karl Popper. Vol. 14, I & II. The library of living philosophers. La Salle, Ill.: Open Court Publishing, 1974, 413—463, Vol. 14-1.
23. C a m p b e l l D. T. «Degrees of Freedom» and the case study *Comparative Political Studies*, Sept., 1975, 178—193. (См. перевод в данной книге.)

273

24. C a m p b e l l D. T., E r l e b a c h e r A. E. How regression artifacts in quasi-experimental evaluations can mistakenly make compensatory education look harmful. — In: J. H e l l m u t h (ed.). Compensatory education: A national debate. Vol. III. N. Y., Brunner/Mazel, 1970, 185—210.
25. C a m p b e l l D. T., L e V i n e R. A. Field-manual anthropology. — In: R. N a r o l l, R. C o h e n (eds.). A handbook of method in cultural anthropology. Garden City, N. Y., Natural History Press, 1970, 366—387.
26. C a m p b e l l D. T., S t a n l e y J. C. Experimental and quasi-experimental designs for research. Chicago, Rand McNally, 1966. (См. перевод в данной книге.)
27. C a m p b e l l R. The chasm: The life and death of a great experiment in ghetto education. Boston, Houghton Mifflin Co., 1974.
28. C a w s P. The structure of discovery. *Science*, 1969, 166, 1375—1380.
29. C h e i n I. On evaluating self-surveys. *J. of Soc. Iss.*, 1949, 5, 56-63.
30. C h e i n I. Community self-surveys. Lecture given at University of Rochester. Nov. 5, 1956, 23.
31. C h e i n I., C o o k S. W., Harding J. The field of action research. *The Amer. Psychologist*, 1948a, 3, 43—50.
32. C h e i n I., C o o k S. W., Harding J. The use of research in social therapy. *Human Relations*, 1948b, 1, 497—510.
33. C o o k S. W. The systematic analysis of socially significant events: A strategy for social research. *The J. of Soc. Iss.*, 1962, 18, 66—84.
34. C o o k T. D., C a m p b e l l D. T. The design and conduct of quasi-experiments and true experiments in field settings. — In: M. D. D u n n e t t e (ed.). Handbook of industrial and organizational psychology. Chicago, Rand McNally, 1976.
35. D o u g l a s J. D. The social meanings of suicide. Princeton, Princeton Univ. Press, 1967.
36. D u n c k e r K. Uber induzierte Bewegung. Ein Beitrag zur Theorie optischwahrgenommener Bewegung. *Psychologische For-schung*, 1929, 12(6), 180—259.
37. E t z i o n i A., L e h m a n E. W. Some dangers in «valid» social measurement. *Annals of the Amer. Acad. of Polit. and Soc. Science*, 1967, 373, 1—15.
38. E v e r h a r t R. B. Problems of doing fieldwork in educational evaluation. *Human Organization*, 1975, 34, 205—215.
39. F e y e r a b e n d P. K. Against method: Outline of an anarchistic theory of knowledge. — In: R a d n e r M., W i n o k u r S. Analyses of theories and methods of physics and psychology. Vol. IV, Minnesota Studies in the Philosophy of Science. Minneapolis, Univ. of Minnesota Press, 1970, 17—130.
40. F r e e m a n H. E., S h e r w o o d C. C. Social policy and social research. Englewood Cliffs, Prentice Hall, 1970.
41. F r e y P. W. Student ratings of teaching: validity of several rating factors. *Science*, 1973, 182, 83—85.
42. G a r f i n k e l H. «Good» organizational reasons for «bad» clinic records. — In: Garfinkel H. Studies in ethnomethodology. Englewood Cliffs, Prentice Hall, 1967, 186—207.

274

43. G i l b e r t N., S p e c h t H. The model cities program: A comparative analysis of participating cities process, product, performance, and prediction. U. S. Government Printing Office, Washington, 1973.
44. G l a s e r B. G., S t r a u s s A. L. The discovery of grounded theory: Strategies for qualitative research. Chicago Aldine, 1967.

45. Goodman N. Fact, fiction, and forecast. Cambridge, Harvard Univ. Press, 1955.
46. Gordon A. C., Campbell D. T., et al. Recommended accounting procedures for the evaluation of improvements in the delivery of state social services. Duplicated manuscript, 1971, Northwestern University.
47. Greenstone J. D., Peterson P. E. Race and authority in urban politics: Community participation and the war on poverty. N. Y., Russell Sage Foundation, 1973.
48. Guttentag M. Models and methods in evaluation research. *J. for the Theory of Soc. Beh.*, 1971, 1, 75—95.
49. Guttentag M. Evaluation of social intervention programs. *Annals of the N. Y. Academy of Sciences*, 1973, 218, 3—13.
50. Hanson N. R. Patterns of discovery. Cambridge, Univ. Press, 1958.
51. Harding J. Community self-surveys: A form of combating discrimination. — Congress Weekly: A Review of Jewish Interests. 1948, March 5.
52. Jastrow J. Fact and fable in psychology. Boston, Houghton Mifflin, 1900.
53. Jones R. W. Principles of biological regulation: An introduction to feedback systems. N. Y., Academic Press, 1973.
54. Kaplan M., et al. The model cities program: A comparative analysis of city response patterns and their relation to future urban policy. U. S. Government Printing Office, Washington, D. C., 1973.
55. Kitsuse J. K., Cicourel A. V. A note on the uses of official statistics. *Social Problems*, 1963, 11, 131 — 139.
56. Kordig C. R. The justification of scientific change. N. Y., Humanities Press, 1971.
57. Кун Т. Структура научных революций. М., 1977.
58. Kuhn T. S. Logic of discovery or psychology of research? & Reflections on my critics. — In: I. Lakatos, A. Musgrave (eds.). Criticism and the growth of knowledge. Cambridge, Cambridge Univ. Press, 1970, p. 1—23, 231—278.
59. Lakatos I. Falsification and the methodology of scientific research programmes. — In: Lakatos I., Musgrave A. (eds.), Criticism and the growth of knowledge. Cambridge Univ. Press, 1970.
60. Levine M. Scientific method and the adversary model: Some preliminary thoughts. *Amer. Psychologist*, 1974, 29 (9), 661—677.
61. Lewin K. Action research and minority problems. *J. of Soc. Issues*, 1946, 2, 34—46. Reprinted in: Lewin K. Resolving social conflicts, N. Y., Harper, 1948.
62. Lewin K. Frontiers in group dynamics, Part II—B. *Human Relations*, 1947, 1, 147—153; reprinted as «Feedback problems in social diagnosis and action». — In: Buckley W. (ed.). Modern Systems Research for the Behavioral Scientist. Chicago, Aldine, 1968, 441—444.
63. Lippitt R., Radke M. New trends in the investigation of prejudice. *The Annals of the Amer. Acad. of Polit. and Soc. Science*, 1946, 244, March.

64. Marrow A. J. The practical theorist: The life and work of Kurt Lewin. N. Y., Basic Books, Inc., 1969.
65. Morrissey W. R. Nixon anti-crime plan undermines crime statistics. *Justice Magazine*, 1972, 1, 8—11, 14.
66. Murchison C. (ed.) Handbook of Child Psychology. Worcester, Mass., Clark Univ. Press, 1931.
67. Nelson H., Giannolla F. J. Research methodology in alternative education settings: The met plan. Minneapolis, Minnesota, Aries Corporation, 1974a.
68. Nelson H., Giannotta F. J. (Draft) MET interim report on southeast alternatives ESP project, vol. 1, Minneapolis, Aries Corp., 1974b.
69. Nelson H., Reynolds J., French L. R., Giannotta F. J. (Draft) MET interim report on southeast alternatives ESP project, vol. IIA; vol. IIB, Minneapolis, Aries Corp., 1974.
70. Poincaré H. Mathematical creation. — In: Poincaré H. The foundations of science. N. Y., Science Press, 1913.
71. Polanyi M. Personal knowledge. London, Routledge & Kegan Paul, 1958.
72. Popper K. R. The logic of scientific discovery. N. Y., Basic Books, 1959.
73. Popper K. R. Conjectures and refutations. N. Y., Basic Books, 1963.

74. Popper K. R. Objective knowledge: An evolutionary approach. Oxford, Clarendon, 1972.
75. Quine W. V. From a logical point of view. Cambridge, Harvard Univ. Press, 1953.
76. Quine W. V. Ontological relativity. N. Y., Columbia Univ. Press, 1969.
77. Raser J. R., Campbell D. T., Chadwick R. W. Gaming and simulation for developing theory relevant to international relations. — In: A. Rappoport (ed.). General Systems: Yearbook of the Society for General Systems Research, Ann Arbor, Mich. 1970, 15, 183—204.
78. Ridgeway V. Dysfunctional consequences of performance measures. *Administrative Science Quarterly*, 1956, 1, 240—247.
79. Salasin S. Experimentation revisited: A conversation with Donald T. Campbell. *Evaluation*. 1973, 1, 7—13.
80. Schulberg H. C., Baker F. Program evaluation models and the implementation of research findings. *Amer. J. of Public Health*, 1968, 58, 1248—1255 (reprinted in: Schulberg, Sheldon, Baker, 1969, p. 562—572).
81. Schulberg E. H., Sheldon A., Baker F. Introduction. — In: Program Evaluation in the health fields. N. Y., Behavioral Publications, 1969, p. 3—28.
82. Segall M. H., Campbell D. T., Herskovits M. J. The influence of culture on visual perception. Indianapolis, Ind., Bobbs-Merrill, 1966.
83. Sellitz C. The use of survey methods in a citizens campaign against discrimination. *Human Organization*, 1956, 14, 19—25.
84. Sellitz C., Cook S. W. Can research in social science be both socially useful and scientifically meaningful? *Amer. Soc. Rev.*, 1948, 13, 454—459.

276

85. Seidman D., Couzens M. Crime statistics and the great American anticrime crusade: Police misreporting of crime and political pressures. Paper presented at the meeting of the American Political Science Association, Washington, C. D., Sept. 1972.
86. Sheldon A., Baker F., McLaughlin C. P. (eds.). Systems and Medical Care. Cambridge, The MIT Press, 1970, 1—26.
87. Shimony A. Scientific inference. — In: R. Colodny (ed.). Pittsburgh studies in the philosophy of science. Vol. IV. Pittsburgh, Univ. of Pittsburgh Press, 1970, 79—172.
88. Stewart V. M. Tests of the «carpentered world» hypothesis by race and environment in America and Zambia. *Intern. J. of Psychol.*, 1973, 8, 83—94.
89. Tolman E. C. Purposive behavior in animals and men. N. Y., The Century Co., 1932.
90. Toulmin S. E. Foresight and understanding: An inquiry into the aims of science. Bloomington. Indiana Univ. Press, 1961.
91. Toulmin S. E. Human understanding. Vol. I. The evolution of collective understanding. Princeton, Princeton Univ. Press, 1972.
92. Twigg R. Downgrading of crimes verified in Baltimore. *Justice Magazine*, 1972, 1, 15, 18.
93. Vidich A., Bensman J. The validity of field data. *Human Organization*, 1954, 13 (1), 20—27.
94. Weiss R. S., Rein M. The evaluation of broad-aim programs: Experimental design, its difficulties, and an alternative. *Admin. Science Quart.*, 1970, 15, 97—109.
95. Werner O., Campbell D. T. Translating, working through interpreters, and the problem of decentering. — In: R. Naroll, R. Cohen (eds.). A handbook of method in cultural anthropology. Garden City, N. Y., Natural History Press, 1970, 398—420.
96. Wittgenstein L. Philosophical investigations. N. Y., Macmillan, 1953.
97. Wolcott H. F. The man in the principal's office: An ethnography. N. Y., Holt, Rinehart and Winston, 1973.
98. Wolf R. L. The application of select legal concepts to educational evaluation. Ph. D. Diss. Univ. of Illinois at Urbana-Champaign, 1974.

277

**«СТЕПЕНИ СВОБОДЫ»  
И ИЗУЧЕНИЕ ОТДЕЛЬНЫХ СЛУЧАЕВ<sup>1</sup>**

Преобладающим способом исследования в антропологии, сравнительной политической науке и сравнительной социологии остается интенсивное изучение отдельной иноземной культуры человеком извне, для которого она является единственной глубоко познанной чужеродной культурой. Подобные исследования могут быть выполнены специалистами в области социальных наук или «любителями» (такими, как миссионеры, дипломаты, журналисты, бизнесмены, легионеры или туристы). Наблюдения и досуг могут побудить их записать свои впечатления о чужеродной для них культуре. Даже если эти наблюдатели-любители не ведут записей, они оказывают сильное влияние на аккумуляцию социального исследователя в чужой культуре или на его приспособление к этой культуре как экспатрианта (Kidder [18]). Другим сходным жанром являются описания своей собственной страны, сделанные во время или после длительного пребывания в какой-нибудь другой стране. Так, например, Кениата в 1938 г. описал Кикуйю<sup>2</sup>, находясь в Англии в качестве ученика Малиновского. Подобное знание, записанное или незаписанное, я буду обозначать как «обыденное знание» в сравнительных социальных науках. Если мы и добьемся какой-нибудь значимой количественной корреляции для 100 наций, то только благодаря тому, что на каждом шагу будем опираться на такого рода знания, но не потому, что нам удастся заменить его

---

<sup>1</sup> Campbell D. T. «Degrees of Freedom» and the Case Study. — «Comparative Political Studies», 1975, vol. 8, № 2, p. 178—193. © 1975 Sage Publications, Inc.

<sup>2</sup> Кикуйю — народ, живущий в центральной части Кении.

какой-то «научной» количественной методологией. Количественное обобщение на материале многих наций в деталях будет вступать в противоречие с подобным натуралистическим наблюдением отдельных случаев, но только потому, что мы примем на веру гораздо более значительное число таких же наблюдений (Campbell [5]).

Нельзя сказать, что такое обыденное натуралистическое наблюдение является объективным, надежным или беспристрастным. Но это все, чем мы располагаем. Это единственный путь к знаниям, сколько бы ни ждало нас на нем помех, оплошностей и пристрастных суждений. Мы должны отдавать себе отчет в его слабостях, но в то же время и доверять ему, если мы собираемся вообще оставаться в рамках сравнительной (или монокультурной) социальной науки. Я вернусь к упомянутым пристрастным суждениям позднее, а сначала хотел бы попытаться внести коррективы в некоторые из моих собственных недостаточно сдержанных замечаний в адрес метода исследования отдельных случаев. Карикатурное изображение этого подхода — изучения отдельных случаев (то, что я ранее имел в виду) — включает наблюдателя, который отмечает какую-то одну поразительную особенность той или иной культуры, после чего приступает к перебору всех прочих различий по всем другим переменным в поисках какого-либо объяснения. К его услугам почти все каузальные понятия его родного языка, которые он использует. То, что он найдет «объяснение», которое кажется вполне удовлетворительным, становится неизбежным, поскольку им полностью утрачены «степени свободы». (Как если бы он пытался совместить две точки наблюдения посредством формулы, содержащей тысячу подходящих обозначений, тогда как в хорошей науке следует иметь в нашей формуле меньше обозначений, чем наши данные содержат точек.)

Эта ориентация получила отражение в работе Кэмпбелла и Стэнли [10]<sup>1</sup>, после чего многие рассматривали межкультурное сравнение как слабую форму квази-экспериментального плана (см., например, Lijphart [23, с. 683—685]; Boesch, Eckensberger [3]).

<sup>1</sup> См. перевод этой работы в данной книге (с. 34—191).

Мое сугубо отрицательное отношение к изучению отдельных случаев — или же, на самом деле, к сравнению двух ареалов — нашло свое выражение и в написанной мною главе в книге, вышедшей под редакцией Шу (см. Campbell [8]):

*«Неинтерпретируемость сравнений двух, и только двух, естественных случаев. Принимая во внимание важность критики Малиновским<sup>1</sup> интерпретации Эдипова конфликта в терминах любви — ревности, непростительно, что его наблюдения больше не повторялись. Как бы основательны ни были его полевые исследования других проблем, опубликованные им данные по этому вопросу на самом деле весьма скудны. Он ссылается на данные, связанные с явным содержанием сновидений... тогда как мы нуждаемся в значительных выборках детально записанных сновидений мальчиков, девочек, мужчин и женщин.*

Но, несмотря на острую потребность в верификации и обобщении данных Малиновского относительно внутрисемейных установок в изучавшемся им племени на острове Тробриан, ценность такого повторения для проверки гипотезы Фрейда невелика. Нам, заинтересованным в использовании такого рода данных скорее для изображения процесса, чем для исчерпывающего описания отдельных случаев, нужно принять следующее правило:

*Никакое сравнение отдельно взятой пары естественных объектов не может быть интерпретировано.* Имеется множество параметров, по которым Тробриан и Вена различаются между собой и которые могли бы лечь в основу потенциальных конкурентных объяснений, и мы никоим образом не можем исключить их. При сравнении подобной пары требование *ceteris paribus*<sup>2</sup> неприемлемо. Но нет нужды отказываться от дальнейшего сбора данных. Как авантюлярная<sup>3</sup>, так и европейская традиция столь широко распространены по

---

<sup>1</sup> Автор имеет в виду работу известного культурантрополога Б. Малиновского: Malinovsky B. *Sex and Repression in Savage Society*. London, Humanities Press, 1927. — *Прим. ред.*

<sup>2</sup> При прочих равных условиях (*лат.*).

<sup>3</sup> То есть традиция, основанная на соблюдении семейно-родового правила, согласно которому мальчик воспитывается дядей, а не отцом. — *Прим. перев.*

всему миру, что, если бы нашей целью была проверка теорий Эдипова комплекса, мы могли бы набрать для сравнения дюжину пар родо-племенных общин из самых разных культурных ареалов, различающихся между собой по одному признаку, а именно по тому, кто из мужчин воспитывает и наставляет мальчика, но сходных, насколько возможно, в иных отношениях. Предположим, что сновидения, рассказанные мальчиками, продемонстрировали ожидавшиеся различия в каждой паре; в таком случае, чем больше оказалось у нас таких пар, тем меньше осталось бы приемлемых конкурентных гипотез и, таким образом, тем надежнее было бы наше подтверждение» (Campbell [8, с. 344—345]).

И далее: «Рассмотренная в этом свете, проблема «лаборатории» актуальных международных отношений состоит в том, что мы имеем здесь слишком малое число действующих лиц, слишком мало событий, слишком мало «повторных демонстраций», в которых поддающиеся классификации как сопоставимые группы действующих лиц вступают во взаимодействие с сравнимых исходных позиций, и еще слишком много релевантных теоретических соображений. Число степеней свободы для последующих суждений, для проверки теории на материале данных, которые к ней привели, гораздо меньше, чем для *предсказаний* (это обстоятельство получило известное отражение в различении одностороннего и двустороннего критерия значимости). Для очерков, написанных с позиций качественного подхода знающим политологом или историком, эти

проблемы столь же реальны и чреватые опасностями, как и для специалиста по мультивариативной статистике. Быть может, довести эту проблему до сознания последнего легче (хотя, впрочем, часто довольно трудно, когда он утверждает, что описывает всю генеральную совокупность), потому что заботится об экспликации переменных, тогда как ученый-гуманитарий располагает неэксплицированным, но очень большим числом возможных «соображений», которые он связывает — или мог бы связать — с тем или иным случаем. (Это скорее, чем признаки, используемые при распознавании образа, составляет суть проблем, связанных с «понимающим» подходом, один из источников его удовлетворительности

281

в конкретных случаях и его неудовлетворительности как процесса испытания реальности» (Raser et al. [29, с. 186-187]).

Хотя вполне вероятно, что многие исследования отдельных случаев, дающие или предполагающие какие-то интерпретации и объяснения или соотносящие эти случаи с теорией, обладают этими недостатками, теперь мне, пожалуй, ясно, что не все они таковы и не обязательно должны быть таковыми и что я недооценил один из главных источников дисциплинированности (иначе говоря, степеней свободы, если я настаиваю на использовании этого статистического понятия для аналогичной проблемы в нестатистических ситуациях). В исследовании отдельных случаев, проведенном «бдительным» социальным исследователем, хорошо знакомым с местными условиями, теория, которую он использует для объяснения основного различия, порождает также предсказания и ожидания относительно десятков других аспектов данной культуры, и он не принимает теорию до тех пор, пока большинство этих последних тоже не получит подтверждения. Он в каком-то смысле проверяет теорию с помощью степеней свободы, образуемых многочисленными импликациями, характерными для любой единичной теории. Этот процесс представляет собой один из видов сопоставления образов (pattern matching) (Campbell [6], Raser [28]), при котором рассматриваются многочисленные аспекты данного образа, требуемые теорией, чтобы совместить их с наблюдениями социального исследователя относительно местных условий.

Опыт социальных исследователей подтверждает это. Даже в качественном исследовании отдельных случаев добросовестный социальный исследователь часто не находит объяснения, которое представлялось бы удовлетворительным. Подобный результат был бы невозможен, если бы карикатурное изображение исследования отдельных случаев, каким оно выглядит в приведенных цитатах, было бы верным: вместо этого мы имели бы избыток субъективно приемлемых объяснений. Хотя у меня и нет сомнений, что у людей существует статистически значимая склонность к заключениям в отсутствие определенности, когда налицо, по существу, одни только случайные данные (Campbell [9]) (у животных

282

нечто аналогичное было продемонстрировано Толменом и Кречевским [34], а также Б. Ф. Скиннером в его исследовании предрасположенности у голубей), эта склонность не может быть доминирующей, поскольку как биологическая, так и социальная эволюция ограничили бы такое легкое верие в пользу мутантов с более тонкой различительной способностью. Иначе говоря, механизмы нашего обыденного знания должны были в конечном счете обладать адаптивной ценностью по крайней мере в той экологической нише, в которой они складывались.

Беккер (личное сообщение, см. также Becker [2, с. 39—62, 25—38]) уверяет, что социальный исследователь, предпринимающий интенсивное изучение отдельных случаев

путем включенного наблюдения или, используя для знакомства с ситуацией какие-либо другие качественные подходы, связанные с обыденным знанием, в конце концов, обнаруживает ошибочность своих первоначальных представлений и теорий. Если это так, то это важный факт, и он заслуживает систематического подтверждения. Если это так, то это означает, что интенсивное сравнительно-культурное исследование отдельных случаев подчиняется известной дисциплине и обладает способностью отвергать теорию, что отрицалось в моем карикатурном изображении этого метода.

Нейролл [26], один из главных сторонников квантификации в области антропологии, располагает мощным средством контроля за качеством количественных данных, определяя «качество» этнографа. Примечательно, что критерии качества связаны не с использованием этнографом тех или иных специальных средств количественных социальных наук (таких, как процедуры случайных выборок, планы и наборы структурированных интервью, психологические тесты и т. д.), а, скорее, с высоким уровнем качественного знакомства с описываемой культурой, определяемым, например, длительностью проживания в данном ареале и знанием местного языка.

Проиллюстрировать принцип дисциплинированности и степеней свободы, связанных с многочисленными теоретическими импликациями, возможно, поможет наложение некоторых непосредственных личных наблю-

283

дений, сделанных на ранних стадиях одного знаменитого исследования отдельных случаев. Осенью 1940 г. я был в Беркли участником семинара «Культура и личность». Этим семинаром руководили А. Л. Крёбер и Э. Г. Эриксон, и он был посвящен изучению индейцев племени юрок в Северной Калифорнии, среди которых Эриксон проводил в то время исследование. На основе этих материалов Эриксон [13] в конце концов написал свой классический труд «Описание индейцев юрок» [13]. Крёбер изучал это племя примерно лет за 20 до этого и вдохновил Эриксона на дальнейшее его исследование. Отправляясь к индейцам юрок, Крёбер взял с собой Эриксона и познакомил его со своими оставшимися в живых друзьями, бывшими 20 лет назад его информантами. Крёбер открыл семинар описанием индейцев юрок. Это описание содержало не только сведения об их географическом положении, экономике, включенности в культурный ареал, социальной организации, физических артефактах и т. п., но и утверждение, что они представляют собой типичное воплощение структуры анального характера по Фрейду. Я вспоминаю, как он кратко обрисовал эту структуру и индейцев юрок с помощью длинной цепочки аллитераций, используя прилагательные, начинающиеся на «п»: что-то вроде «прижимистый», «педантичный», «пуриганский», «подозрительный», «придирчивый», «параноидный» и т. д., хотя деталей я не помню. Во всяком случае, Крёбер приготовил сцену для ортодоксальных, не допускающих сомнений упражнений в психоаналитической антропологии. Крёбер сам подвергался психоанализу и проводил психоанализ в течение нескольких лет, с 1918 по 1922 г., в Сан-Франциско (см. Т. Kroeber [21]). Его верность антропологической приверженности фактам, очевидно, была сильнее его преданности теории Фрейда, что видно из знаменитых обзоров Крёбера в книге «Тотэм и табу» (А. Кроебер [19, 20]), но в данном случае он придерживался ортодоксальной фрейдистской ориентации. Несомненно, он вдохновил Эриксона на посещение индейцев юрок, поскольку был уверен, что они представляют собой некий идеальный тип во фрейдовской типологии культур и характеров.

Эриксон был в то время вполне ортодоксальным

284

фрейдистом. Он разработал психоаналитическую этнографию индейцев племени сиу [14], у которых удлинённый период кормления грудью, сопровождаемый запрещением кусать материнскую грудь, приводил к компенсаторной орально-дентальной фиксации, выражавшейся в язвительном юморе и нервной привычке кусать палки, соломинки и ногти. Он творчески разработал теорию детского развития Фрейда, создав свою детализированную типологию форм и зон либидональной фиксации в детском возрасте [12, с. 44—92], но она была всего лишь ортодоксальной разработкой, хотя и отличавшейся известной оригинальностью. Если бы в качественном исследовании случаев в рамках отдельно взятой совокупности культурных условий дисциплина степеней свободы отсутствовала в той же мере, что и в моих карикатурных представлениях, то в этом случае ортодоксальный фрейдовский анальный синдром был бы, конечно, подтвержден. Но этого не произошло. Ни в настоящих, ни в прошлых приемах обучения детей племени юрок навыкам ухода за собой Эриксон не смог найти ничего такого, что должно было бы вызывать анальную фиксацию какого бы то ни было типа. Это яркий пример *исследования отдельных случаев, подрывающего авторитет теории*, если воспользоваться терминами, предложенными в анализе Лиджфарта [23], однако это опровержение оказывается значительно более действенным, чем то, которое он допускает, поскольку он тоже игнорирует дисциплинированность, проистекающую из богатства релевантных деталей.

Эриксон не только отказался (по крайней мере, в случае с индейцами юрок) от гипотезы об анальном типе характера, он выдвинул другую гипотезу и проверил ее на материале этого племени. Из бесчисленных альтернатив, которые приходили ему на ум в качестве возможных решений загадок, связанных с данным материалом, *большинство оказывались по той или иной причине несостоятельными*. Я подчеркиваю эту фразу, поскольку в ней, как мне кажется, кратко выражено представление о пользе многочисленных импликаций, многочисленных наблюдений («по той или иной причине») в концептуальной сфере некоей многочисленной совокупности импликаций любой конкретной теории: степеней свободы на основе многочисленных имплика-

ций. Та единственная теория, на которой он, в конечном счете, остановился, была подкреплена многочисленными разнообразными наблюдениями того же рода, что избавили его от прочих гипотетических решений. У меня здесь слишком мало места, чтобы с достаточной убедительностью изложить решение, к которому пришел Эриксон, и оно, само по себе неправдоподобное, в следующей крайне упрощенной формулировке прозвучит еще менее правдоподобно: вместо оральной и анальной фиксаций у индейцев юрок была фиксация на всем пищеварительном тракте. Это положение придало организованность поразительному обилию необычных черт этой культуры, таких, как повышенное внимание индейцев юрок к реке Кламат, протекающей через всю их маленькую долину, их парильни с отдельными входами и выходами, их назидательные сказки для детей, включая историю о птичке колибри, которая влетела в рот медведю и вылетела через анальное отверстие, и т. д. Чтение самой монографии убеждает в еще большей степени. (Моя собственная семинарская работа, выполненная в соавторстве с Ф. М. Гейером, была попыткой психологического осмысления 30 или 40 мифов индейцев юрок с помощью прожективных тестов. Во время работы семинара мы, насколько я помню, новой теорией Эриксона не располагали. Результат этой попытки был столь удручающим, что мы опустили до студенческого юмора и в качестве эпиграфа поместили на титульном листе следующий афоризм, приписав его Конфуцию: «One speaks verily of myth-interpretation only if one lispeth»<sup>1</sup>.)

Этот дисциплинирующий аспект добросовестных, тонких, самокритичных исследований отдельных случаев, который я пытался связать здесь со статистическим представлением о наличии достаточного для проверки пригодности гипотез числа степеней свободы,

является, несомненно, одним из аспектов принципов сопоставления образов и зависимости от контекста (Campbell [6]). Кроме того, я с опозданием присоединяюсь к тем же доводам, что и Лассуэлл [22] и другие (например, Рейзер [28]), настаивая на «конфигу-

---

<sup>1</sup> Уста глаголят истину о толковании мифов, только если они фепелявят (англ.). — Прим. перев

рационном подходе». Тщательный эпистемологическо-статистический анализ ситуации все еще является делом будущего. Это мое частичное отречение не что иное, как начало, и, желая вновь привлечь внимание к проблеме, я, вероятно, чрезмерно ее защищаю. Конечно, теоретический ум способен на удивительно гибкие рационализации *post hoc* по поводу любого результата и может в ходе таких рационализаций чрезмерно увлекаться истолкованиями, спекулировать на случайностях и исчерпывать степени свободы. В качестве примера в той области исследований, которую представляет Эриксон, можно обратить внимание на контраст между тем, как представлены у Кардинера [16] данные, собранные Дюбуа о племени алор, и как это сделано у нее самой [11]. Развивая свою тему оральной депривации, Кардинер уделяет большое внимание ребенку, который, получив для игры сосуд с ручками, попытался сосать одну из ручек и был сфотографирован за этим занятием. Кардинер игнорирует более релевантные факты, но свидетельствующие о противном, в том числе отсутствие у этих детей привычки сосать большой палец. В общем, когда тщательное количественное исследование не подтверждает какое-нибудь широко распространенное представление, то количественному исследованию верят обычно больше. Но оно в свою очередь отвергает это единичное убеждение здравого смысла, лишь опираясь на множество других. Конечно, обыденное социальное знание не свободно от многих специфических предубеждений; некоторыми из них я сейчас займусь. Но этот путь остается единственным. Даже когда мы заняты его совершенствованием, мы должны пройти его и принять за основу.

Свидетельство Беккера и случай Эриксона противоречат хорошо известной всеобщей склонности видеть вещи такими, какими мы ожидаем их увидеть. Фрэнсис Бэкон считал это первым из «идолов [призраков] рода». «Человеческий разум в силу своей склонности легко предполагает в вещах больше порядка и единообразия, чем их находит... Разум человека все привлекает для поддержки и согласия с тем, что он однажды принял... уму человеческому постоянно свойственно заблуждение, что он более поддается положительным доводам, чем отрицательным, тогда как по справедли-

вости он должен был бы одинаково относиться к тем и другим...»<sup>1</sup>. Лично я принимал участие в одном эксперименте, четко подтверждающем это положение на материале перцептивных показаний при восприятии расфокусированных картинок (Wyatt, Campbell [35]); в литературе найдется еще много других таких свидетельств (Campbell [9]). Но, хотя эта склонность в какой-то степени является всеобщей, это все-таки маргинальное свойство. В нашем эксперименте точность оценок, с которой должен был бы восприниматься сигнал в отсутствие соответствующих ожиданий, снижалась на 15% или около того; этот эффект исчезал, когда картинка попадала в фокус (Wyatt, Campbell [35]). Это, конечно, некая предрасположенность, свойственная как восприятию физических объектов, так и социальной перцепции. Она часто вступает в конфликт с другой предрасположенностью, связанной с колебаниями в уровнях адаптации (Nelson [15]), которая заставляет наблюдателя подмечать неожиданные контрасты ценой отказа от восприятия тождества (Campbell [8, с. 341—344; 91]).

Поскольку исследования отдельных случаев в границах одной культуры по-прежнему будут оставаться главной формой сравнительной социальной науки, не мешало бы внести ряд предложений, направленных на совершенствование дисциплины, которая необходима в таких исследованиях для опробования теории. Эти предложения основываются на аналогиях из области количественных исследований. Как мы видим, количественные исследования в том виде, в каком они публикуются, ни в коей мере не свободны от этих проблем. Однако количественные исследования ведут к попыткам установить уровни значимости, которые в свою очередь приводят к осознанию сложных аспектов проблемы. Основная общая рекомендация состоит в том, чтобы исследователь, который проводит изучение отдельных случаев в одной ситуации (ареале), вел более эксплицитную, развернутую по форме регистрацию аналогичных аспектов всей своей деятельности по решению задач.

---

<sup>1</sup> Бэкон Фрэнсис. Соч. в двух томах, т. 2, М., 1978, с. 20, 21.

Самой известной проблемой, касающейся степеней свободы при проверке значимости, является проблема числа наблюдений, в отношении которых проверяется гипотеза. В случае сходных повторных наблюдений в некотором ограниченном культурном ареале полевой исследователь обычно обращает на это внимание, например на число наблюдаемых деревень, в которых осуществляется данная форма лидерства, и деревень, в которых она отсутствует. Здесь предлагается, чтобы регистрация аналогичных процессов служила для накопления импликаций данной теории, приближаясь к полной регистрации всех связанных с проверкой теории мыслей и исследований, которые принимают следующую форму: если теория *A* верна, то должны следовать *B*, *C* и *D*. Там, где эти импликации побудили исследователя к активному поиску, их регистрация не составит труда. Там, где проверка была адекватно выполнена в ходе самого мыслительного процесса, многие из этих принятых во внимание импликаций будут опущены. Но если даже проверка протекает в форме мысленных проб, внимание, уделяемое данной проблеме в самосознании, может привести к гораздо более полной регистрации, чем это имеет место в теперешних монографиях, посвященных исследованиям отдельных случаев. Несомненно, что этот мучительный поиск разгадки во многом бессознателен в буквальном смысле слова и/или, как утверждает Пуанкаре [27], протекает во сне. Полная регистрация, таким образом, вещь недостижимая. Вероятно, наиболее распространенным упущением должно быть игнорирование неподходящих импликаций, особенно наблюдателями, которые в большей мере ориентированы на теорию. Накопленная литература, содержащая сведения о других исследованиях и критику их, способна сдерживать эту тенденцию не сразу и лишь частично. Нам необходима традиция сознательного использования противоположного процесса, в котором другие эксперты побуждаются включаться в поиск иных импликаций данной теории и иных фактов, противоречащих ей или поддерживающих ее. Кое-что в этом направлении уже сделано, но не в той форме, которая могла бы привести к образованию фонда «попаданий и промахов». Дальнейшим ограничением неизбежных остатков этноцентризма у различных социаль-

ных исследователей могла бы служить практика использования критических или одобрительных примечаний и комментариев местных социальных исследователей и публикация этих комментариев наряду с основным текстом исследований. Это способствовало бы также и расширению фонда попаданий и промахов. Почти для каждого связанного с какой-либо теорией классического исследования отдельных случаев имеются

до сих пор не проверенные предсказания, которые могли бы быть использованы в целях перекрестной валидации. (Возможно, мифы индейцев юрок, над которыми трудились мы с Гейером, могли бы послужить таким полигоном для гипотезы Эриксона.) Если исследование важно, оно заслуживает подтверждающего изучения отдельных случаев.

Другой вопрос, касающийся степеней свободы и проверки значимости, который стоит позаимствовать из количественных исследований, — это различие одностороннего и двухстороннего критериев значимости. Его, по-видимому, можно перевести на язык исследования отдельных случаев как различие подтверждающей ценности, которую имеет согласие между импликацией и фактами, когда теория была выбрана в свете этих фактов, и более высокой подтверждающей ценности, если теория выбиралась без учета этих подтверждений. Конечно, в нашем фонде эти два типа исследований должны были бы храниться отдельно. Более того, когда, например, теория верно предсказывает факт, очень неожиданный с точки зрения здравого смысла или других теорий или культур, то в этом случае подтверждение, строго говоря, является гораздо более убедительным, чем когда предсказание банально. В традиционных проверках значимости этот момент осознается не слишком четко (Meehl [25]), но в истории науки он известен по таким фактам, как использование Галилеем фаз спутников Юпитера. Вероятно, можно воспользоваться Байесовой ориентацией и говорить о начальных вероятностях, приписываемых теории *A*, в отличие от начальных вероятностей, приписываемых иным теориям. Такое применение формулы Байеса вряд ли типично: она больше применима к случаю одной теории, когда предсказанные начальные вероятности сравниваются с вероятностями, полученными

290

после сбора данных. Однако это возможно, и можно смело рекомендовать эпистемологу, занимающемуся Байесовой статистикой, обратить внимание на проблему проверки теории в исследованиях случаев. Эту проблему следует, кроме того, разрабатывать отдельно для каждого из шести типов исследования отдельных случаев, предложенных Лиджфартом [23] (то есть для гипотезо-образующего, подтверждающего теорию, подрывающего теорию и т. д.), уделяя особое внимание различию в степенях свободы между такими ситуациями, когда ареал, от изучения которого ждут подтверждения теории, определяется после явного или неявного поиска среди многих культур с целью обнаружить именно такой случай и когда в противоположность этому на какое-нибудь подтверждение наталкиваются вслепую. Предварительное знание об экстремальности либо одной из двух релевантных переменных, либо обеих, либо ни одной из них влияет на проблему степеней свободы. Анализ Лиджфарта имеет самое непосредственное отношение к рассматриваемому здесь вопросу, хотя он, как и я в своих более ранних работах, игнорирует проверку теории на основе многочисленных импликаций в условиях отдельно взятого случая.

Наконец, еще одна проблема, касающаяся степеней свободы, пока что недостаточно освещена в опубликованных статистических исследованиях, но интерес к ней растет. Она связана с проверкой многочисленных гипотез и последующим оформлением в качестве выводов тех из них, «статистическая значимость» которых установлена путем применения критерия значимости, предполагающего, что исследование проводилось на основе только одной этой гипотезы. Так, если данные получены фактически на основе случайных чисел и если исследуются все возможные взаимосвязи между 15 переменными, что дает 105 гипотетических связей из двух переменных каждая, то  $\frac{1}{20}$ , или более 5 из них, будет признана «значимой» на уровне значимости 5%. Эта проблема встречается в литературе под именем «появление ошибки, связанной с экспериментом» (Ryan [30]) и «проблема множественных сравнений» (Scheffé [31]) или проблема «вычерпывания данных» (Selvin, Stuart [32]). В сочетании с проблемой числа наблюдений ее можно выделить в особенно острой форме в

политологических исследованиях, посвященных коррелированию показателей, характеризующих нации. Если у нас столько же переменных, сколько наций, множественная корреляция, связывающая любую переменную с другими, будет равна 1,00 даже в том случае, если данными служат случайные числа. (В этом можно убедиться на материале формулы «сокращения» значения  $R$  (Lord, Novick [24, с. 286]). Значения  $R$ , вычисляемые на малых выборках, как правило, оказываются меньше, когда их проверяют на новых выборках. С помощью упомянутой формулы определяется подобное сокращение при отсутствии перекрестной валидации. В исследованиях, использующих факторный анализ, обычно наблюдается, что при меньшей выборке наций вес большего фактора и большее число факторов зависят от более высокой выборочной варибельности коэффициентов корреляции с малым значением  $n$ . Тем самым получает надлежащее объяснение неоднократно наблюдаемый результат: структуры факторов оказываются более четкими и больше объясняют тогда, когда данные анализируются по континентам, чем когда они собраны со всего света.

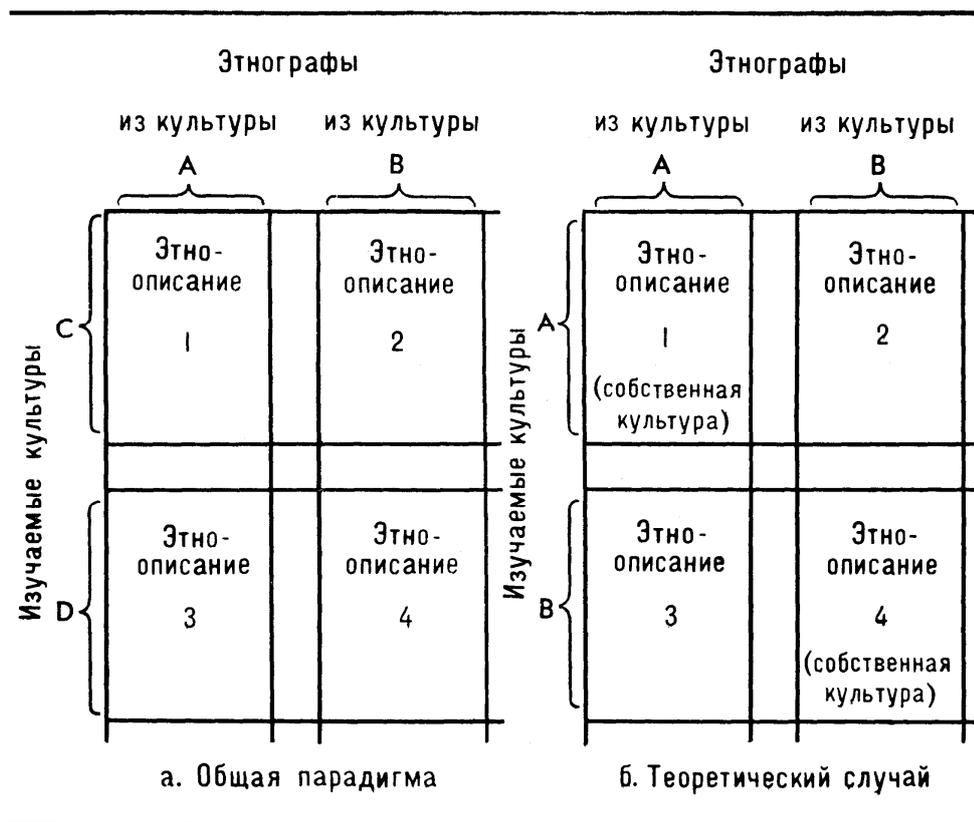
Чтобы учесть эту проблему в ходе гипотезообразующего изучения случаев, каковым является, например, вторая фаза исследования Эриксоном индейцев юрок, следует регистрировать все теории, рассматриваемые исследователем в процессе творческого решения стоящих перед ним задач. Чтобы представить степени свободы, возникающие на основе множественных импликаций, следует также регистрировать импликации, относительно которых проверялась каждая из теорий, и вести счет попаданиям и промахам. Лично я убежден, что теория «пищеварительного тракта», выдвинутая Эриксоном, лучше, чем псевдосовершенная множественная корреляция, там, где исчерпаны степени свободы путем проверки слишком большого количества гипотез на слишком незначительном числе случаев или импликаций. Я убежден и в том, что исследования случаев с точки зрения их ценности для проверки теорий могут быть в этом отношении усовершенствованы. Кроме того, когда признается реальность взаимодействия более высоких порядков между переменными (например, то обстоятельство, что связь между переменными  $A$  и  $B$

различна для разных уровней  $C$ ,  $D$ ,  $E$  и т. д.), следует признать и тот факт, что избыток правдоподобных гипотез и малочисленность или высокая стоимость эмпирических примеров создают реальные пределы для сравнительной социальной науки. Однако познание эмпирических примеров в рамках пространства импликаций, чтобы дополнить эти примеры знанием людей, поселений, наций, эпох и т. д., хотя бы в некоторой степени уменьшает остроту проблемы.

Выскажем еще одно предположение в целях усовершенствования исследования случаев. Изучение отдельных случаев, как показано выше, является в действительности сравнением двух случаев — местной и иноземной культуры. Но это крайне асимметричное сравнение, асимметричное по многим существенным параметрам: одна из культур осваивается так, как осваивалась бы ребенком, у которого нет альтернатив, и ее освоение идет параллельно с овладением местными культурными устоями, предрасположенностью и языком социального знания. Другая культура осваивается так, как осваивалась бы она взрослым иностранцем. Детали этой новой культуры, составляющей предмет исследования, находятся в центре внимания и точно определены. ИмPLICITным остается только тот факт, что основанием для сравнения служит собственная культура. Если бы черты собственной культуры были выделены столь же рельефно и эксплицитно, если бы они «изучались» непосредственно, к имPLICITному сравнению прибегали бы гораздо реже. Но хотя это и было бы известным усовершенствованием, черты своей собственной культуры все равно казались бы более разумными, обоснованными, более доступными интуитивному

постижению и более моральными. Черты же иноземной культуры представлялись бы, как правило, произвольными, странными, загадочными и чуть ли не безнравственными. Чтобы исправить это положение, я высказал в одной из своих более ранних работ [7] следующее предположение:

*«Триангуляция как преодоление приверженности наблюдателя своей собственной культуре».* В науке создание полезных «реалистических» конструкторов требует множества методов, сконцентрированных на диагностировании одного и того же конструктора с ряда неза-



Р и с. 1. Схемы множественных этноописаний, позволяющие вычлнить из общего описания изучаемой культуры содержание, вносимое тем или иным этнографом.

висимых позиций наблюдения с помощью своего рода триангуляции. Это необходимо потому, что чувственные данные или показания измерительного прибора рассматриваются в настоящее время как результат взаимодействия, в который каждый его участник — наблюдатель (или измерительный прибор) и объект исследования — вносит свой вклад при определении формы этих данных. Располагая единичным наблюдением, невозможно отделить субъективный компонент от объективного. Когда же наблюдения, полученные с помощью различных инструментов и с разных наблюдательных пунктов, могут быть сопоставлены как отражающие «одни и те же» объекты, появляется возможность выделить в наших данных компоненты, обязанные своим происхождением наблюдателю (инструменту) и наблюдаемому. Оказывается, процесс такого «разведения»

требует как множества наблюдателей (методов), так и множества различных объектов исследования.

В применении к исследованиям философии культуры это означает, что наше типичное исследование одной отдельно взятой культуры одним наблюдателем по природе своей весьма неопределенно. По отношению к любой данной характеристике отчета нельзя с уверенностью утверждать, является ли она характеристикой наблюдателя или характеристикой наблюдавшегося объекта. Быть может, идеальная парадигма, позволяющая исправить это положение, выглядит так, как показано на рис. 1 а.

Согласно самой общей схеме, два антрополога, представители двух разных культур, занимаются каждый изучением третьей и четвертой культуры. В полученных в результате такого исследования четырех этноописаниях 1-е и 3-е характеризуются общими свойствами, которые отсутствуют во 2-м и 4-м, и могут быть, таким образом, приписаны этнографу *A*, тогда как свойства, объединяющие 2-е и 4-е и отсутствующие в остальных описаниях, — этнографу *B*. Прослеживая соответствия по рядам на рисунке, мы можем приписать свойства, объединяющие этноописания 1-е и 2-е и не представленные в 3-м и 4-м, культуре *C* как «объективно» известные. Свойства, общие для всех четырех этноописаний, по природе своей неоднозначны, их можно интерпретировать и как результат ошибок — отклонений, общих для обоих этнографов, — и как итог культурной общности изучаемых сообществ. Отметим в этой связи желательность взаимного сравнения этнографов, различающихся, насколько возможно, по своей культурной основе. Если же этнографы принадлежат к одной культуре, повторение результатов становится скорее делом надежности, чем валидности, в том смысле, в каком эти термины употребляются при обсуждении психологических тестов. Если подобное исследование проводится четырьмя этнографами (по двое из культуры *A* и культуры *B*), которые изучают в культурах *C* и *D* разные поселения, чтобы избежать взаимных помех и сговора, то свойства, присущие только одному какому-то из полученных этноописаний, могут быть приписаны неопределенному сочетанию специфики данного поселения в рамках данной культуры, личностных

особенностей этнографа и взаимодействия специфической культуры этнографа с изучаемой культурой. (Если бы из каждой культуры привлекался только один этнограф и если бы эти два этнографа по очереди изучали в исследуемой культуре одно и то же поселение, то в таком случае было бы неясно, чему обязаны своим происхождением свойства, характерные исключительно лишь одному из четырех полученных этноописаний: взаимодействию этнографа и культуры или влиянию временной последовательности, в какой этнограф изучал свою вторую культуру, или влиянию временной последовательности, в какой сообщество реагировало на работу второго исследователя, или фоновым изменениям, или взаимодействию между всеми этими факторами.) Наличие подобной неопределенности не следует замалчивать, нельзя также допустить, чтобы остались в тени те огромные достижения в понимании явлений, к которым подводят такие исследования с участием многих этнографов.

Хотя множественность культур, как тех, к которым принадлежат этнографы, так и тех, которые становятся предметом исследования, и является идеалом, значительных достижений можно было бы добиться даже в том случае, если бы нам удалось реализовать одну лишь верхнюю половину схемы, изображенной на рис. 1а, то есть когда этнографы из двух культур сосредоточивают свое внимание на изучении одной-единственной культуры. При всех подобных триангуляциях мы вновь сталкиваемся с парадоксальной неспособностью воспользоваться различиями, когда они столь существенны, что делают невозможным взаимное совмещение соответствующих аспектов сравниваемых отчетов. Необходимость в этом общем знаменателе подтверждает справедливость защищаемого

Хоккетом положения о включении материальных и поведенческих деталей культуры даже в этноописания, сконцентрированные на вопросах детерминации философии культур.

Другой вариант модели с использованием многих этнографов и множества объектов описания предполагает взаимное изучение двух культур, как это показано в схеме на рис. 1б. Обычно внимание сосредоточивается на этноопнеаниях 2 и 3 (отчет *A* о *B* и отчет *B*

296

об *A*). Однако описание *A*, имеющееся у *A*, и описание *B*, имеющееся у *B*, содержатся в сравнении как неявные основания отсчета и соотношений. Такие отчеты представляют, вероятно, определенную научную ценность, хотя бы как совокупности стереотипных представлений друг о друге или как показатели уровня консензуса соседних общин. Как только оценочный компонент (каждое племя рассматривает себя как наилучшее) устраняется, такие совокупности взаимных стереотипов обнаруживают удивительное согласие в подтверждении характера групповых различий» (Campbell [7, с. 331-333]).

В плане настоящего обсуждения я бы расширил теперь эти требования, предложив, чтобы во второй фазе полевого исследования каждый этнограф проводил проверку и подвергал сомнению валидность интерпретации, которую давал исследуемой культуре параллельно изучавший ее этнограф.

#### ЗАКЛЮЧИТЕЛЬНЫЕ ЗАМЕЧАНИЯ

Поисковый характер этой статьи очевиден, она является резким отказом от моего прежнего догматического пренебрежения к исследованиям случаев и демонстрирует другую крайность. Хотя эта попытка не окажет существенного влияния на мое же собственное учение о квазиэкспериментальных планах и методах исследования, в настоящий момент положения этой статьи кажутся мне правильными. В конце концов, человек в своих обыденных делах является весьма компетентным, и качественное обыденное знание не замещается знанием количественным. Скорее напротив, количественное знание должно доверять качественному, в том числе и обычному восприятию, и основываться на нем (Campbell [5]). Мы, методологи, должны прийти к прикладной эпистемологии, объединяющей в одно целое и то и другое.

#### ЛИТЕРАТУРА

1. Бэкон Ф. Новый органон. Сочинения в 2-х томах. Т. 2., 1970.
2. Вебер Н. Sociological Work. Chicago, Aldine, 1970.
3. Вoesch E. F., Eckensberger L. H. Methodische Probleme des interkulturellen Vergleichs. - In: C. F. Graumann (ed.).

297

- Socialpsychologie, vol. 7 of Handbúch der Psychologie. Gottingen, Verlag fur Psychologie, 1969.
4. Boring E. G. The nature and the history of experimental control. *Amer. J. of Psychology*, 1954, 67, 573—589.
  5. Campbell D. T. Qualitative knowing in action research Kurt Lewin Award Address. New Orlean, 1974.
  6. Campbell D. T. Pattern matching as an essential in distal knowing. — In: K. R. Hammond (ed.). *The psychology of Egon Brunswik*. N. Y., Holt, Rinehart & Winston, 1966.
  7. Campbell D. T. Distinguishing differences of perception from failures of communication in cross-cultural studies. — In: F. S. C. Northrop, H. H. Livingston (eds.). *Cross-Cultural Understanding: Epistemology in Anthropology*. N. Y., Harper & Row, 1964.

8. Campbell D. T. The mutual methodological relevance of anthropology and psychology. — In: F. L. K. Hsu (ed.). *Psychological Anthropology: Approaches to Culture and Personality*. Home-wood, Ill. Dorsey Press, 1961.
9. Campbell D. T. Systematic error on the part of human links in communication systems. *Information and Control*, 1959, 1, 334—369.
10. Campbell D. T., Stanley J. C. *Experimental and Quasi-Experimental Design for Research*. Chicago. Rand McNally, 1966.
11. DuBois C. A. *The People of Alor*. Minneapolis, Univ. of Minnesota Press, 1944.
12. Erikson E. H. *Childhood and Society*. N. Y., W. W. Norton, 1950.
13. Erikson E. H. Observations on the Yurok. *American Archaeology and Ethnology*, 35, № 10. Berkeley, Univ. of California Press, 1943.
14. Erikson E. H. Observations on Sioux education. *J. of Psychol.* 1939, 7, 101 — 156.
15. Helson H. *Adaptation-Level Theory*. N. Y., Harper, 1964.
16. Kardiner A. *The Individual and His Society*. N. Y., Columbia Univ. Press, 1939.
17. Kenyatta J. *Facing Mount Kenya*. London, Secker & Warburg, 1938.
18. Kiddler L. H. Foreign visitors: a study of the changes in selves, skills, and attitudes of Westerners in India. Ph. D. diss. Northwestern University, 1971.
19. Kroeber A. L. Totem and taboo in retrospect. *Amer. J. of Soc.* 1939, 45, 446—451.
20. Kroeber A. L. Totem and taboo: an ethnologic psychoanalysis. *Amer. Anthropologist.*, 1920, 22, 48—55.
21. Kroeber T. Alfred Kroeber: A Personal Configuration. Berkeley, Univ. of California Press, 1970.
22. Lasswell H. D. The future of the comparative method. *Comp. Politics*, 1968, 1, 3—18.
23. Lijphart A. Comparative politics and the comparative method. *Amer. Pol. Sci. Rev.*, 1971, 65, 682—693.
24. Lord F. M., Novick M. R. *Statistical Theories of Mental Test Scores*. Reading, Mass., Addison-Wesley, 1968.
25. Meehl P. E. Theory-testing in psychology and physics: a methodological paradox. *Philosophy of Set.* 1967, 34, 103—115.

26. Natio R. *Data Quality Control: A New Research Technique*. N. Y., Free Press, 1962.
27. Poincaré H. *Mathematical creation*. — In: H. Poincaré, *The Foundations of Science*. N. Y., Science Press, 1919.
28. Raser J. R. *Simulation and Society*. Boston, Allyn & Bacon, 1969.
29. Raser J. R., Campbell D. T., Chadwick R. W. Gaming and simulation for developing theory relevant to international relations. *General Systems*, 1970, 15, 183—204.
30. Ryan T. A. Significance tests for multiple comparisons of proportions, variances and other statistics. *Psych. Bull.* 1960, 57, 318—328.
31. Schéffé H. A method for judging all contrasts in analysis of variance. *Biometrika*, 1953, 40, 87—104.
32. Selvin H. C., Stuart A. Data-dredging procedures in survey analysis. *Amer. Statistician*, 1966, 20, 20—23.
33. Stouffer S. A. (ed.). *The American Soldier*. Princeton, Princeton Univ. Press, 1949
34. Tolman E. C., Krechevsky I. Means-end-readiness and hypothesis. *Psych. Rev.*, 1933, 40, 60—70.
35. Wyatt D. F., Campbell D. T. On the liability of stereotype or hypothesis. *J. of Abn. and Soc. Psychol.*, 1951, 46, 496—500.

## ДЕЦЕНТРИРОВАННЫЙ ПЕРЕВОД ВОПРОСНИКОВ И ТЕСТОВ<sup>1</sup>

Проблема перевода, особенно в контексте сравнительных социальных исследований, имеет самостоятельное эпистемологическое значение. Если гипотезу Сепира — Уорфа понимать буквально, то перевод с одного языка на другой невозможен. В эпистемологическом плане Куайн [39] детально разработал это положение. Однако в наши цели не входит теоретическое обсуждение гипотезы лингвистической относительности. На практике в сравнительных межкультурных исследованиях никто не придерживается этого крайнего взгляда и никого не останавливают выводы эпистемологов о том, что достоверное знание здесь невозможно.

Перевод с одного языка на другой — важнейшая процедура для всех, кто интересуется межкультурными исследованиями. В этой работе мы рассмотрим специальный, но в методическом плане очень важный вопрос о переводе психологического инструментария — вопросников и тестов. Остановимся на одном специфическом различии целей перевода. С одной стороны, существует *симметричный*, или *децентрированный*, перевод, стремящийся сохранить как верный смысл, так и привычность и разговорность как в том, так и в другом языке. С другой стороны, существует *асимметричный*, или *уницентрированный*, перевод, в котором доми-

---

<sup>1</sup> Campbell D. T. Decentering the Translation of Questionnaires and Tests. Переработанный для данного издания вариант статьи: Werner O., Campbell D. T. Translating, Working Through Interpreters, and the Problem of Decentering. — In: A handbook of method in cultural anthropology. Ed. by R. Naroll, R. Cohen N. Y. The Natural History Press, 1970. © The Natural History Press. 1970.

нирует верность одному из языков, обычно языку оригинала. И хотя в обоих случаях идеал недостижим, на практике мы сталкиваемся с большими различиями.

Возьмем, к примеру, такой нежелательный случай асимметричного перевода, как распространенная практика перевода английских версий «личностных тестов» на иностранные языки. Если требуется эквивалентность перевода оригинальной «стандартизированной» форме, на свет появляется неуклюжая, экзотическая версия теста на языке, на который осуществляется перевод, и здесь существует только одна возможность: рассматривать саму английскую версию в процессе подготовки «эквивалентных» тестов как постоянно доступную пересмотру, то есть осуществлять децентрированный перевод. В то же время приемлемый этнографический перевод песен,

церемоний, устных преданий и мифов, как отмечает Боханнэн [7], в принципе асимметричен. Произведение, звучащее на языке оригинала как разговорное и привычное, соответственно переводится на язык исследователя как нечто экзотическое и неестественное. Эта экзотичность может заключаться в сохранении местных терминов и, таким образом, предполагает знакомство читателя с новым словарем, когда соответствующие понятия в его языке отсутствуют. Стремление же к достижению эквивалентной типичности путем буквального перевода не позволило бы передать важнейшую информацию о культурных различиях. Конечно, как указывают С. Фёгелин и Ф. Фёгелин [45], экзотический материал отчасти, несомненно, обязан недостаточному пониманию переводчиком терминов социальных наук, имеющихся в языке исследователя, — это не необходимый и дезориентирующий источник ошибок.

Целью симметричного перевода является идеал, достигнутый в большинстве литературных переводов. Он, как таковой, почти непременно предполагает супралингвистический перевод культурных символов. Найда [35, с. 166] обозначает перевод такого типа термином «динамическая эквивалентность». Иными словами, культурный символ языка оригинала переводится таким культурным символом другого языка, который вызывает ту же функциональную реакцию. Боханнэн [7, с. 815] приводит в качестве примера историю с

301

расследованием таинственного убийства, в которой английский детектив сумел в темноте узнать некую женщину по ее произношению, в американской же версии для детектива пришлось придумать уличный фонарь, так как в Америке нельзя заговорить с женщиной из общества в темноте. Точно так же язык навахо может отождествлять койота с колдовством, в то время как в европейских языках и культуре ту же функцию могла бы выполнить черная кошка. Другими словами, перевод, помимо всего прочего, предполагает наличие известной культурной эмпатии, когда понадобится приспособить его содержание, контекст и стиль к условиям культуры-приемника.

В то время как этнографический перевод должен быть асимметричным, для большинства других целей антропологические полевые исследования нуждаются в децентрирующем и симметричном переводе. Децентрация предполагает снятие основного акцента с языка исследователя: вводится система символов, принадлежащая как той, так и другой культуре. Лучше всего, если децентрация стирает различия между языком оригинала и языком перевода и подчеркивает моменты сходства. Интралингвистический парафраз как основа для перевода, многоступенчатый перевод и обратный перевод — таковы некоторые из методов децентрации.

### **ПРИМЕНЕНИЕ ОБРАТНОГО ПЕРЕВОДА В ПОЛЕВОЙ РАБОТЕ С ИНТЕРВЬЮЕРАМИ**

Социальный исследователь, который в процессе интервьюирования вынужден прибегать к посредничеству переводчиков, почти или вовсе не зная языка, на который осуществляется перевод, практически беспомощен: он полностью зависит от своих переводчиков, не способен оценить их квалификацию и проинструктировать их относительно качества перевода, в котором он нуждается. В этих условиях методика обратного перевода дает ему в руки некоторую возможность контроля. После того как исследователь подготовил программу интервью на своем самом что ни на есть доступном переводе языка, он сажает за работу двух переводчиков, из которых один переводит на местный язык первую половину, а другой — вторую. Затем каждый из

них берется за половину, переведенную другим, и снова переводит ее на язык оригинала. (Это могут быть письменные переводы, если переводчики грамотны, или магнитофонные записи, обеспечивающие возможность обратного перевода.) В итоге исследователь располагает двумя версиями текста интервью на своем собственном языке и за счет их триангуляционной съемкой версии на местном языке, которая почти наверняка адекватна, если верны обе версии на языке оригинала.

Исследователю в какой-то мере непосредственно открывается качественный уровень его переводчиков. Если верен обратный перевод, то они оба заслуживают доверия. Если же нет, то трехстороннее обсуждение каждого из пунктов прямого и обратного перевода приведет к известному согласию по поводу частных проблем. Все это послужит свидетельством общей способности переводчиков, которое ими самими будет воспринято как достоверное. Вероятно, нет лучшего способа отобрать кандидатов на должность переводчика.

Результаты первых попыток таких обратных переводов, как правило, отнюдь не обнадеживают, о чем сообщает в одном из очень немногих опубликованных обсуждений этой методики Филлипс [38] и что подтверждает наш собственный опыт. В отличие от Филлипса мы, однако, принимаем это обычно низкое качество как нечто, с чем исследователь не может не столкнуться, и как доказательство ценности обратного перевода, а не как причину от него отказаться. Конечно, более неформальные способы использования одного переводчика для проверки переводов другого дают исследователю больше уверенности в его компетентности, но возможно, что это впечатление обманчиво. Исследователь не может контролировать такие показатели, как общераспространенные заблуждения, склонность профессионала воздерживаться от критики другого профессионала, местную круговую поруку, когда дело идет об эксплуатации чужеземца, и т. д. Обратный перевод обеспечивает исследователю весьма значительный (хотя, надо сказать, неполный) контроль над качеством прямого перевода, даже если он совсем не знает местного языка.

Обратный перевод позволяет исследователю получить важную информацию о том, какая часть его

интервью может с успехом использоваться и какая часть его интересов как социального исследователя непередаваема, по крайней мере, при тех способностях, которыми обладают его переводчики. Он будет вынужден пренебречь многими тонкими различиями, которые не могут быть переданы. Он внесет свою лепту в активный пересмотр «оригинала».

Обратный перевод — это самое мощное орудие подготовки переводчиков, какое только доступно исследователю. В целях обучения необходимо иметь возможность выявить ошибки, и исследователь получает ее, лишь обратившись к методике обратного перевода, если только он не крупный знаток местного языка. Неделя или около того, потраченная на обратный перевод по всему ряду вопросов, представляющих интерес для исследователя, в значительной мере повышает уровень знаний интервьюера и его готовность обсуждать альтернативные решения специфических проблем перевода. Иначе переводчик надевает амбициозную маску догматической профессиональной самоуверенности.

Даже для того большинства антропологов, которые не пользуются никакой предварительно заготовленной программой опроса, прибегая к услугам переводчиков в целях проведения импровизированных интервью, ход которых определяется инициативой респондентов (для основной массы полевых исследований это, несомненно, наилучший метод), дисциплина обратного перевода представляется желательной по упомянутым выше причинам. При отсутствии таких проверок зависимость от переводчиков неизбежно ведет к

изрядной доле самообмана в отношении качества перевода. (Этнонаучные методики, сводя весь вопрос к местному языку, избегают этих проблем ценой столь сильного утомления, что они редко применяются в качестве единственного метода.) Даже если усилия, затраченные на обратный перевод, не имеют своим итогом стандартных вопросов, которые можно было бы использовать при интервьюировании, они дают некоторую приемлемую совокупность переводов ключевых понятий исследователя на местный язык.

Еще одно специфическое применение обратного перевода — его использование в роли сокращенного приема исследования этнотаксономии в данной области

304

(LeVine, Campbell [30]). Исследователь подготавливает список терминов и приблизительных синонимов, относящихся к данной области (например, к межличностной агрессии). Одно лицо, знающее оба языка («билингв»), подбирает для *каждого* термина как можно большее число переводов на местный язык. Другой билингв подбирает как можно большее число английских переводов каждого местного термина. После чего к исходной таксономии языка оригинала, которой располагал исследователь, добавляются гетерогенные совокупности обратных переводов на язык оригинала; их-то он и пытается осмыслить, выдвигая гипотезы относительно таксономии в местном языке.

Методика обратного перевода, при всех ее достоинствах, не гарантирует от ошибок. Даже после того, как программа опроса успешно подвергнута обратному переводу, дальнейшее ее применение в полевых условиях может выявить вопросы, переведенные до смешного нелепо. Нелишне иметь некоторое представление о том, как они обнаруживаются. В этом пункте мы зависим от контекста и от предположений, связанных с уже полученным знанием, а также от того, насколько основательны наши суждения о данном ответе как о странном, отклоняющемся от темы, не имеющем отношения к делу. Источником таких ошибок часто оказываются установленные по словарю межъязыковые эквиваленты, которые использовались в процессе как прямого, так и обратного перевода.

Эти словарные правила перевода — результат навеянных школьным образованием и чтением Библии стандартизированных «решений», принимаемых по поводу трудных, но часто повторяющихся проблем перевода. Переводчику, менее знакомому с традицией, который никогда не видел словарей для перевода, который изучал каждый язык скорее путем участия в общении, чем по инструкции, и поэтому непосредственно переводит с языка оригинала на местный язык, удастся избежать подобных ошибок. (Другие источники ошибок могут корениться в многозначности местного языка, которая может быть особенно велика в его письменной форме.) Отметим, что, чем шире контекст и чем выше избыточность, тем больше шансов обнаружить ошибки, допускаемые переводчиком. Связные рассказы-описания

305

больше всего подходят для этих целей, вопросы, предполагающие краткий ответ, менее пригодны. Ответы «да — нет» дают столь малую возможность установить меру понимания респондента, что при работе через переводчика их ценность сводится к нулю.

Методика обратного перевода в целом нуждается в более обширном исследовании, равно как и в более скрупулезном накоплении имеющегося опыта и информировании о нем.

Необходимо также изучение истории этого метода. Он, вне всякого сомнения, изобретался несколько раз самостоятельно. Антропологи, проводившие опрос общественного мнения в Японии для нужд военного руководства США, применяли его примерно в 1946 г. (личное сообщение Дж. Беннета). Стерн и Д'Эппине [42] опубликовали самое первое его описание из всех нам известных. Этот метод часто применялся, хотя и без особого методологического обсуждения (Jacobson [23]; Schachter [40]; Duijker [13]; Ervin,

Bower [15]; Cantril [11]; Mitchell [33]; Almond, Verba [1]). Насколько нам известно, единственное обсуждение данного метода в рамках антропологии принадлежит Филлипсу [38].

### ПЕРЕВОД ЛИЧНОСТНЫХ ТЕСТОВ, ТЕСТОВ СОЦИАЛЬНЫХ УСТАНОВОК И СПОСОБНОСТЕЙ

Сравнительно-культурным исследованиям, использующим психологический инструментарий, немалый вред принесла идея «стандартизированных тестов», то есть убеждение, что существуют некие детально разработанные, пригодные в любых целях средства для сравнения, которым соответствуют наборы значимых «популяционных норм»<sup>1</sup>. Это понятие стандартизированных тестов —

---

<sup>1</sup> Детально методологические проблемы использования вопросников и тестов освещаются в ряде работ автора. См., например, Campbell D. T., Fiske D. W. Convergent and discriminant validation by the multitrait-multimethod matrix. — *Psychol. Bull.*, 1959, vol. 56, p. 81—105; Campbell D. T. Recommendations for APA test standards regarding construct, trait, or discriminant validity. — *Amer. Psychologist*, 1960, vol. 15, p. 546—553; Brewer M. B., Campbell D. T., Crano W. D. Testing a single factor model as an alternative to the misuse of partial correlations in hypothesis-

чуть ли не магическая посылка, согласно которой психологические тесты могут измерять непосредственно и успешно то, что они должны измерять, — ведет при переводе таких средств к прискорбной этноцентрической асимметрии. Версия теста на языке исследователя (например, на английском) считается неприкосновенной. Логический позитивизм даже рассматривал ее как операциональное определение того научного конструкта, который предполагается данной методикой (Campbell [9], Campbell, Fiske [10], Webb et al. [46]). В результате все усилия переводчиков направлялись на то, чтобы сохранить верность оригиналу. Образный, метафорический перевод, при котором идиомы языка, на который осуществляется перевод, заменяют собой аналогичные идиомы оригинала, не допускался из боязни нарушить взаимоотношения идентичности вопросов. Иллюстративные ссылки сохранялись даже в тех случаях, когда в них упоминались вещи, привычные с точки зрения языка оригинала и экзотические с точки зрения языка перевода. Отсюда и существенная асимметрия, вследствие которой привычные, разговорные, понятные вопросы теста на языке оригинала в тесте на другом языке становятся необычными, неуклюжими и трудными.

Ясно, что в случае интеллектуальных тестов одинаково одаренные люди будут хуже справляться с версией на местном языке. В случае других типов тестов характер нарушений нельзя предсказать без учета их конкретного содержания, но и здесь они будут, несомненно, играть столь же важную роль. Чтобы исправить дело, нужно рассматривать версию на языке оригинала как саму по себе подлежащую пересмотру. Такой пересмотр в сочетании с процедурами отбора вопросов приведет, очевидно, к тому, что оригинальная версия станет более банальной, менее трудной, более эксплицированной, менее разговорной, менее идиоматической, менее метафорической и не будет включать необычные вопросы. В итоге может быть получен тест, уступающий в отношении надежности и валидности первоначальной версии. Необходимо будет создать и сравнить

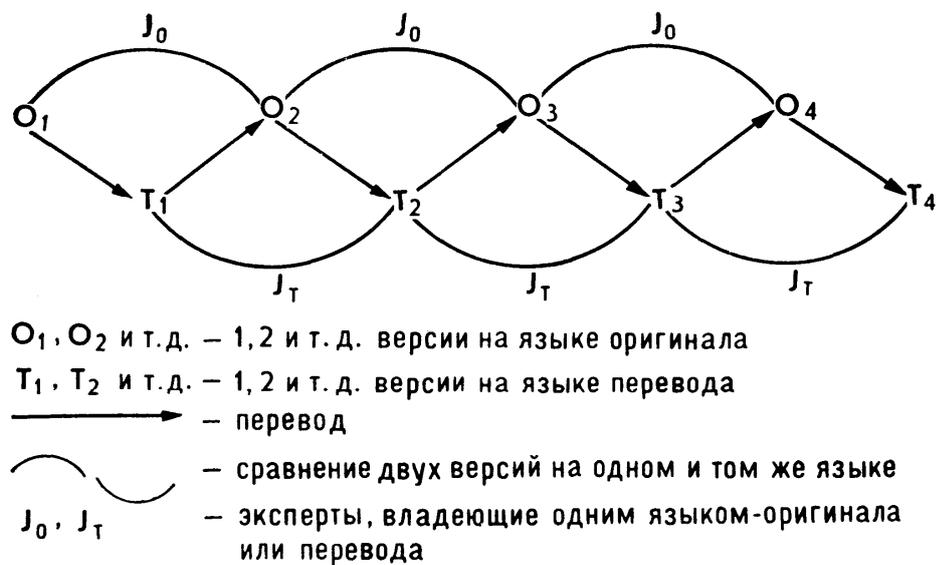
---

testing research. — *Sociometry*, 1970, vol. 33, p. 1—11; Crano W. D., Kenny D. A., Campbell D. T. Does intelligence cause achievement? A cross-lagged panel analysis. — *Journ. of Educational Psychology*, 1972, vol. 63, p. 258—275.

между собой новые группы обследуемых по новой методике испытуемых, говорящих на языке оригинала. Но, разумеется, лучше иметь сопоставимые данные по двум банальным тестам, нежели данные, несопоставимые из-за того, что банален только один из них.

Преобладание неосознанного предположения о неизменяемости исходной версии можно проиллюстрировать следующим рассуждением Беррина [6], взятым из одной его в целом блестящей статьи: «При переводе суждений Эдвардса, взятых из набора суждений для определения личностных предпочтений, на японский язык Ивахара столкнулся еще с одной трудно разрешимой проблемой. Вопросы, затрагивающие гетеросексуальные интересы, были слишком грубы и откровенны для особой чувствительности японцев. Он поэтому вписал новые суждения по теме гетеросексуальных интересов, более приемлемые в данной культуре. Но в таком случае имеем ли мы право, сравнивать японские данные с ответами американцев?» (Bergien [6, с. 39]). В своем пространном ответе на этот вопрос, предполагающем, что решением является процедура валидизации бикультурных конструктов, он упускает из виду другое решение — сбор в США новых данных, полученных при использовании хорошего обратного перевода на английский язык более деликатных суждений японской версии.

Обратный перевод служит чрезвычайно полезным методом для возможного пересмотра оригинала, равно как и первой версии перевода. Он, кроме того, служит эпистемологической моделью для сложного процесса децентрации. Рассмотрим оригинальный тест, его первый перевод и его первый обратный перевод. У нас теперь две версии на языке оригинала. Какую из них следует использовать при сборе данных на этом языке? Исследователь, сравнив оригинал с первоначальным обратным переводом, остановится, конечно, на оригинале. Но который из этих тестов более сопоставим с переводной версией? Без сомнения, обратный перевод — прежде всего вследствие своего сравнительно *низкого* качества. Если только перевод оказался возможным, у исследователя должно появиться желание собрать данные с помощью этого обратного перевода. Какие бы сомнения ни вызывал у него этот последний, те же самые чувства должен испытывать он и к переводной версии.



Р и с. 1. Многократно повторяющаяся последовательность обратных переводов.

Понятие обратного перевода, рассматриваемого как многофазовый повторяющийся процесс, служит идеальной концептуализацией децентрированного перевода. В дополнение к значительному контингенту компетентных двуязычных переводчиков, для части которых в идеале исходным является один язык, а для какой-то другой их части — другой, следует привлечь одноязычных экспертов по переводу на каждом из языков. Эти эксперты должны судить об адекватности каждого двойного перевода, сравнивая соответствующий исходный текст и его обратный перевод на их родной язык. Попытки пересмотра перевода должны продолжаться до тех пор, пока его качество не удовлетворит эксперта. Такой процесс иллюстрирует рис. 1. Для процесса децентрации важным моментом является присутствие эксперта, владеющего одним только языком перевода. Без него, при наличии одного-единственного эксперта (обычно самого исследователя), который владеет только языком оригинала, этот процесс будет по-прежнему приводить к уницентрированному продукту, хотя и не в такой степени, как в случае, когда версия на языке оригинала не подвергается вообще никаким модификациям.

Стратегия сравнения культур должна состоять в том, чтобы наблюдаемые нами культурные различия

309

получили одновременно несколько конкурентных интерпретаций и чтобы при этом различие в культурных установках или в личностных переменных было лишь *одним* из возможных объяснений. Недостатки перевода также будут фигурировать в этом ряду в качестве одной из правдоподобных конкурентных гипотез. Как и вообще в случае таких конкурентных гипотез (Segall, Campbell, Herskovits [41]; Frijda, Jahoda [16]), этот фактор можно исключить путем изучения влияния, оказываемого сознательно вносимыми вариациями. Так, если ответы на все вопросы теста для определения установок указывают на различия, несовершенство перевода перестает быть правдоподобной гипотезой. Обратите внимание здесь на следующие две рекомендации: во-первых, когда вопросов много, это лучше, чем когда задан один-единственный вопрос, и, во-вторых, анализировать результаты следует по отдельным вопросам, а не только по суммарному показателю. Вероятность несовершенного перевода столь велика, что даже в случае опроса общественного мнения, где, как правило, каждый отдельный вопрос рассчитан на выявление какого-то общего представления, следует использовать, по крайней мере, по два вопроса на каждое такое представление. Эти вопросы не должны содержать общих ключевых слов, но должны быть как можно более идентичны по смыслу.

Один из приемов проверки эквивалентности перевода и оригинала состоит в том, чтобы получить от билингов ответы на одни и те же вопросы, заданные им на обоих языках (см., например, Schachter [40]). Используемый как формально-статистический подход, этот прием на деле имеет целью установление не столько взаимоднозначного тождества вопросов, сколько эквивалентности средних дисперсий, равно как и соответствующих корреляций между показателями по двум формам вопросника или теста. Бенц [4] и Лоннер [32] сообщают о переводах стандартных тестов личности и интересов, которые приближаются к этим требованиям. В исследовании Бенца, в котором принимали участие 60 двуязычных латиноамериканских должностных лиц, межъязыковые корреляции по шести шкалам теста ценностей Олпорта — Вернона, полученные при пятимесячном перерыве между обследованиями, равнялись в

310

среднем 0,57, тогда как корреляция оценок первого и повторного тестирования с применением (на случайной популяции) одной только исходной англоязычной версии теста после трехмесячного перерыва составила 0,74. Аналогичные цифры для «шкал

интереса» Кудера были соответственно 0,75 и 0,90 при перерыве в два месяца. В исследовании Лоннера средняя межъязыковая корреляция, полученная на основе опроса 18 европейских психологов по 62 шкалам анкеты на выявление профессиональных интересов, которая предлагалась сначала на немецком, а через месяц на английском языке, равнялась 0,80, в то время как соответствующие корреляции между данными, полученными на гораздо менее однородной группе, которой дважды предлагалась английская версия, составили в среднем 0,91. И Лоннер, и Бенц утверждают, что оптимальный перевод должен был бы привести к тому, чтобы корреляции между разноязычными версиями приближались к данным двукратного опроса с применением одной из них. Но это требование, конечно же, нецелесообразно, поскольку, как мы уже подчеркивали, перевод в лучшем случае отличается от оригинала как его парафраз на том же самом языке. Если у нас есть парафраз каждого вопроса на том же языке и если мы в таком случае обладаем формой А и формой Б, то межъязыковая корреляция для формы А в лучшем случае должна быть не выше, чем внутриязыковая корреляция для форм А и Б при сопоставимом временном интервале между опросами. А эта последняя будет, разумеется, ниже, чем обычная корреляция между данными первого и повторного тестирования. Если у нас нет специально подготовленных парафразов, с этой целью можно использовать разбиение всех вопросов на пары с нивелированными различиями. Чтобы усложнить данную эпистемологическую проблему, отметим, что Эрвин [14] приводит правдоподобный случай для билингов, которые по-разному проявляются как личности в каждом из двух языков (см. также Anderson [2]).

Чаще же всего обращение к адекватному переводу тестов сводится к поиску сходства между разноязычными формами по уровню надежности и валидности, по факторной структуре и другим статистическим параметрам. Это превосходно, когда показатели хорошие,

и они могут служить совершенно недвусмысленным свидетельством, как в пользу универсальности данной черты, так и в пользу адекватности перевода. Весьма впечатляющие параллельные оценки валидности получил Гау [19, 20]. Однако даже здесь возникает неопределенность, связанная с недостатками уницентрированной привязанности к американской версии теста. Так, например, хотя шкала «женственности» была предназначена для измерения определенной черты личности для каждого из полов, ее валидность исчисляется уровнем межполовых различий. Для американской версии этот уровень, определяемый бисериальным коэффициентом корреляции, равен 0,71. Для других стран соответствующие величины были значимыми, но более низкими: для Франции — 0,52, Италии — 0,47, Норвегии — 0,62, Турции — 0,47, Венесуэлы — 0,58. Если бы этот тест был децентрирован (то есть если бы американская версия рассматривалась как столь же нуждающаяся в редактировании, как и все остальные, если бы вопросы предполагали межполовые различия в интересах, наблюдаемые в равной мере в каждой стране, если бы статистические операции по выбору вопросов осуществлялись в едином ключе для всех стран и т. д. и если бы проводилась полицентрированная редакция с помощью обратного перевода), то полученный с его помощью результат мог бы означать, что наибольшие межполовые различия в интересах имели место в США, а наименьшие — в Италии и Турции. Однако наши антропологические знания наводят на мысль, что дело обстоит как раз наоборот, так что значение 0,71 отражает просто-напросто культурное смещение теста: будучи построен в расчете на максимизацию межполовых различий в США, в других культурах он, естественно, справлялся с этой задачей менее успешно.

Если методика разработки децентрированного теста включает в себя такие процедуры, как факторный анализ или выбор вопросов, эти последние должны проводиться в обоих языках, и для вычисления факторной оценки должны использоваться только те вопросы,

которые в обеих культурах предполагают одни и те же факторы (Gordon, Kikuchi [18], Kikuchi, Gordon [26]). И снова на пути к осуществлению этой программы стоит нежелание пересмотреть американскую версию.

312

Уницентрические смещения могут принимать крайне изоцированные формы и с трудом поддаваться искоренению. Гордон (см., например, Gordon, Kikuchi [18]), тонко воспринимающий эти проблемы и указавший на измерения, которые остаются в стороне, когда исследования ценностей имеют своим отправным пунктом только американский вопросник («ни один из опубликованных американских тестов не может измерить очень важную японскую ценность — «гири», или «сыновнюю почтительность»), в одной из более поздних работ [17] допустил ошибку: пользуясь переводом вопросов Q-сортинга, рассчитанного на максимизацию индивидуальных различий между американцами, он сделал вывод, что личности китайцев фактически более однородны, чем личности американцев. До тех пор пока он не осуществит противоположный процесс, то есть не разработает в Китае вопросы Q-сортинга, чтобы максимизировать различия между китайцами, и не переведет их затем для предъявления в США, подобные выводы ничем не оправданы.

Особые проблемы возникают в связи со специфическими характеристиками тестов, о которых, вероятно, можно говорить как о проблемах перевода разве что метафорически: они иллюстрируют собою фундаментальные проблемы из области межкультурной эквивалентности. Шкалы личностных предпочтений Эдвардса, подобно многим более хорошим тестам личности, построены так, чтобы удержать респондента от соблазна льстить самому себе, заставляя его выбирать между двумя или тремя утверждениями, выбранными для оценки социальной желательности. Конструируя японскую версию теста, Беррин [5], собрав переведенные вопросы, включил их в шкалы социальной желательности в Японии, а затем воспользовался этими совсем иными ценностями, чтобы по-новому объединить свои вопросы в пары. Так, несомненно, и следовало поступить, если бы речь шла о тесте для установления различий между японцами. Однако с точки зрения разработки теста для сравнения США и Японии это был неверный шаг: безукоризненно выполненная процедура перевода должна была бы привести к уравниванию различий в ценностях или потребностях между культурами (должны были бы совпасть средние значения

313

каждой ценности или потребности и внутри каждой из культур [26]). Фактами, дающими ключ к описанию культурных различий личности, были у Беррина его промежуточные данные, свидетельствовавшие о дифференциальной оценке японцами и американцами социальной желательности одних и тех же ответов. Точно так же шкалы, предложенные Кэнтрилом [11] и создающие, по-видимому, дополнительные возможности для межкультурного сравнения, без сомнения, полезны с точки зрения сравнительного исследования внутри одной страны, но, поскольку в них берется в качестве основания для сравнения локальный уровень адаптации и локальная система отсчета, в них не выявляются ясно различия, которые могли бы быть здесь обнаружены в противном случае.

В последние годы подчеркивается важность невербальных индикаторов социальной установки (см., например, Webb et al. [46]) и применения в межкультурном исследовании невербальных задач как одного из способов избежать проблем перевода, связанных с вербальным материалом (Frijda, Jahoda [16]). Указываются, в частности, преимущества задач, которые могут быть предъявлены посредством жеста. Разрабатываемые с целью избежать чрезмерной зависимости от вербальных методов, эти методики, обеспечивая перекрестную валидизацию с помощью альтернативных средств, заслуживают всяческой поддержки. Невербальные методы и методы наблюдения имеют изъяны иного рода,

нежели те, которые присущи вербальным процедурам, и это делает возможной известную методологическую триангуляцию, которая отфильтровывает нерелевантные, систематически возникающие погрешности, связанные с каждым отдельным методом (Campbell, Fiske [10]; Webb et al. [46]). Следует, однако, отказаться от какой бы то ни было защиты невербальных задач и наблюдений в роли заменителей (а не дополнений) вербальных методов, поскольку эти невербальные процедуры имеют свои собственные нерелевантные и дезориентирующие компоненты, учитывать которые столь же важно, как и в случае вербальных процедур [46].

Следует, прежде всего, отказаться от мысли, что невербальные задачи и инструкции посредством жеста не

314

требуют перевода. Жесты действительно часто используются для простейших коммуникаций людьми, говорящими на разных языках; тем не менее, жест сам культурно-специфичен: так, члены некоторых этнических групп Новой Гвинеи указывают на предметы глазами и не понимают указания, когда оно делается рукой; болгары заявляют о согласии, качая головой из стороны в сторону, движением, которым мы пользуемся для выражения отрицания; тибетцы выражают одобрение, высовывая язык, и т. д. Невербальные средства исследования нуждаются в столь же полном переводе, что и вербальный материал, но искусство перевода в этом случае совсем еще не развито. Вопросы невербального интеллектуального теста в целом зависят от культуры ничуть не меньше, чем вербальные. Обычно они обладают более высокой культурной специфичностью и уницентрированностью. Содержание используемых в тесте личности ТАТ картинок, к которым придумываются рассказы, в очень большой мере обусловлено культурой. Попытки исправить это положение путем использования новых картинок, соответствующих специфике данной культуры, ничуть не приблизили к решению проблемы сравнительного исследования, так как при этом остается возможность, что различия в ответах вызваны различиями между картинками (Lindzey [31]; Doob [12]). В конце концов, прожективный тест личности с вербальной инструкцией «Расскажи мне историю про маленького мальчика и его мать» поддается переводу лучше, чем методика ТАТ, стимулирующая рассказы при помощи картинок. (Этот простой пример помогает нам осознать те трудности, которые ожидают нас на этом пути.)

## ЛИТЕРАТУРА

1. Almond G., Verba S. The civic culture: political attitudes and democracy in five nations. Princeton, N. J., Princeton Univ. Press, 1963.
2. Anderson R. B. W. On the comparability of meaningful stimuli in cross-cultural research. *Sociometry*, 1967, 30, 124—136.
3. Varoux M. Experience in the Time International Survey: techniques used in France. *Public Opinion Quarterly*, 1949, 12, 715—718.

315

4. Bentz V. I. A comparison of the Spanish and English versions of the Sears Executive Battery. Chicago, National Personnel Department, Sears, Roebuck and Company. 1955 (Mimeo).
5. Berrien F. K. Japanese and American values. *International J. of PsychoL*, 1966, 1, 129—141.
6. Berrien F. K. Methodological and related problems in cross-cultural research. *International J. of Psychology*, 1967, 2, 33—43.
7. Bohannan P. J. Translation: a problem in anthropology. *The Listener*, 1954, May 13, 815—816.

8. Campbell D. T. Distinguishing differences of perception from failures of communication in cross-cultural studies. — In: F. S. C. Northrop, H. H. Livingston (eds.). *Cross-cultural understanding: epistemology in anthropology*. N. Y., Harper and Row, 1964.
9. Campbell D. T. Pattern matching as an essential in distal knowing. — In: K. R. Hammond (ed.). *Egon Brunswik's psychology*. N. Y., Holt, Rinehart and Winston, 1966.
10. Campbell D. T., Fiske D. W. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol Bull.*, 1964, 56, 81—105.
11. Cantril H. *The pattern of human concerns*. New Brunswick, N. J., Rutgers Univ. Press, 1965.
12. Doob L. W. *Psychology*. — In: R. Lystad (ed.). *The African world: a survey of social research*. N. Y., Praeger, 1965.
13. Duijker H. C. J. Comparative research in social science with special reference to attitude research. *International Social Science Bulletin*, 1955, 7, 555—556.
14. Ervin Susan M. Language and TAT content in bilinguals. *J. of Abn. and Soc. Psychol*, 1964, 68, 500—507.
15. Ervin S., Bower R. T. Translation problems in international surveys. *Public Opinion Quarterly*, 1953, 16, 595—604.
16. Frida N., Jahoda G. On the scope and methods of cross-cultural research. *International J. of Psychology*, 1966, 1, 109—127.
17. Gordon L. V. Q-typing of Oriental and American youth: initial and clarifying studies. *J. of Soc. Psychol*, 1967, 71, 185—195.
18. Gordon L. V., Kikuchi A. American personality tests in cross-cultural research — a caution. *J. of Soc. Psychol*, 1966, 69, 179—183.
19. Gough H. G. A cross-cultural analysis of the CPI Femininity Scale. *J. of Consulting Psychol*, 1966, 30, 136—141.
20. Gough H. G. An appraisal of social maturity by means of the CPI. *J. of Abnormal Psychol*, 1966, 71, 189—195.
21. Herskovits M. J. *Man and his works*. N. Y., Knopf, 1949.
22. Hyman H. World surveys — the Japanese angle. *International J. of Opinion and Attitude Research*, 1947, 1, 18—30.
23. Jacobson E. H. Methods used for producing comparable data in the OCSR: Seven-Nation Attitude Study. *J. of Soc. Iss.*, 1954, 10, 40—51.
24. Jacobson E. H. Cross-cultural contributions to attitude research. *Public Opinion Quarterly*, 1960, 24, 205—223.
25. Kennedy D. A., Lighton A. M. Pilot study of cultural items. — In: *Medical diagnosis, a field report*. USPMS Division of Indian Health, Health Education Branch, 1957.

26. Kikuchi A., Gordon L. V. Evaluation and cross-cultural application of a Japanese form of the Survey of Interpersonal Values. *J. of Soc. Psychol*, 1966, 69, 185—195.
27. Kluckhohn F. H., Strodtbeck F. *Variations in value orientations*. Evanston, Ill., Row, Peterson, 1961.
28. Lambert W. E. Measurement of the linguistic dominance of bilinguals. *J. of Abn. and Soc. Psychol*, 1955, 50, 197—200.
29. Lambert W. E., Havelka J., Crosby C. The influence acquisition contexts on bilingualism. *J. of Abn. and Soc. Psychol*, 1958, 56, 239—244.
30. Levine R. A., Campbell D. T., *Ethnocentrism field manual* — In: Levine R. A., Campbell D. T. *Ethnocentrism*. N. Y., Wiley, 1972, p. 249—296.
31. Lindzey G. *Projective techniques and cross-cultural research*. N. Y., Appleton-Century-Crofts, 1961.
32. Lonner W. J. The SVIB visits German, Austrian and Swiss psychologists. *Amer. Psychologist*, 1968, 23, 164—179.
33. Mitchell R. E. Survey materials collected in the developing countries: sampling measurement, and interviewing obstacles to intra-and international comparisons. *International Social Science Journal*, 1965, 17, 677.
34. Nida E. Linguistics and ethnology in translation. *Word*, 1945, I, 194—208.

35. Nida E. Toward a scientific theory of translating. N. Y., J. J. Brill., 1964.
36. Osgood E. E. Studies on the generality of affective meaning systems. *Amer. Psychologist*, 1962, 17, 10—28.
37. Osgood E. E., Sebeok T. A. (eds.). Psycholinguistics: a survey of theory and research problems. A Morton Prince Memorial Supplement to the *Journal of Abnormal Social Psychology*, 1954, 49.
38. Phillips H. P. Problems of translating and meaning in field work. *Human Organization*, 1959, 18, 184—192.
39. Quine W. V. Word and object. N. Y., Wiley, 1960.
40. Schachter S. Interpretative and methodological problems of replicated research. *J. of Soc. Iss.*, 1954, 10, 52—60.
41. Segall M. H., Campbell D. T., Herskovits M. J. The influence of culture on visual perception. Indianapolis, Bobbs-Merrill, 1966.
42. Stern E., D'Epina y R. L. Some polling experiences in Switzerland. *Public Opinion Quarterly*, 1948, 11, 553—557.
43. Voegelin C. F. Multiple stage translation. *International J. of American Linguistics*, 1954, 20, 271—280.
44. Voegelin C. F. Anthropological linguistics and translation. To honor Roman Jakobson. Paris, Mouton, 1967, p. 2159—2190.
45. Voegelin C. F., Florence M. Voegelin. Selection in Hopi ethics, linguistics, and translation. *Anthropological Linguistics*, 1960, 2, 48—77.
46. Webb E. J., Campbell D. T., Schwartz R. D., Sechrest L. B. Unobtrusive measures: nonreactive research in the social sciences. Chicago, Rand McNally, 1966.

**ИЗМЕРЕНИЕ ВОЗДЕЙСТВИЙ,  
ОКАЗЫВАЕМЫХ ЗАПЛАНИРОВАННЫМ  
СОЦИАЛЬНЫМ ИЗМЕНЕНИЕМ<sup>1</sup>**

Одна из отличительных особенностей всех современных обществ состоит в том, что мы сознательно выбираем и составляем проекты, которые призваны усовершенствовать наши социальные институты. Наше общее затруднение заключается в том, что наши проекты не всегда дают ожидаемые эффекты. Вполне вероятно, на практике все согласится, что часто мы не можем определить, оказал ли вообще данный проект какое-либо воздействие, — настолько сложен поток исторических изменений, идущих своим чередом, и так велико число других проектов, от которых можно ждать изменений тех же самых показателей.

Эта общая совокупность проблем в сочетании с явной релевантностью процедур социального исследования должна, по-видимому, неизбежно вести к появлению в большинстве стран особой отрасли методологии, а также специалистов-методологов, которые сосредоточивают свое внимание на проблеме измерения воздействий, оказываемых запланированным социальным изменением. Настоящая работа основывается на допущении, что эта методология в значительной своей части может быть с пользой для дела принята всеми, несмотря на различия в формах правления, а также в под-

<sup>1</sup> Campbell D. T. Assessing the impact of planned social change. — In: Social Research and Public Policies. Ed. by G. M. Lyons. Hanover, New Hampshire, Univ. Press of New England, 1975. © University Press of New

England, 1975. Один вариант работы был представлен на конференцию по социальной психологии в Вышереграде (Венгрия), 1974 г. Публикуется с сокращениями.

318

ходах к социальному планированию и выработке решений. Методология оценки эффективности социальных проектов является одной из тех областей науки, которые обладают достаточной универсальностью, чтобы сделать обмен научным опытом взаимовыгодным. Как часть этого обмена, настоящая работа информирует о методологии измерения эффективности программ в том ее варианте, который разрабатывается сейчас в США.

Наиболее распространенным в США наименованием этой развивающейся области исследования является *«исследование по оценке эффективности»*, что в настоящее время почти всегда предполагает *«оценивание программ»* (хотя термин «оценивание» уже давно употребляется для обозначения процедур определения пригодности тех или иных лиц для выполнения тех или иных социальных ролей). В этой области уже имеется ряд публикаций — сборников и пособий (Suchman [87]; Caro [28]; Weiss [94, 95]; Rivlin [74]; Rossi, Williams [77]; Glaser [43]; Fairweather [39]; Wholey et al. [100, 101]; Caporaso, Roos [27]; Riecken, Boruch, Campbell, Caplan, Glennan, Pratt, Rees, Williams [73]). Издаются также журнал «Evaluation» и журнал, посвященный оценке программ в сфере образования, — «Evaluation Comment: The Journal of Educational Evaluations». Ряд других журналов, например «Social Science Research», «Law and Society Review» и др., публикует данные об исследовательских методах в этой области.

Специалисты приходят в эту область из самых разных социальных дисциплин. Широко представлены экономисты. Вносят свой вклад специалисты в области исследований операций и других сфер «научного управления». В работе участвуют статистики, социологи, психологи, специалисты в области политических наук, исследователи в сфере социального обслуживания, педагогики. То, что их работы и рекомендации, в конечном счете, сходны, свидетельствует о стремительном становлении новой самостоятельной дисциплины, которая, отделившись от своих столь различных «родителей», может вскоре обрести свою собственную индивидуальность.

Поскольку сам я пришел в эту область из социальной психологии, я, хотя и рассматриваю то, над чем работаю сейчас, скорее как «прикладную социальную

319

науку», нежели как собственно социальную психологию, испытываю тем не менее потребность прокомментировать тот особый вклад, который может внести социальная психология в рассматриваемую область. Во-первых, из всех дисциплин, связанных с интересующей нас сферой исследований, одна лишь психология характеризуется ориентацией на лабораторный эксперимент, и именно социальные психологи располагают наибольшим опытом распространения лабораторных экспериментальных моделей на социальные ситуации. С тех пор как для уменьшения неопределенности того, «что чем вызвано», в исследованиях по оцениванию программ появляется модель экспериментальной науки (см. основополагающую работу Шухмана [87]), это наиболее значительный вклад, как в отношении общей ориентации, так и в отношении специфических умений.

Во-вторых, психологи лучше всего подготовлены к работе с критическими и аналитическими понятиями в области измерений. Психология благодаря давней традиции построения и критического анализа своих собственных мер разработала понятия и математические модели надежности и валидности, в которых очень нуждается практика оценивания программ, даже, несмотря на то, что эти меры, вероятно, все еще непригодны для изучения некоторых специальных психологических проблем. Понятие отклонения, или «погрешности», в то виде, в каком оно разработано старой психофизикой при различении

«постоянной ошибки» (bias) и «переменной ошибки» (ненадежности), а также более поздние труды по измерению личностных переменных и установок, включая измерение установок на ответ, галоэффектов, факторов социальной желательности, индексной корреляции, факторных методов и т. д. (Cronbach [33, 34]; Edwards [38]; Jackson, Messick [54]; Campbell, Siegman, Rees [25]; Campbell, Fiske [24]), также очень важны и не находят отражения в понятии валидности, если оно определяется как коэффициент корреляции с некоторым критерием. Конечно, это не является нашей монополией. В самом деле, кто, как не представители качественной социологии, исследует условия, при которых складывается оружие социальная статистика (см., например, Becker et al. [6, 7]; Douglas [37]; Garfinkel [42]; Kitsuse, Cicourel [59]; Beck [5]), от кого, как не

320

от них, исходит необходимый скептицизм по поводу таких мер, как данные о частоте самоубийств и преступлений? Однако даже здесь именно у психологов оказалось достаточно опыта, чтобы различить степени валидности, располагающиеся между полной незначимостью и высшим совершенством, и именно у психологов явилось желание использовать, хотя бы критически, те меры, которые, как они знали, были отчасти ненадежны и ошибочны.

В-третьих, многие методологические проблемы осуществления социальных проектов и измерения их воздействий связаны с изучением социально-психологических аспектов взаимодействия между гражданами и проектами, или между гражданами и способами экспериментального воплощения проектов (рандомизация, контрольные группы), или между гражданами и специальными измерительными процедурами, используемыми при оценивании. Это специальные проблемы формирования социальных установок и их влияния на ответы, явно находящиеся в сфере нашей компетенции.

После этих замечаний о профессиональных сторонах исследований по оценке эффективности программ в США я хотел бы перейти к проблемам, с которыми мы сталкивались до сих пор, и к решениям, которые нами предлагались. Замечу с сожалением, что со времени моей более ранней работы (Campbell [16]) мы ушли не слишком далеко; однако я попытаюсь привести новые иллюстрации.

Последующее изложение сосредоточено, таким образом, на трудностях и проблемах, о которых я считаю необходимым предупредить. Если мы претендуем на то, чтобы быть *методологами*, мы тем самым намерены быть экспертами по этим проблемам, и, хотелось бы надеяться, теми, кто вырабатывает решения. Не было бы нужды в такой специальности, если бы не существовали эти проблемы. С этой точки зрения не нужно никаких оправданий. Но мне хотелось бы привлечь новых участников в новую профессию и вдохновить их на то, чтобы они посвятили свои усилия делу, которое может принести плоды только с большой отсрочкой. На потенциальных новых участников, а также и на тех, кто уже занялся этим делом вплотную, полный перечень стоящих перед нами трудностей, включая проблему такого

321

использования наших умений, которое у нас самих не вызвало бы сожалений, может подействовать только расхолаживающе. Мы пока не можем предложить полный перечень профессиональных умений, гарантирующих необходимую эффективность. В нескольких случаях успешного осуществления немногих выгодных программ, оцененных однозначно, общество вполне обходится (или могло бы обойтись) без нашей помощи. Мы все еще не можем привести примеров значительных вкладов в общественные нововведения, которыми эти последние были бы обязаны нашим методологическим умениям. Потребность в нашей специальности и конкретные рекомендации, исходящие от нас, все еще должны быть

обоснованы скорее обещаниями, чем уже достигнутыми результатами. Наши рекомендации априорны в том смысле, что они представляют собой экстраполяции в новом контексте, не прошедшие, однако, в этом контексте перекрестной валидации. Что касается меня, то я склонен считать проблему проверки социальных образований реальностью настолько важной, что наши усилия и профессиональные обязанности полностью оправдываются обещанием. Я полагаю, что проблемы неоднозначности свидетельств в пользу эффективности программ настолько близки общим проблемам научного вывода, что указанные экстраполяции в разработке рекомендаций относительно процедур оценивания программ при наличии соответствующей взаимной критики могут быть хорошо обоснованы. Как бы то ни было, исходя отчасти из представления о том, что слишком многочисленные обещания оказываются одним из основных препятствий для осмысленного оценивания программ, я стремлюсь, однако, с двойственных позиций дать честную пессимистическую картину этих проблем.

Второй круг проблем связан с тем, что множество методологических трудностей неизбежно создается взаимодействием различных аспектов политической ситуации, образующих контекст социальных программ и их оценивания.

Именно в сфере методологических проблем, вызванных к жизни политическими соображениями, теряют силу исходные посылки об универсальности наших методологических принципов, едва только мы сопоста-

322

вим практику резко отличающихся друг от друга социальных, экономических и политических систем. Ученым, собирающимся на международные конференции, большей частью удастся избежать разобщающих политических вопросов путем концентрации внимания на научных задачах, которые объединяют их как ученых. Когда темой является измерение воздействия, оказываемого запланированными социальными изменениями, мы не можем позволить себе такую роскошь. И все же я не теряю надежды на разработку такой технологии, которая принесет пользу любой политической системе. Я полагаю, что методология оценивания программ будет во многом независима от содержания самой программы. В пользу этого свидетельствует подчеркивание роли специалистов в области социальных наук, когда дело касается их помощи обществу при выявлении эффектов, вызванных происходящими в этом обществе политическими процессами, и преуменьшение роли этих специалистов в разработке новых программ. Возможна ли такая независимость от идеологии, и если да, то, как совместима она с долгом социального ученого способствовать развитию более аутентичного человеческого сознания и созданию более человеческих форм общественной жизни, — вот вопросы, которые я пока что не рассматривал должным образом, не говоря уже об их решении.

В последующих разделах я сгруппировал наши проблемы под тремя общими заголовками, однако не очень заботился о том, чтобы строго придерживаться этой рубрикации при обсуждении. Вначале идут специальные вопросы нашего научного сообщества: их следовало бы поставить даже в том случае, если бы ученые, оценивающие программы, стремились лишь к одноплановому оцениванию программ. Это «Метанаучные проблемы» и «Статистические проблемы». Последний раздел включает взаимодействие с общественным контекстом. В разделе «Проблемы политической системы» я обращаюсь к вопросам, которые касаются преимущественно политических процессов и правительственных институтов. Некоторые из этих вопросов являются, вероятно, общими для всех крупных бюрократических обществ, а другие — специфическими для условий США.

323

*Количественная и качественная методология.* Противоположность «качественной» и «количественной» форм знания, *geisteswissenschaftlich* (науки о духе) и *naturwissenschaftlich* (науки о природе), «гуманистического» и «сциентистского» подходов характерна сегодня для большинства социальных наук в США. В таких областях, как социология и социальная психология, многие из наших самых талантливых и преданных науке исследователей все больше склоняются к качественному, гуманистическому подходу. В политической науке также существует разделение по этому признаку. Кажется, только экономика и география относительно свободны от этого разделения.

Этот раскол проник, конечно, и в исследования по оценке эффективности, приняв здесь форму спора по поводу законности количественно-экспериментальной парадигмы при оценивании программ (см., например, Weiss, Rein [96, 97]; Guttentag [48, 49, 50]; Campbell [17, 19]). Разумеется, спор по данному вопросу ведется в несколько иных терминах. Критики, занимающие позицию, которую я называю гуманистической, довольно часто сведуци в количественных экспериментальных методах. Их специфические критические замечания нередко хорошо обоснованы именно с позиций экспериментатора: эксперименты, связанные с использованием некоей частной процедуры в каких-то частных условиях, совершенно неопределенны в плане установления того, что чем вызвано; системам измерения свойственна ничем не обоснованная жесткость, ограничивающая регистрируемые выходные данные заранее ожидавшимися параметрами; экспериментальная программа, сосредоточенная на общем эффекте некоего комплексного воздействия, часто игнорирует сам процесс, поэтому знание таких эффектов имеет сомнительное значение для повторного осуществления или улучшения программы; программы широкого масштаба часто бывают безнадежно неопределенны во всем, что касается целей и релевантных индикаторов; изменения программы воздействия в ходе какого-нибудь улучшающего эксперимента, хотя и имеют решающее значение в практическом отношении, делают невозможным истолкование соотношения входа

и выхода; осуществление социальных программ часто обеспечивается несостоятельными с точки зрения экспериментальных моделей способами; даже в тщательно контролируемых ситуациях экспериментирование является делом крайне утомительным и неопределенным; экспериментирование — слишком медленный процесс, чтобы быть полезным в политическом плане и т. д. Все эти соображения соответствуют истине и подчас достаточно справедливы для того, чтобы побудить к энергичному поиску альтернатив. Однако предлагавшиеся до сих пор альтернативы качественного знания (см., например, Weiss, Rein, [96, 97]; Guttentag [48, 49, 50]) для меня неубедительны.

Я полагаю, что натуралистическое наблюдение за событиями является, в сущности, неопределенным основанием для каузальных выводов, с помощью каких бы средств — качественных или количественных — такое наблюдение ни проводилось, ибо имеет место смешивание фактора отбора и воздействий. Всякая попытка уменьшить неопределенность будет приводить к тому, что условия будут становиться более «экспериментальными». «Эксперименты» — это фактически и есть особый тип специально разработанных условий наблюдения, оптимальный для каузального вывода. Проблемы вывода, связанные с оцениванием программ, присущи самой природе программ, осуществляемых в ходе социальных процессов. Экспериментальные планы не являются источником этих проблем и фактически облегчают их решение, хотя часто лишь в небольшой степени.

В таких возражениях часто содержится неявный протест против подмены качественного знания косвенными и гипотетическими процессами научного познания. Но хотя я вынужден отвергнуть этот аспект гуманистической критики, существуют иные аспекты,

побуждающие к такой критике, к которым я могу всей душой присоединиться. Эти критические положения можно обозначить как «пренебрежение релевантными качественными контекстуальными данными» или «чрезмерная зависимость от небольшого числа квантифицированных абстракций, ведущая к игнорированию противоречащих им и дополняющих их качественных данных».

Под влиянием миссионеров из лагеря логического позитивизма представители количественной социальной

325

науки слишком часто исходят из предположения, что в истинной науке качественное обыденное знание заменяется знанием количественным. В действительности дело обстоит совсем иначе. Скорее, наука зависит от обыденного качественного знания, даже если она в лучшем случае и выходит за его пределы.

Чтобы подвергнуть сомнению обыденное понимание какой-то частности и внести в него поправки, следует доверять обыденному пониманию в целом.

С этим связано также подчеркивание в эпистемологии качественного распознавания целостного образа как исходного по отношению к распознаванию атомарных частиц, могущих быть квантифицированными, в противоположность все еще широко распространенной логической атомистической интуиции (Campbell [13]). Такая эпистемология скорее фаллибилист<sup>1</sup>, чем провидец: она делает акцент на предположении о том, что при подобном распознавании образов допускаются ошибки, а не на признании способности восприятия служить надежной основой достоверности. Но она также видит в этом подверженном ошибкам, интуитивном, гипотетическом обыденном восприятии единственный возможный путь. Дело не в том, что перцептивные данные выше всякой критики (Campbell [15]), но они доступны критике лишь благодаря доверию к другим многочисленным перцептивным данным того же гностического уровня.

Если мы применим такую эпистемологию для оценки эффективности программ, это сразу узаконит «описательную» часть большинства отчетов и подскажет, что такая деятельность получает формальное признание в планировании и проведении исследований. Без этих материалов исследования по оценке эффективности недоступны интерпретации и в большинстве своем интерпретировались бы лучше, если бы этих материалов было больше. Тот факт, что содержание их субъективно и грешит ошибками, должен привести нас к более тщательному подбору лиц, привлекаемых для регистрации событий, и к подготовке формализованных процедур, с помощью которых все заинтересованные участ-

---

<sup>1</sup> Фаллибилизм (от лат. fallere) — воззрение, основывающееся на утверждении подверженности познания разного рода ошибкам. — *Прим. перев.*

326

ники могли бы внести дополнения и исправления в официальную версию. Следует предусмотреть использование профессионально подготовленных историков, антропологов и специалистов в области качественной социологии. Без описательной части не обходится ни один заключительный отчет, и для ее подготовки должны использоваться самые совершенные качественные методы.

Следует также осознать, что как участники разработки и внедрения планируемых изменений, так и наблюдатели на протяжении веков не прибегали для оценки их ни к квантификации, ни к научному методу. Наши научные данные должны основываться на обыденном знании, научное знание должно превзойти, но не заменить обыденное знание. Однако при количественных оцениваниях это обстоятельство обычно игнорируется, разве что к делу приобщается несколько случайно подобранных эпизодов. Согласно отстаиваемой мной эпистемологии, исследователю следует попытаться систематически

учитывать все критические замечания и оценки с позиций здравого смысла в адрес программы, исходящие от клиентов, их семей и членов общины. Наряду с количественными процедурами, такими, как вопросники и оценочные шкалы, которые из соображений удобства будут вводиться на этой стадии исследования в целях сбора и суммирования данных, должны быть предусмотрены и неколичественные методы сбора данных — например, работа с иерархически организованными дискуссионными группами. Там, где подобные оценки приходят в противоречие с количественными результатами, последние следует брать под сомнение, пока не будут поняты причины расхождения. Конечно, ни те, ни другие методы не непогрешимы. Но вот что следовало бы запомнить многим из нас: количественные результаты могут быть столь же мало застрахованы от ошибок, как и качественные. В конце концов, измерительные приборы в физических лабораториях часто работают неадекватно, и неисправность обнаруживается лишь благодаря качественному знанию и предположениям по поводу того, что должен был бы показывать этот прибор. (Это весьма далеко от мифа о том, что метрические показатели операционально определяют теоретические параметры.)

327

Приходится, к сожалению, констатировать, что в американских работах по оценке эффективности программ такое тонкое сочетание способов познания пока что не практикуется. Вместо этого наблюдается переход от одной крайности к другой. Когда для оценивания используются наблюдения антропологов, как это имеет место, например, при оценках моделей городов, они скорее замещают собой, чем дополняют количественные показатели, предусматривающие предварительное и итоговое тестирование и сравнения с контрольными группами. Актуальный пример — использование антропологов в программе «Экспериментальные школы». В одном случае предполагалось участие группы из пяти антропологов, в течение пяти лет занимавшейся изучением системы школьного образования для города с населением в 100 тысяч человек. Эти антропологи не имели ранее опыта изучения какой-либо другой системы школьного образования в США. Им не предоставили времени для специальной подготовки до введения программы, вместо этого они приступили к работе уже после того, как программа была начата. Их планы не предусматривали изучение какой-либо иной сравнимой системы школьного образования, не претерпевшей соответствующего изменения. Для того чтобы поверить, что в таких неблагоприятных для наблюдения условиях эти наблюдатели, использующие качественный подход, смогут прийти к заключению о том, какими именно характеристиками наблюдаемые ими процессы обязаны нововведениям, предусмотренным данной программой, нужно больше веры, чем есть у меня, хотя, пожалуй, следовало бы воздержаться от суждений до ознакомления с результатами. Более того, в исследовании делается упор на непосредственные наблюдения самих антропологов, а не на их роль в отборе и использовании в качестве информантов тех лиц, которые сами вовлечены в изменения, осуществляемые по данной программе<sup>1</sup>. В результате —

---

<sup>1</sup> Подробнее о методе использования информантов в полевых исследованиях см. в следующей работе: S a m p b e l l D. T. and L e V i n e R. A. Field — Manual Anthropology. — In: A Handbook of Method in Cultural Anthropology. Ed. by R. Naroll and R. Cohen, Garden City, N. Y., The Natural History Press, 1970, p. 366—387.— *Прим. ред.*

328

тенденция игнорировать данные других наблюдателей, придерживающихся качественного подхода и находящихся в более благоприятных условиях, чем эти антропологи. К таким наблюдателям относятся родители, имевшие других детей, которые обучались в школе до ее преобразования, учителя, наблюдавшие именно эту систему до, во время и после

преобразования, учителя, у которых до прихода в эту школу уже имелся опыт работы в школах, сравнимых с новой по каким-то другим параметрам, и сами учащиеся. Желательно, вероятно, располагать большим массивом таких наблюдений, полученных с помощью вопросников. Поэтому естественным было бы и желание задавать соответствующие вопросы до осуществления экспериментальной программы и, кроме того, провести этот опрос в какой-нибудь школьной системе, не подвергшейся такой реформе, восстановив, таким образом, экспериментальный план и количественные обобщения качественных оценок. (Подробное обсуждение вопроса о качественном и количественном подходах читатель может найти в работе: Campbell [20] <sup>1</sup>.)

Хотя вопрос о количественной и качественной ориентациях имеет большое практическое значение, он, тем не менее, как мне кажется, является, прежде всего, вопросом, входящим в компетенцию специалистов в области социальных наук и относительно независимым от более глобального политического процесса. То, какого подхода мы придерживаемся, во многом зависит от того, из какого подразделения в сообществе социальных исследователей затребована соответствующая рекомендация, что в свою очередь отчасти объясняется разочарованием в ранее применявшейся модели. Иными словами, решение данного вопроса принадлежит нам.

Прочие вопросы, относящиеся к числу метанаучных, больше связаны с вненаучными вопросами о человеческой природе, социальных системах и политическом процессе. Здесь я лишь классифицирую их, поскольку считаю, что первый шаг в их решении должен был бы состоять в достижении взаимопонимания между методологами — специалистами по оцениванию программ, а единство взглядов предполагает согласие не столько

---

<sup>1</sup> См. перевод этой работы в данной книге.

по поводу деталей метода, сколько по поводу метанаучных вопросов.

*Осуществление и оценивание программ.* Одна из стратегий, утвердившихся в правительственных учреждениях США, наиболее тесно связанных с оцениванием программ, состоит в организационном отделении процесса осуществления программ от их оценивания. Эта рекомендация исходит от академической группы теоретиков научного управления, она получила распространение в конце 60-х годов в правительственных кругах как «Система программирования, планирования и финансирования», или «PPBS» <sup>1</sup>, согласно которой эти функции, как и функции надзора за программой или ее оценивания, должны были быть поручены отдельной организационной единице, не зависящей от учреждений, решающих практические задачи. (Соответствующее изложение этой стратегии принадлежит Уильямсу и Эвансу [103].) Эта рекомендация основывается на теории организационного контроля, известной как принцип проверок и балансов. Она подтверждается не только общими наблюдениями, свидетельствующими о нежелании человека заниматься самокритикой, но особенно наблюдениями из области давно принятой в США самоохранительной практики, при которой на отчеты о достижениях и другие оценки программ в первую очередь с необходимостью возлагается задача оправдать бюджет следующего года. Как полагает типичный администратор программы совершенствования в США, будь это экспериментальная программа или одна из давно принятых программ, бюджеты должны получать непрерывное оправдание. Для подобного администратора оценки программы вряд ли могут быть оторваны от этой непрекращающейся отчаянной борьбы. В такой ситуации имеется глубочайший смысл в том, чтобы функцию оценивания программ осуществляло отдельное подразделение, которому никакие бюджетные ограничения не мешают давать честные оценки. В этом отношении такая стратегия не вызывает возражений.

Мои собственные наблюдения, однако, приводят меня к выводу, что и эта стратегия не работает. Рассма-

---

<sup>1</sup> См. о системе «PPBS» в работе: Г в и ш а н и Д. М. Организация и управление. М., «Наука», изд. 2-е, 1972, с. 462—471. — *Прим. ред.*

триваемое разделение направлено против тех форм осуществления программ, которые оптимизировали бы интерпретацию данных оценивания. Такие формы существуют, и они к тому же дешевы, но они требуют усовершенствованного планирования и тесной кооперации исполнителя и оценщика. Внешние оценщики, помимо всего прочего, лишены, как правило, фундаментального качественного знания о происходящем. Постоянный конфликт между оценщиками и исполнителями, который примет достаточно острые формы при общем локальном руководстве, имеет тенденцию к углублению. Пригодность мер для целей локальных программ и опасности особенно уменьшаются в условиях принятых в США процедур выполнения договорных исследований. Оценивание становится деморализующим фактором и источником отвлекающего конфликта. Можно было бы надеяться, что специализация способствовала бы притоку более опытных методологов. Если и есть такой выигрыш, он полностью уничтожается недостаточным экспериментальным контролем.

Конечно, поднятые проблемы обязаны своим происхождением не только разграничению осуществления и оценивания. И доводы в пользу такого разграничения остаются весьма вескими. Тем не менее, эти проблемы вызывают беспокойство, и они достаточно взаимосвязаны, чтобы служить обоснованием для пересмотра этого принципа, особенно если учесть, что такое разграничение, видимо, полностью отсутствует в экспериментальной науке. Возникает метанаучный вопрос о том, каким образом в науке, несмотря на погрешности исследователя, достигается объективность, а также об уместности этой модели для объективности в оценивании программ.

Обычно в науке тот, кто задумывает эксперимент, сам же регистрирует показания прибора. Соответственно существует проблема пристрастной мотивации. Исследователь почти неизбежно является пристрастным защитником какого-то одного результата опыта. Появляются неопределенности в интерпретации. Под угрозой оказывается честное имя и дело всей жизни. Не все допущенные ошибки исправляются перед публикацией, причем ошибки, которые подтверждают гипотезу автора, выявить гораздо труднее. Загадка о том, как наука обретает свою объективность (если только она ее

обретает), представляет собой метанаучный вопрос, который все еще не решен. Хотя ученые, по-видимому, честнее, осторожнее и самокритичней большинства других социальных групп, эти их качества являются скорее побочным продуктом социальных сил, порождающих научную объективность, чем ее источником. Главными факторами служат, вероятно, традиция и возможность независимого повторения. Хотя философы и социологи науки лучше проясняют эти вопросы, методологи исследований по оценке эффективности программ должны проявлять внимание к возможностям моделей, применимых в их области. Забегая несколько вперед, отмечу, что я прихожу к следующей предварительной точке зрения.

Осуществление и оценивание программ совершенствования в США требуют сегодня большего энтузиазма, преданности и более высоких моральных качеств. Они могут быть усилены путем принятия в качестве модели экспериментатора-оценщика образа ученого. Если бы могли быть созданы условия для повторной перекрестной валидации, а бюджет мог быть избавлен от угроз, проистекающих от негативных оценок (например, если бы

исполнители программы могли получить право переходить для достижения той же самой цели к альтернативным программам), то от политики разграничения осуществления и оценивания можно было бы отказаться.

Наш вопрос не требует сугубо однозначного ответа. Внешние оценки могут сочетаться с внутренними оценками. Вероятно, внешнее оценивание можно было бы вполне безболезненно свести к проверке внутренних оценочных данных, поручив его кому-нибудь из методологов по оцениванию.

*Максимизация повторения и критики.* Продолжая разговор на ту же метанаучную тему, которая была затронута в предыдущем разделе, можно привести еще целый ряд рекомендаций относительно исследовательской стратегии. Некоторые из них противоречат ортодоксальным взглядам и практике, распространенным в США сегодня.

В настоящее время предпочтение отдается единичным, согласованным в масштабе страны оценкам даже в том случае, если нововведения по программам осуществляются во многих районах. Если бы задача состо-

332

яла в имитации научного подхода к объективной реальности, оптимальным было бы, вероятно, разделить большие эксперименты и общие оценки на части и заключить вместо одного два или более контрактов, преследующих общую цель, с тем чтобы получить какую-то возможность одновременного повторения. Компенсаторные образовательные программы (например, такие, как «Head Start», «Follow Through» и др.<sup>1</sup>) иногда принимают, согласно нашим основным оценкам, такой размах, что подобное разделение не способствовало бы значительному увеличению расходов. Мы часто могли бы предусматривать нашими планами сравнительную проверку при повторении, которая помогает науке оставаться объективной.

Положительным моментом исследований по оценке эффективности в США является широко распространенная защита и проведение от случая к случаю *повторного анализа* данных оценивания программ другими исследователями. «Фонд Рассела» финансирует целый ряд таких работ, в том числе оценку эффективности телевизионных образовательных программ для дошкольников (Cook et al. [31]). Первоначальную правительственную оценку (Cicirelli [29]) программы компенсаторного дошкольного образования повторно проанализировали Смит и Биссел [85] и Барноу [3]. Ведется дальнейшая работа в этом направлении. Так же обстоит дело и в отношении некоторых других классических исследований по оценке эффективности, хотя они все еще не получили широкого распространения и многие данные недоступны.

Одно из изменений, в которых нуждаются традиции исследования, или его этика — это отношение к «показаниям меньшинства» из состава исследователей. Этическое правило, согласно которому данные оценивания должны быть доступны для критического анализа, следует распространить и на сотрудников, которые проводили сбор данных и их анализ и которые очень часто прекрасно знают, как можно организовать эти данные, чтобы они подтверждали совсем иные выводы, а не те,

---

<sup>1</sup> Здесь и далее автор использует принятые в США условные обозначения специальных образовательных программ, призванных компенсировать отставание в развитии и образовании, в частности, детей из плохо обеспеченных семей. — *Прим. ред.*

333

которые представлены в официальном отчете. В настоящее время любые действия подобного рода рассматривались бы как достойная осуждения нелояльность по отношению к организации. Из-за этого оказывается утраченным один из особенно компетентных источников критики, а тем самым и объективности. Официальное предложение

подготовить отчет, полученное каждым членом группы специалистов по оценке эффективности от учредителя или администратора, в значительной мере способствовало бы в этом отношении освобождению, как от вины, так и от цензурных ограничений.

В этой связи нам следует иметь в виду две ценные модели социального экспериментирования. С одной стороны, модель большой науки, образец которой мы находили в эксперименте «Негативный подоходный налог», он будет обсуждаться ниже. С другой стороны, модель недорогого «административного эксперимента» (Campbell [14], Thompson [89]), который состоит в том, что какая-нибудь административная единица, например город или штат (или предприятие, или школа), применяет новую стратегию таким образом, чтобы иметь возможность экспериментальной или квазиэкспериментальной проверки ее эффективности. «Практическое оценивание программ для местного управленческого персонала и персонала штатов» (Hatry et al. [51]) служит примером этого подхода. Чтобы административное экспериментирование давало объективное знание, нужна *диффузная перекрестная валидизация*, при которой города, штаты, и т. п., приняв обещанные нововведения, подтверждают их эффективность посредством своей собственной деятельности по оцениванию.

Децентрализация принятия решений обладает тем преимуществом, что она создает большее число социальных единиц, способных повторять и проверять социальные преобразования или одновременно изучать множество оценок различных альтернативных решений. Даже в отсутствие планирования само существование в США правления штатов создает возможность квази-экспериментальных сравнений, что было бы недоступно в более интегрированной системе. Эту точку зрения хорошо обосновал Цейзель [105, 106], а примером может служить исследование Болдуса [2], о котором ниже говорится более подробно. Если предприятия,

школы или другие подобные им по размерам единицы могут осуществлять независимый выбор программ и если заимствуемые программы оцениваются столь же тщательно, как новые, цепная реакция заимствований самых многообещающих программ обеспечивала бы что-то наподобие валидизации науки.

*Исследования по оцениванию — скорее нормальная, нежели экстраординарная наука.* До сих пор, рассматривая метанаучные вопросы, мы не обращались к острым метанаучным проблемам, обсуждающимся в настоящее время в США. Из них в центре внимания оказываются все еще те вопросы, которые были подняты в книге Т. Куна «Структура научных революций» [60]. Хотя я больше, чем Кун, подчеркиваю последовательность и относительную объективность науки, я нахожу в том, что он говорит, немало ценного и имеющего отношение к нашей проблеме. А именно: в научном развитии бывают нормальные периоды, когда существует общее согласие в отношении правил, с помощью которых решается, какая теория более валидна. В развитии науки имеют место и экстраординарные, или революционные, периоды, когда ученые должны делать свой выбор на основе правил, которые не являются элементами прежней парадигмы. После такой революции выбор новой преобладающей теории первоначально не обосновывается в терминах правил решения, принятых в предшествующий период нормальной науки.

В случае исследований по оценке эффективности куновская метафора «революция» может быть возвращена в политическую сферу. Ясно, что исследования по оценке эффективности осуществляются правительством или по крайней мере не вызывают у него возражений. Это предполагает наличие устойчивой социальной системы, характеризующейся социальными показателями, которые остаются сравнительно постоянными, так что ими можно пользоваться для измерения воздействия, оказанного программой. Осуществленные программы должны быть достаточно малы по объему, чтобы не вызвать серьезного нарушения в общей социальной системе. Обсуждаемая мной

технология непригодна для измерения социальных воздействий какой-либо революции. Даже в рамках стабильной политической ситуации ее приложения следует, возможно, ограничить сравнительно

второстепенными нововведениями. Это показывает, например, Цейзель, рассматривая экспериментирование с правовой системой США. (Само собой разумеется, я не собираюсь строить на этом валидную аргументацию против осуществления изменений, масштабы которых делают невозможным их оценивание.)

### Статистические проблемы

В этом разделе я займусь более земными проблемами, в которых мы, методологи количественного оценивания, чувствуем себя наиболее уверенно. Это проблемы, явно требующие профессионального умения, проблемы, которые нуждаются в решении и могут быть решены. Такие статистические проблемы предполагают при решении метанаучных вопросов предпочтение количественного экспериментального подхода. Я начну с полезного обыденного метода — с прерывных временных серий. Затем я остановлюсь на некоторых популярных, но неприемлемых регрессионных подходах к квазиэкспериментальным планам. Потом будут обсуждены проблемы, связанные с экспериментами, предполагающими рандомизацию, а в заключение — оригинальный компромиссный план<sup>1</sup>.

*План прерывных временных серий.* Этим термином я обозначаю широко распространенную практику построения на основе той или иной социальной статистики временных рядов и попыток их интерпретировать. Эта практика, как и проблемы, с которыми она столкнулась, и решения, к которым она привела, независимо возникли во многих странах. Я отмечу некоторые примеры использования этого метода вне США: Kutchinsky [61]; David, Wright [36]; Glass, Willson, Gottman [44] и др.

В работе Росса [75] приводятся статистические дан-

---

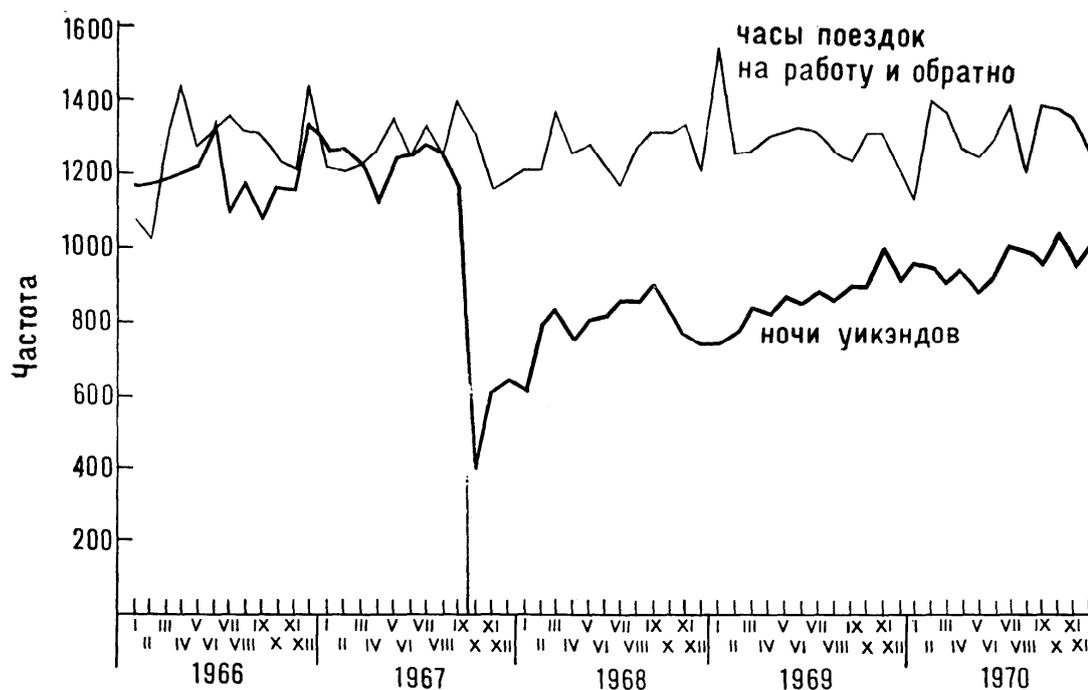
<sup>1</sup> Рассматриваемые в данном разделе модели и планы квазиэкспериментальных исследований, а также методологические и методические вопросы, связанные с разработкой и использованием этих моделей, более детально изложены в специальных работах Д. Т. Кэмпбелла (см., например: Campbell D. T. *Quasi-Experimental Designs*. — In: *Social Experimentation: A method for planning and evaluating social intervention*. N. Y., Academic Press, 1974; Campbell D. T. *Reforms as experiments*. — In: *Readings in Evaluation Research*. Ed. by F. G. Caro, N. Y., Russell Sage Foundation, 1977; Campbell D. T. *Focal local indicators for social program evaluation*. — *Social indicators research*, 1976, vol. 3. — *Прим. ред.*

ные об эффекте особых мер по обеспечению безопасности движения, принятых в Англии в 1967 г.<sup>1</sup> На приведенном графике (рис. 1) этот эффект получил более наглядное отражение, чем в какой бы то ни было английской публикации. Британское министерство транспорта на протяжении последующего года добросовестно обнародовало впечатляющие результаты. В отчете приводились проценты сокращения числа дорожных происшествий в данном месяце по сравнению с тем же месяцем предыдущего года. Это лучше, чем полное игнорирование зависимости результатов от времени года, но и этот метод неэффективен, поскольку необычные «эффекты» часто в той же мере обязаны своим происхождением превратностям предшествующего периода времени, что и текущего периода. Кроме того, этот метод не позволяет создать общую картину. Газеты с полным основанием отмечали успех мероприятия, однако интерес к нему вскоре угас, и в настоящее время английские социологи в большинстве своем не имеют представления об эффективности этой программы. На рис. 1 данные приводятся с поправкой на сезонные тенденции, на неодинаковое

число дней и уикэндв в разных месяцах и с учетом того факта, что меры оценивались лишь после даты их введения. Всякая такая подгонка (adjustment) по-своему проблематична и может быть решена по-разному. В данном случае последствия принятых мер были настолько значительны, что их можно было бы выявить при любом подходе, но во многих случаях дело будет обстоять иначе. (Отмеченные на графике данные о часах поездок на работу и обратно служат контрольными по отношению к вечерним и ночным часам уикэндв.)

Болдус [2] проанализировал серьезные последствия одного закона, который он считает вредным

<sup>1</sup> Изучение эффективности принятых мер предполагало оценку причин уменьшения числа дорожных происшествий, вызванных водителями в нетрезвом виде. Более детально исследование и методика описаны в работах: R o s s H. L., C a m p b e l l D. T., G l a s s G. V. Determining the social effects of legal reform: The British «breathalyser» crackdown of 1967. — *Amer. Behavioral Scientist*, 1970, vol. 13, NO 4. C a m p b e l l D. T., R o s s H. L. The Connecticut crackdown on speeding, *Time — Series Data in Quasi-Experimental Analysis*. — *Law and Society Review*, 1968, vol. III, № 1; C o o k T. D., C a m p b e l l D. T. *Op. cit.* — *Прим. ред.*



Р и с. 1. Дорожные происшествия (со смертельным исходом или с тяжелыми телесными повреждениями) в Англии до и после введения программы по обеспечению безопасности дорожного движения в октябре 1967 г. (упорядоченные посезонно). (По: Ross [75].)

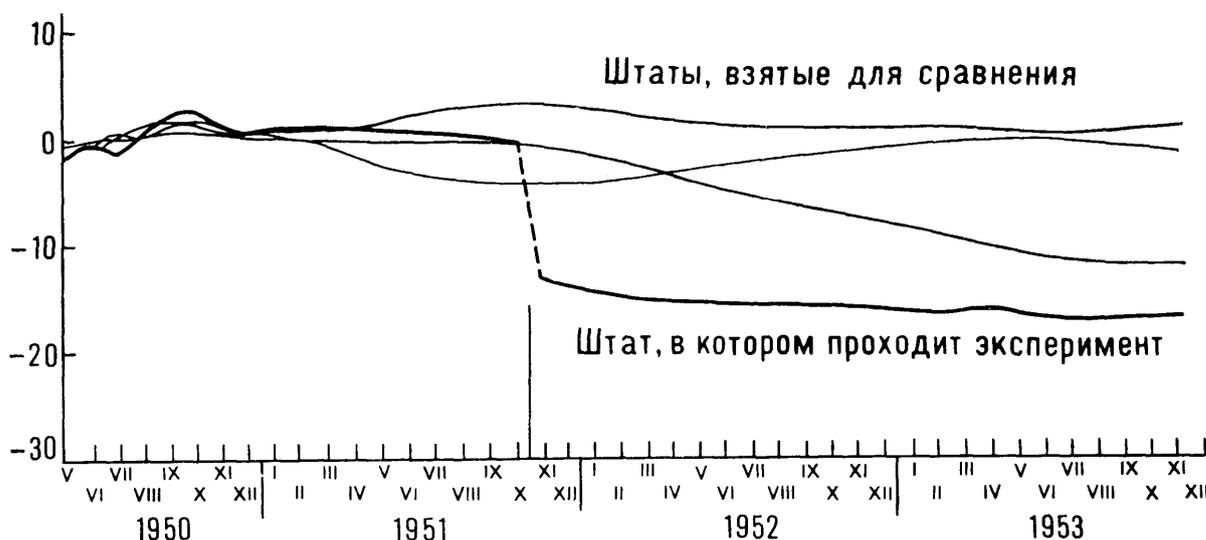
именно потому, что он эффективен. Если лицо, получавшее пособие по старости (вспомоществование для бедных, назначаемое правительством), умирая, оставляет деньги или какую-нибудь собственность, то, согласно этому закону, расходы правительства должны быть возмещены. При капиталистической идеологии, усвоенной даже бедными, многие старики будут умирать голодной смертью, лишь бы иметь возможность оставить свое жилище детям. Болдус изучал последствия таких законов на материале примерно 40 случаев, когда они вводились в штатах, и примерно 40 других случаев, когда их действие

приостанавливалось. В каждом случае он отыскивал для сравнения какие-то похожие штаты, которые не меняли своих законов (см. рис. 2).

План прерывных временных серий имеет исключительно важное значение для оценивания программ. К нему можно обращаться, когда новая программа оказывает влияние на всех, вследствие чего нельзя составить настоящую контрольную группу. Если есть возможность привлечь для сравнения данные какой-ни-

338

будь другой группы, этот план лучше любого другого квазиэкспериментального плана (Campbell, Stanley [26]<sup>1</sup>). Нередко он может быть реконструирован на основе архивных материалов. Будучи представлен графически, он легко доступен пониманию администраторов и законодателей. Вот почему он вполне заслуживает методического совершенствования. Далее приводится краткий перечень методологических проблем, с которыми мы сталкиваемся при разработке этого метода.



Р и с. 2. Влияние введения в штате А закона, требующего возмещения получаемой помощи после смерти получателя, на поступление в учреждения социального обеспечения запросов о пособиях престарелым. (Модификация по: Boldus [2, с. 204].) Ежемесячные данные, где все значения выражены в процентах от количества запросов за 18 месяцев до изменения в законе.

1. Все еще остается проблемой проверка значимости. Обычный метод наименьших квадратов здесь, как правило, неприменим вследствие авторегрессионной ошибки, поэтому наиболее приемлемой представляется модель «подвижных средних». Гласс, Уиллсон и Готтмен [44] разработали самый лучший подход к проблеме, который основывается на трудах Бокса и Тьяо [11] и Бокса и Дженкинса [12]. Эти модели требуют, чтобы в данных отсутствовали систематические циклы, однако все методы их устранения, как правило, не полностью обеспечивают подгонку данных. Для них требуется, кроме того, большое число временных точек, и иной

<sup>1</sup> См. перевод этой работы в данной книге.

339

раз не удастся выявить эффект, который воспринимается визуально, когда он представлен графически. Но иногда с их помощью выявляется значимое воздействие там, где визуально нельзя ничего обнаружить.

2. Проблемой остается и устранение сезонных тенденций. Последние сами по себе нестабильны и предполагают обращение к модели подвижных средних. Изменение из месяца в месяц, совпадающее с предусмотренной программой, не может считаться чисто сезонным; ряд, следовательно, должен быть разделен в данной точке на части для оценки характера сезонных изменений. Поэтому точки, непосредственно предшествующие осуществлению программы и непосредственно следующие за ним, становятся крайними точками ряда и поддаются коррекции хуже, чем точки в середине ряда (Керка [55]; McCain [66]).

3. Новая администрация, которая предлагает новые программы, склонна вносить изменения и в систему регистрации процесса. Тем самым изменения в показателях часто становятся не поддающимися интерпретации (Campbell [16]). Этого следует по возможности избегать.

4. Когда введенные программы являются реакцией на какую-нибудь острую проблему (например, резкое изменение к худшему, фиксируемое каким-нибудь социальным показателем), благотворные воздействия программы смешиваются с «регрессионными артефактами», которые обязаны своим возникновением тому факту, что в нестабильном ряду точки, следующие за экстремальным отклонением, располагаются, как правило, ближе к линии общей тенденции (Campbell [16, 413—414]).

5. С помощью этого плана обычно невозможно уловить изменения, вводимые постепенно. Если администратор, пользуясь этим планом, хочет сделать программу оптимально доступной оцениванию, не следует вводить ее до тех пор, пока не будут сделаны все необходимые приготовления для резкого ввода. Британская программа мер по обеспечению безопасности движения, о которой мы говорили выше, может служить примером этой оптимальной практики (см. рис. 1).

6. Ввиду того, что необходимы длинные серии наблюдений, мы, как правило, вынуждены пользоваться

только теми индикаторами, которые регистрировались когда-то для иных целей. И хотя часто они релевантны (например, данные о рождении и смерти), тем не менее, даже самые что ни на есть намеренно выбираемые индикаторы никогда полностью не релевантны, это — серьезное ограничение. Отсутствуют, в частности, отчеты о переживаниях участников и их опыте. В то же время невозможно и нежелательно пытаться предвидеть все будущие потребности и заводить книги для их учета. Необходим известный компромисс между крайностями, пусть даже ценой умножения бланков, подлежащих заполнению, и записей, которые надо регулярно вести. На институциональном уровне было бы важно получить от каждого участника «Годовые отчеты для оценивания программы» (Gordon, Campbell [46]). В системе школьного образования такие отчеты могли бы представлять учителя, учащиеся и родители. Отметим, что в настоящее время школьная система фиксирует успехи учеников, но еще ни разу не интересовались отчетами учеников об успехах школы. Учителя из года в год оцениваются по эффективности их работы, но им ни разу еще не представился случай оценить стратегию, которую они должны проводить в жизнь. Изучаются некоторые шаги в этом направлении (Weber, Cook, Campbell [93]; Anderson [1]). В системе социального обеспечения в США такие отчеты могли бы представлять как социальные работники, так и те, кто получает помощь (Gordon, Campbell [46]). Все показатели должны оценивать лишь программы и стратегии, но не конкретных лиц — по причинам, которые будут рассмотрены ниже.

*Корректирование регрессий как замена рандомизации.* Наиболее распространенная в американской практике модель оценивания состоит во введении новой программы в каком-то одном институте или административной единице с обследованием до, и после ее

введения. Хотя при этом оставляет желать лучшего способ контроля, этот метод часто оказывается достаточно информативным для того, чтобы его стоило практиковать. Почти столь же часто этот план усовершенствуется за счет привлечения сравниваемой группы, которая также подвергается измерениям до и после введения программы в основной группе. Это, как пра-

341

вило, какая-то иная, свободная от воздействий социальная единица, которая не получает новой программы и признается сопоставимой в других отношениях. Обычно оказывается, что эти две группы отличаются друг от друга уже до работы с ними, и одна из естественных тенденций состоит в том, чтобы устранить эти отличия. На мой взгляд, средства, используемые с этой целью в практике применения статистических методов в США, почти всегда неверны. Получается, что некий набор статистических инструментов, разработанных и пригодных для предсказания, применяется в целях каузального вывода, которым он не соответствует. Регрессионный анализ, мультивариативная статистика, ковариационный анализ — вот наименования некоторых статистических приемов, которые я имею в виду. Идет ли речь о статистике в области образования или экономики, подбор методов остается, по-видимому, тем же. У экономистов есть выражение, которым они обозначают данную проблему, — «ошибка в переменных», или, точнее, «ошибка в независимых переменных». Но, учитывая эту проблему в теории, они так привыкли считать свои индикаторы, по существу, свободными от ошибки, что на практике ее игнорируют. Они забывают, что нерелевантные систематические компоненты дисперсии создают ту же проблему, что и случайная ошибка, ведущая к такой же погрешности недостаточной коррекции. Отметим, что наличие ошибки и единичная дисперсия имеют систематический эффект, то есть служат скорее источником погрешности, чем источником нестабильности оценок. Этот факт также игнорируется экономистами и другими исследователями. Таким образом, усилия, направленные на то, чтобы внести поправку на отличия, имевшие место до воздействия, с помощью «регрессионных коррекций» средних, либо путем «отделения» отличий, установленных при предварительном тестировании, либо путем ковариационных коррекций, приводят к недостаточной коррекции, если только показатели предварительного тестирования (или другие сопутствующие переменные) не являются совершенными мерами того, что оказывается общим для показателей предварительного и итогового тестирований. Хорошо известно, что старая методика использования только тех случаев, для которых имеются соответствия показателей, выявленных в предвари-

342

тельном и итоговом тестировании, порождает «регрессионные артефакты» (Thorndike [90]; Campbell, Stanley [26]). Ковариация дает ту же погрешность, ту же степень недостаточной коррекции, но более прецизионную (Lord [64, 65]; Porter [71]; Campbell, Erlebacher [23]), и точно так же обстоит дело со множественной регрессией и частной корреляцией (см., например, Cook, Campbell [32]). В сущности, та же проблема возникает в исследованиях *ex post facto*, где хотя и не бывает предварительного тестирования, но коррекции поддаются другие сопутствующие переменные. Широко распространенный вариант данной проблемы имеет место в случае, когда некоторые лица подвергаются какому-то воздействию, а наряду с ними имеется обширная популяция лиц, не подвергающихся этому воздействию, из которых составляется «контрольная» группа для сравнения.

В практике США стало крайне важно различать два вида ситуаций, в которых используется этот тип квазиэкспериментального плана и эти типы коррекции, так как недостаточная коррекция в этих ситуациях имеет противоположный социальный смысл. С

одной стороны, имеются специальные программы, обеспечивающие благоприятные возможности — вроде университетского образования,— предназначенные для тех, кто менее нуждается, или, согласно более распространенной формулировке, кто больше всех их заслуживает, или для тех, кто скорее, чем другие, сумеет извлечь из них какую-то пользу. Будем называть их «дистрибутивными» программами в отличие от «компенсаторных» программ, то есть от тех, в которых благоприятные возможности предоставляются тем, кто более всего в них нуждается.

В случае регрессивных программ группа, выбранная для воздействия, будет, как правило, превосходить контрольную группу или ту популяцию, из которой выбираются квазиэкспериментальные контрольные группы. В этих условиях неизбежная недостаточная коррекция, вызванная единичной дисперсией и ошибкой в показателях предварительного тестирования и/или другой сопутствующей переменной («регрессионные артефакты»), способствует тому, чтобы воздействие казалось эффективным, тогда как на самом деле оно не имеет

343

смысла и ведет к преувеличению его эффективности во всех случаях. Многим из нас такая ошибка кажется вполне простительной, поскольку она укрепляет нашу веру в воздействия, в пользу которых мы убеждены всей душой. (Это может показаться удивительным, но американская телевизионная образовательная программа для дошкольников «Sesame Street» является «дистрибутивной» в том смысле, что ее больше смотрят дети более образованных родителей) (Cook et al. [31].)

В компенсаторных программах контрольные группы обычно, хотя и не всегда, вначале превосходят те группы, которые подвергаются воздействию, либо подбираются из какой-нибудь более обширной популяции, в среднем превосходящей обследуемую. В этих условиях погрешности недостаточной коррекции, регрессионные артефакты, ведут к недооценке эффективности программы и к тому, что наши программы кажутся вредными, в то время как они попросту бесполезны. Такие условия для квазиэкспериментальных исследований имеют место в наших важнейших работах по оцениванию компенсаторных образовательных программ, известных под названиями «Head Start», «Follow Through», «Performance Contracting», «Job Corps» и др. Можно считать это почти определенно установленным для значительных нежелательных эффектов, обнаруживаемых при оценивании короткой трехмесячной программы «Head Start», на которую отводится 10 часов в неделю (Cicirelli [29]; Campbell, Erlebacher [23]). Я убежден, что преобладание этих квазиэкспериментальных условий и процедур коррекции является одним из главных источников пессимистических отчетов о результатах подобных компенсаторных образовательных программ. Те весьма немногочисленные исследования в сфере компенсаторного образования, которые обходят эту проблему путем случайного распределения детей по экспериментальным и контрольным группам, дают более оптимистические результаты.

В области компенсаторного образования имеется еще ряд проблем, способствующих тому, чтобы в квазиэкспериментальных исследованиях программа выглядела как вредная. Они включают тесты, которые слишком сложны, дифференциальные показатели развития в сочетании с возрастными, эквивалентными школьной оцен-

344

ке, абсолютными, или необработанными, показателями, а также тот факт, что надежность теста выше для итоговых проверок, чем для предварительного тестирования, а надежность оценок контрольной группы превышает надежность оценок экспериментальной группы (Campbell [19]). Эти обстоятельства требуют коренного пересмотра нашей практики

тестовых оценок. Когда к одной популяции в каком-то единичном частном случае применяются разные модели оценивания, все процедуры оценивания так сильно коррелируют между собой, что можно с тем же успехом воспользоваться наипростейшей из них. Но когда две группы, с самого начала отличающиеся друг от друга, в период ускоренного развития подвергаются тестированию в два разных срока, наша обычная практика оценивания дает видимое увеличение разрыва, если, как это обычно бывает, повышается надежность теста. Использование коррекции для предсказаний становится крайне важным. Следует отказаться от общепринятой модели, основывающейся на допущении, что «истинная оценка» и «ошибка» не зависят друг от друга, и заменить ее моделью, которая при сравнении индивидуальных результатов выявляет отрицательную корреляцию между ошибкой и истинной оценкой: чем более значим компонент ошибки, тем менее значим компонент истинной оценки.

*Проблемы, связанные с рандомизированными экспериментами.* Достойным внимания примером хорошего социального эксперимента в США является «Эксперимент с негативным подоходным налогом» в Нью-Джерси (или Программа «НПН»)¹.

В этом эксперименте исследуется вопрос о гарантированном годовом доходе как альтернативе к действующим в США в настоящее время системам социального обеспечения. Своим наименованием он обязан тому, что, если доходы падают ниже некоторого заданного уровня, налог должен становиться негативным, иными словами, государство должно выплачивать какую-то

---

¹ «Эксперимент с негативным подоходным налогом» («New Jersey Negative Income Tax Experiment») описан в ряде работ: Watts, Rees [92]; Kershaw [56, 57, 58], см. также: Social Experimentation: A method for planning and evaluating social intervention. Ed. by H. W. Riecken and R. F. Boruch. N. Y., Academic Press, 1974. — *Прим. ред.*

сумму гражданам вместо того, чтобы взимать ее с них. Кроме того, предполагается изменить и процедуры налогообложения. В этом эксперименте 600 семей, где главой семьи является работающий мужчина, получали денежные пособия, благодаря которым их доход достигал некоторой суммы, от 3000 до 4000 долларов в год, на семью из четырех человек, эта сумма варьировалась в согласии с одним из восьми планов, причем каждый план предусматривал определенный уровень материальной помощи, а также стимулирование участников увеличивать свои собственные доходы. Другие 600 семейств не получали дополнительных доходов, но один раз в квартал у них брали интервью. Эксперимент продолжался три года, и сейчас уже имеются предварительные данные о его результатах. Я хотел бы использовать опыт этого социального эксперимента для того, чтобы осветить методологические проблемы, которые должны быть решены в более совершенных социальных экспериментах. В этом плане мои комментарии кажутся по преимуществу критическими. Однако мое общее отношение к этому эксперименту одобрительное. Я рассматриваю его как наиболее значительный пример практического использования прикладной социальной науки.

Основной результат этого эксперимента состоит в том, что гарантированный доход не ослабляет эффективных усилий неимущих людей в сфере труда. Этот итог, если его признать, опровергает главный аргумент, выдвигаемый против такой программы, поскольку, если исходить из одного только объема затрат, она должна обходиться дешевле, чем нынешняя система социального обеспечения, по крайней мере если она не побуждает многих ныне работающих людей оставить свою работу. В методологическом плане основная критика этого исследования связана с сомнением по поводу правдоподобия утверждений, что этот «лабораторный» результат имел бы место и после того, как данная программа вспомоществований стала в США обычной, постоянной политикой. Это вопросы «внешней валидности» (Campbell, Stanley [26]) или «конструктивной валидности», если использовать термин, первоначально разработанный для теории измерения Куком

(Cook, Campbell [32]). Особенно примечательны два критических замечания. Одно состоит в том, что предполагает-

346

ся, что имел место «эффект Хоуторна» или «эффект морской свинки»<sup>1</sup>. Семьи, участвовавшие в эксперименте, знали, что они являются избранными участниками мероприятия, проводимого с особой целью, и находятся в центре всеобщего внимания. Поэтому они и вели себя как подобает «хорошим» людям, трудолюбивым и респектабельным, что и привело к полученным результатам. Такая мотивация отсутствовала бы, если бы программа была общей для всех. Можно предположить, что этот момент усугублялся двумя обстоятельствами, сопутствовавшими реализации данной программы. Эксперимент с самого начала получил широкую известность и рекламу, включая телевизионные интервью с wybranными испытуемыми, и случайная выборка производилась по семьям, а не по жилым кварталам, так что каждая семья, участвовавшая в эксперименте, была окружена столь же неимущими соседями, на которых программа помощи не распространялась. Второе общее критическое замечание, особенно со стороны экономистов, можно было бы обозначить как эффект ограничения времени. Участникам эксперимента предлагалась материальная помощь ровно на три года. Им объяснили, что эксперимент ограничен этим периодом. В этих условиях предусмотрительные участники должны были держаться за свою работу, разве что они могли получить другую, получше, так что они должны были быть готовы вернуться к своим обычным финансовым трудностям.

Следует отметить, что эти две проблемы ни в коей мере не составляют специфику рандомизированных экспериментов и, более того, должны быть характерны для наиболее поверхностных пилотажных программ. Их можно избежать, лишь оценивая принятие такой программы в качестве общенациональной стратегии. Такое оценивание должно было бы быть по идее квазиэкспериментальным, как при применении временных серий, и, по-видимому, предполагать сравнение с не-

---

<sup>1</sup> Имеются в виду неоднократно описанные в социальной психологии хоуторнские эксперименты и в психологии «эффекты морской свинки», то есть влияние самого факта проведения эксперимента, внимания к испытуемым и наблюдения за ними на их поведение и результаты исследования. О хоуторнских экспериментах см., например: О'Шонессе и Дж. Принципы организации управления фирмой. М., 1979, с. 137—140. — *Прим. ред.*

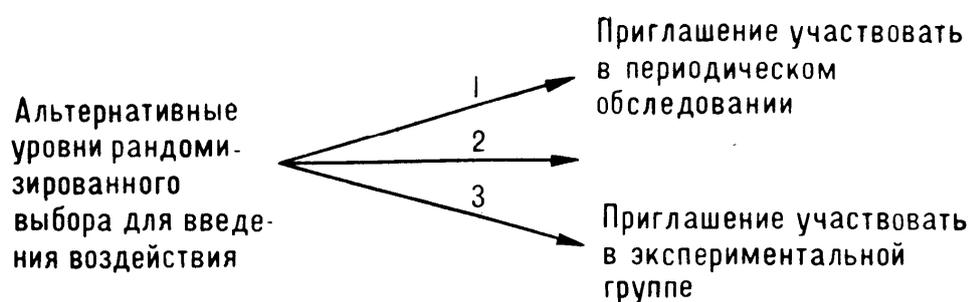
347

сколькими канадскими городами. Такие оценки должны быть более успешными в плане внешней и конструктивной валидности и менее — в плане внутренней валидности. Специфика наших национальных установок заставляет, однако, усомниться в том, что это квазиэкспериментальное оценивание будет проведено, как следует, если только оно вообще будет проведено: уж раз мы избираем какую-то стратегию, мы теряем интерес к ее оценке. Если бы программа привела к снижению трудовых усилий, принятие ее в качестве стратегии в масштабах страны было бы крайне маловероятным. Уже поэтому стоило бы проводить ее оценку, и проводить хорошо.

Детали этого эксперимента указывают на ряд проблем, касающихся метода, которые нуждаются в пристальном внимании со стороны творчески мыслящих статистиков и социальных психологов. Мы здесь лишь отметим их, более подробно они рассматриваются в других работах (Riecken et al. [73]). Вопрос о единице, подлежащей рандомизации, уже поднимался ранее. Часто существует возможность рандомизации более крупных социальных единиц, чем индивиды или семьи, нередко могут быть использованы жилые кварталы, школьные классы, школы и т. д. С точки зрения статистической эффективности

предпочтителен выбор более мелких и более многочисленных единиц, максимизирующий число степеней свободы и действенность рандомизации. Однако использование более крупных единиц часто увеличивает конструктивную валидность. С выбором для осуществления рандомизации уровня, на котором проводится набор респондентов, тесно связаны проблемы, которые вызваны потерями, обусловленными в свою очередь отказами и более поздним отсевом. Чтобы выявить приемлемых участников программы, обращались к статистическим данным переписи в районах бедноты и к приемам формирования опросной выборки. Переосмысление задачи, стоящей перед создателями программы, показывает целесообразность различения двух типов согласия, требуемого для участия в эксперименте, — согласия на участие в обследовании (измерении) и на участие в мероприятии (воздействии); имеются, таким образом, два независимых друг от друга уровня, на которых может быть отказ. При этом

348



Р и с. 3. Три альтернативных уровня, на которых может быть осуществлена рандомизация.

возникают три критические альтернативные точки, в которых могла быть осуществлена рандомизация (см. рис. 3). В программе была использована первая альтернатива. Впоследствии респондентам из контрольной группы было предложено участвовать в измерении, а испытуемым экспериментальной группы — как в измерении, так и в получении помощи (воздействии). Возможно, что в результате в экспериментальной группе оказались лица, которые не стали бы утруждать себя участием в процедурах, если бы их случайно пригласили в контрольную группу. Приглашение участвовать в программе, осуществляемое поэтапно, и проведение рандомизации среди тех, кто согласился на обследование-опрос (то есть на условия контрольной группы), могли бы обеспечить сопоставимость групп. В ходе осуществления программы имели место отказы от экспериментального воздействия, вызванные нежеланием принимать «подавание». Тем самым создается неэквивалентность, однако погрешности, возникающие из-за такого дифференциального отказа, могут быть измерены в том случае, если те, кто отказывается от воздействия, продолжают все же участвовать в обследовании, связанном с измерением. Именно эту вторую альтернативу мы и могли бы здесь рекомендовать.

Можно рассмотреть и дальнейшее отсрочивание рандомизации — третью альтернативу. В этих случаях всех потенциальных участников знакомят с описанием каждого из экспериментальных условий и с их шансами оказаться в этих условиях. Затем следует выяснить согласие каждого на участие в эксперименте — независимо от того, какая функция выпадет на его долю. Рандомизированный выбор следует производить из тех, кто согласился на все условия. Эту альтернативу ждет все более широкое применение. Возможность дифферен-

циального отказа сведена здесь к минимуму (хотя кое-кто все же откажется, когда узнает, какой жребий выпал на его долю). Тем самым доводится, по-видимому, до максимума уровень «осведомленного согласия», который национальные институты здоровья США считают обязательным условием для всех финансируемых ими исследований в области медицины и поведенческих дисциплин. Тем не менее, Комитет по разработке экспериментов как метода планирования и оценки эффективности социальных программ при Совете социальных исследований США (Riesken et al. [73]) в своих рекомендациях не использовал эту третью альтернативу. Члены этого комитета считали, что достигнуто согласие, когда участник полностью осведомлен о предстоящем воздействии. Сообщение участникам контрольной группы о благах, которые предназначены другим и чуть было не достались им самим, могло бы вызвать у них чувство недовольства и превратить контрольное воздействие скорее в особый опыт, чем просто в ситуацию отсутствия воздействия. Это могло бы усилить склонность испытуемых, включенных в контрольную группу, к более частым, чем у испытуемых экспериментальной группы, отказам, и могло бы оказаться, что данный прием повышения сопоставимости групп, в конечном счете, производит обратное действие. Так что имеются веские аргументы и за, и против. Эта проблема нуждается в дальнейшем обсуждении и исследовании.

*Отсев*, и в частности *дифференциальный отсев*, испытуемых становится центральной проблемой, для разработки которой необходимы исследования изобретательных статистиков. В программе «НПН» величина отсева за трехлетний период колебалась от 25,3% в контрольной группе до всего лишь 6,5% в той из экспериментальных групп, которая получала наибольшее вознаграждение. Эти различия достаточно велики, чтобы привести к псевдоэффектам в оценках при итоговом тестировании. Наличие данных предварительного тестирования дает нам некоторую информацию о направлении смещения, однако ковариация, вычисленная на основе полученных оценок, недостаточно установлена и не может, таким образом, служить адекватной коррекцией. Необходимо разработать методы исключения максимальных и минимальных смещений при разного

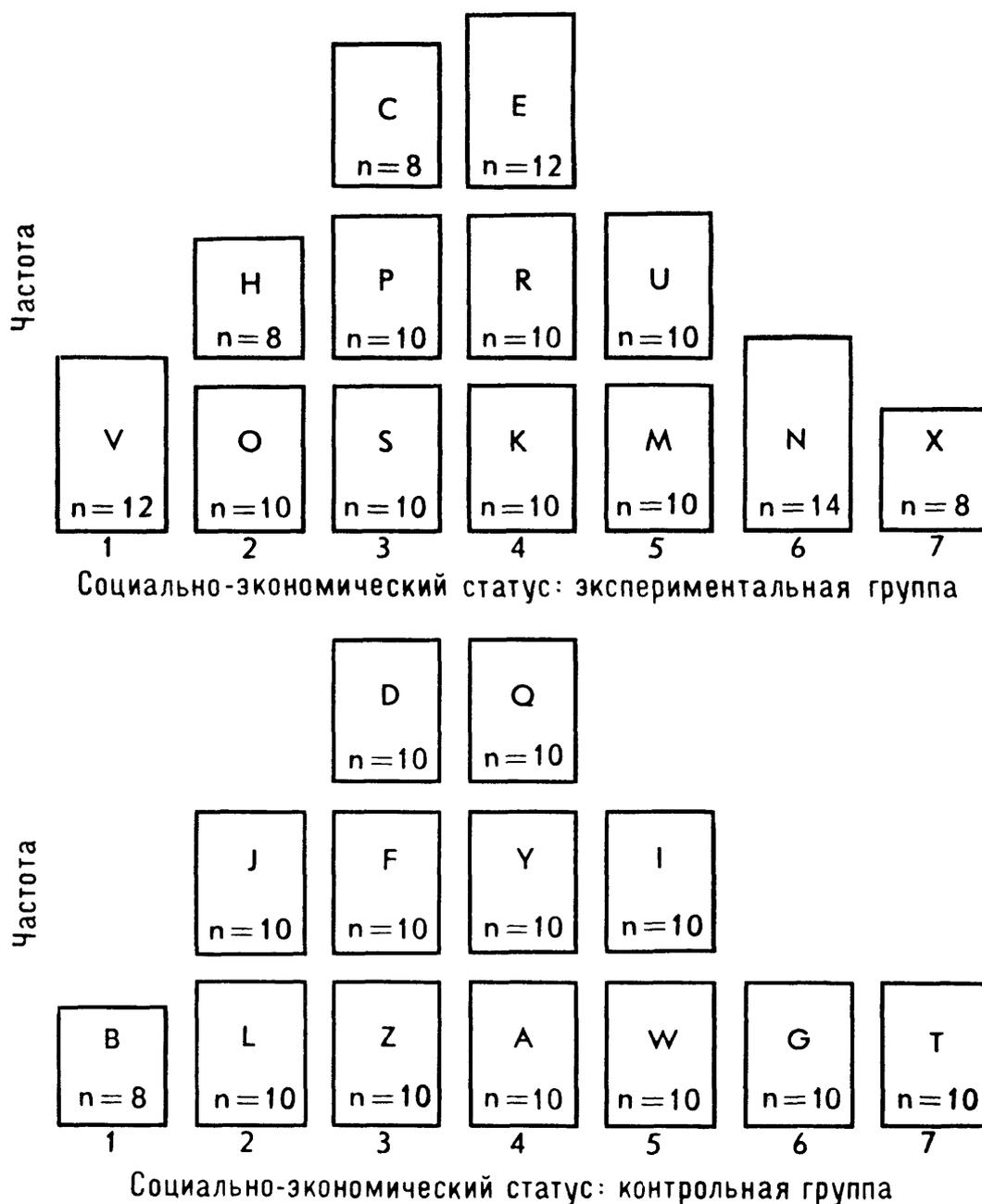
рода специфических допущениях. Там, где имеется система периодических обследований широкого охвата, которая все еще включает людей, отказавшихся от дальнейшего участия в данном эксперименте, существуют другие альтернативы, ждущие своей разработки.

Такие системы измерения, соответствующие программе «НПН», могли бы включать данные Управления социального обеспечения о доходах, подлежащих налогообложению на нужды социального обеспечения, об обращениях за пособием по безработице, данные финансового управления об удержании налогов, сведения о предоставлении больничных услуг и т. д. Такие данные время от времени используются в исследованиях по оценке эффективности (см., например: Levenson, McDill [62]; Bauman, David, Miller [4]; Fischer [40]; Heller [52]), однако возможности для осуществления такой деятельности не получили соответствующего развития, и может показаться, что идея подобного использования данных идет вразрез с существующим сегодня в США требованием сохранения конфиденциальности в отношении административных данных (см., например: Reubhausen, Brim [78]; Sawyer, Schechter [79]; Goslin [47]; Miller [67]; Westin [98]; Wheeler [99]). Поскольку открытый доступ исследователя к административным документам мог бы дать столь ценную возможность снизить стоимость контроля за нововведениями, связанными с программами, в настоящей работе стоило бы обсудить и этот вопрос.

Существует один способ статистического соотношения данных исследования и административных данных без предания гласности конфиденциальной информации об индивидах (Schwartz, Orleans [80]; Campbell, Boruch, Schwartz, Stenberg [22]; Boruch,

Campbell [10]). Обозначим этот способ как «взаимно изолированный обмен документами». Он требует, чтобы последовательный ряд административных документов заключал в себе возможность внутреннего статистического анализа содержащихся в нем данных. Не вдаваясь в детали <sup>1</sup>,

<sup>1</sup> Методологические положения и процедуры такого исследования приводятся в работе: Campbell D. T., Boruch R. F., Schwartz R. D., Steinberg J. Confidentiality-preserving modes of useful access to files and to interfile exchange for useful statistical analysis. — *Evaluation Quarterly*, 1977, vol. 1, № 2, p. 269—300.



Р и с. 4. Гипотетические данные двух групп, участвующих в эксперименте. По социально-экономическому статусу они разбиты на подгруппы и обозначены символами от A до Z.

я попытался бы, тем не менее, передать суть этой идеи. На рис. 4 приводятся данные гипотетического эксперимента с участием одной экспериментальной и одной контрольной группы. В этих условиях мы располагаем достаточным числом случаев, чтобы их можно было подвергнуть дальнейшему дроблению на социально-экономическом уровне. На основе этих данных подготовлено 26 списков, включающих от 8 до 14 человек. Каждому списку случайным образом присваивается ка-

352

кое-нибудь обозначение (в данном случае буквы от *A* до *Z*), которое не содержит никакой информации, если рассматривать административный документ. В списке указывается имя человека, его номер в картотеке бюро социального обеспечения и, возможно, дата и место рождения. Списки отсылаются затем в административную картотеку, где из каждого списка исключают в случайном порядке какое-то лицо, восстанавливают желаемые данные по документам каждого из оставшихся (для кого они доступны) и вычисляют среднее, дисперсию и число случаев по наличным данным для каждого списка и каждой переменной. Затем эти данные с соответствующим обозначением по каждому списку возвращаются исследователям, оценивающим эффективность программы. Они заново составляют статистически осмысленные сочетания, а затем вычисляют для экспериментальных и контрольных групп средние, дисперсии, корреляции, зависимости от социально-экономического уровня и т. д. Таким образом, ни исследовательская, ни административная документация не обогатилась за счет друг друга никакими сведениями индивидуального порядка об участниках обследования, и в то же время эффективность программы может быть оценена статистически.

Что же касается проблемы отсева в рандомизированных экспериментах, то тут мы нуждаемся не только в новых статистических методах, но и в социально-психологических усовершенствованиях. В долгосрочных экспериментах, подобных исследованию Икеды, Ингера и Лейкока [53], в котором университет каждое лето работал с 12-летними детьми из непривилегированных слоев, пытаюсь стимулировать и направить их учебу в старших классах (с 14 до 18 лет), чтобы они были готовы к поступлению в университет, высокая дифференцированность может вызываться наряду с прочим двумя причинами. От испытуемых экспериментальной группы (с которыми имеется постоянный контакт) легче получить домашний адрес, и они охотнее отвечают на контрольные вопросы из чувства благодарности. Это наводит на мысль, что испытуемым контрольной группы, занятой в долгосрочном исследовании, следует оказывать какую-то полезную услугу в течение длительного времени, меньшую, чем экспериментальной группе, но

353

достаточную для того, чтобы побудить их ставить работников, занятых проектом, в известность в случае перемены адреса и не уклоняться от сопутствующих расспросов (Ikeda et al. [53]). Если признается, что сопоставимость экспериментальной и контрольной групп важнее, чем полнота исследования, то вполне возможно, что степень сопоставимости могла бы быть повышена путем намеренного сведения данных экспериментальной группы к уровню контрольной. В целях проверки этой возможности мы с Икедой и Ричардсоном в ходе уже упоминавшегося эксперимента проводили дополнительное обследование, используя адреса пятилетней давности, отдаленный и не связанный якобы с программой исследовательский центр и вопросы, не относящиеся конкретно к программе Икеды, Ингера и Лейкока. (Я упоминаю здесь этот отнюдь не обнадеживающий пример, чтобы выразить свое ощущение: мы нуждаемся в широком поиске возможных решений данной проблемы.)

Очевидно, что в результате отказов и отсевов истинные эксперименты имеют тенденцию превратиться в квазиэксперименты. Еще хуже то, что, если мы начинаем с рандомизации, многие потенциальные источники смещения причиняют больше хлопот тем, что они оказываются в центре нашего внимания. Я, однако, убежден, что, хотя такие смещения весьма очевидны, они на самом деле значительно меньше, чем те, которыми сопровождается более случайный выбор сравниваемых групп. Кроме того, неизмеримо возрастают возможности оценки этих смещений. По моему мнению, нам следовало бы гораздо шире использовать случайное распределение, включая процедуры отсева в текущих программах при наличии некоторого избытка желающих участвовать в эксперименте. Чтобы сделать это, нам нужно разработать практические процедуры и логические обоснования, позволяющие преодолеть сопротивление, которое встречает рандомизация в этих условиях. Чтобы показать, что в этой области имеются свои проблемы, ждущие решения, я кратко укажу некоторые из них.

Рандомизация вызывает многочисленные возражения со стороны администраторов (Conner [30]). К жребию прибегали, считая: «Пусть решает господь бог»; теперь же администратор какой-либо программы чув-

354

ствует, что он сам «играет роль бога», когда применяет процедуру рандомизации, но не тогда, когда пользуется своим собственным некомпетентным и частным суждением, основывающимся на неадекватной и нерелевантной информации (Campbell [18]). Участники эксперимента тоже противятся рандомизации, правда менее упорно, когда они сами тянут жребий, чем в том случае, если администратор сам проводит рандомизацию (Wortman et al. [104]). Составление полного списка приемлемых кандидатов и последующая рандомизация часто вызывают тягостные отсрочки, и бывает, что целесообразно бросать жребий в отношении каждого кандидата и отклонять предложения, как только все вакансии будут заполнены, в тот момент, когда контрольная группа по своей численности будет примерно равна экспериментальной. В таких условиях, как, например, специальные дома для престарелых, контрольная группа перестает быть репрезентативной в отношении неэкспериментальных условий, если тем, кто отпал по жребию, разрешено стать в очередь, — ожидание вакансии опережает нормальное решение проблемы. В этих случаях рекомендуется жребий с тремя исходами: 1) принят, 2) поставлен на очередь, 3) отвергнут. Группа 3 могла бы соответствовать контрольной. Учреждения, в которых примерно каждую неделю открывается несколько вакансий, нуждаются в специальных процедурах «процеживания», а не в крупносерийной рандомизации. Если программа недостаточно поддерживается, тот факт, что большинство людей обходится без нее, мог бы примирить испытуемых контрольной группы с выпавшим на их долю жребием; тем не менее экспериментальные процедуры, включающие в себя рандомизацию и измерение, могут вызывать острую *фокальную депривацию*<sup>1</sup>, превращающую сам по себе статус контрольного испытуемого в воздействие и состояние особого рода. Это может привести к компенсаторным

---

<sup>1</sup> Очевидно, это состояние следует отнести к большому классу состояний, которые мы условно можем обозначить как «состояния социально-психологической недостаточности» (например, «голод общения», ощущение «недостаточности внимания» и т. д.). В данном случае речь идет о состоянии, возникающем у человека, оказавшегося вне (или «не в центре») осуществляемого мероприятия и испытывающего чувство лишения, неполноценности и т. д. — *Прим. ред.*

355

усилиям или к подавленному моральному состоянию (Cook, Campbell [32]).

*Анализ нарушения непрерывности регрессии.* Одним из аргументов против рандомизированного включения в программу улучшений (там, где приемлемых кандидатов

больше, чем может быть включено в программу) является тот факт, что существуют различные степени приемлемости, нуждаемости или заслуг и что специальная программа должна охватывать самых приемлемых, самых нуждающихся в ней или достойных. Если приемлемость может быть выражена количественно (например, посредством рангов, школьных баллов, сводных оценок) и если вопрос о некоторых или всех претендентах может быть решен непосредственно на основе этой оценки, то тем самым возникает возможность использования весьма эффективного квазиэкспериментального плана, в котором нарушение непрерывности регрессии свидетельствует о наличии эффекта. Общее обсуждение деталей можно найти в работах Кэмпбелла [16] и Рикена и др. [73]. Суин [88] предложил для этой модели подходящие критерии значимости. Голдбергер [45], основываясь на эконометрических соображениях, пришел, по существу, к таким же рекомендациям.

Применение процедур количественного определения приемлемости обычно предполагает по крайней мере столь же значительный отход от общепринятых процедур отбора для участия в эксперименте, как и рандомизация. Необходима разработка особых операций, соответствующих данным условиям. Но когда они будут установлены, они окажутся экономически выгодными и окупятся большей справедливостью исследовательских процедур. Возможно, однако, что они встретят некоторое сопротивление. Администраторы любят делать исключения из правил, даже если эти правила придумали они сами. (Количественный критерий приемлемости не нуждается в определении «валидности» или «надежности»; в самом деле, когда его надежность приближается к нулю, он становится эквивалентом рандомизации.)

#### **Политико-методологические проблемы**

*Сопротивление оцениванию эффективности.* В США одной из самых общих причин, по которым доступные

интерпретации примеры оценивания программ представляют собой такую редкость, является широко распространенное сопротивление различных институтов и администраторов попыткам оценить эффективность их программ. Методология исследований по оценке эффективности должна учитывать причины этого сопротивления и способы его преодоления.

В США главным источником сопротивления является идентификация администратора и административной единицы с самой программой. В условиях политического климата США оценка программы становится оценкой данного учреждения и его руководителей. К тому же средства, применяемые для оценки программы, могут намеренно использоваться для оценивания администраторов. Помимо этого, ряд других факторов приводит к тому, что администраторы вполне обоснованно предвидят разочаровывающие результаты. Как указывает Росси [76], для специальных программ, находящихся в центре внимания исследований по оценке эффективности, отводятся обычно хронически неразрешимые проблемы, которые оказались не под силу, как правило, успешно работающим традиционным институтам. Это само по себе обуславливает пессимистический прогноз. Кроме того, финансирование обычно является неадекватным как из-за неизбежного соперничества множества причин для ограничений фондов, так и вследствие имеющейся у части законодательных и исполнительных органов тенденции предпринимать действие лишь на словах либо осуществлять косметические операции, рассчитанные скорее на то, чтобы убедить общественность в том, что меры принимаются, чем на решение проблемы. Даже если речь идет о действительно ценных программах, большое усилие, необходимое, чтобы преодолеть институциональную инерцию, препятствующую принятию каждой

новой программы, ведет к непомерно преувеличенным притязаниям. Тем самым создается «ловушка сверхагитации» («overadvocacy trap») (Campbell [16]), так что даже хорошим и эффективным программам не удается в полной мере то, что было обещано, а это еще больше усиливает страх перед оценкой.

Серьезность этих и связанных с ними проблем трудно переоценить. Хотя я в настоящем очерке и уделил

357

больше места более оптимистическим случаям, для сегодняшних американских исследований по оценке эффективности более типично положение, описанное в предыдущем абзаце. Мы, методологи, работающие в США, вынуждены принимать участие в политическом процессе, чтобы попытаться исправить это положение. Но прежде чем мы осуществим это, мы должны, сидя за столом в своих башнях из слоновой кости, изобрести политические и/или организационные альтернативы, которые позволили бы решить эту проблему. К этой задаче мы едва приступили, и это дело такого рода, где нас вполне может постигнуть неудача. Два небольших предложения послужат иллюстрацией. Я рекомендовал бы всем специалистам по методам оценки эффективности отказаться от использования наших умений в исследованиях *ad hominem*<sup>1</sup>. Хотя дорогостоящая технология социального экспериментирования и может применяться для оценки индивидов, ее не следует использовать для этого. Такие результаты доступны обобщению лишь в весьма ограниченных пределах. Наши умения следовало бы приберечь для оценки стратегий и программ, которые могут применяться более чем в одной совокупности условий, и которые при надлежащем финансировании мог бы освоить любой администратор, если только у него будет на то желание. Мы должны тщательно обрабатывать данные опроса общественного мнения, чтобы с их помощью собрать сведения только об установках в отношении предлагаемых программой альтернатив и исключить такие темы, как эффективность надзора. Этот запрет на исследования *ad hominem* нужно было бы распространить также и на тех, кто участвует в программе. Следовало бы оценивать не учащих или тех, кто получает помощь, а альтернативный подход к их проблемам. Я чувствую, что такой запрет морально оправдан, это ясно. Но я все же должен сознаться, что в условиях США он рекомендуется еще и из опасений. Администраторы программы и ее участники в силах саботировать наши усилия по оценке

---

<sup>1</sup> «К человеку» (*лат.*). Средство убеждения, основанное не на объективных данных, а рассчитанное на чувства убеждаемого. В данном случае автор отмечает, что в исследованиях описываемого типа недопустимо использовать информацию о конкретных испытуемых. — *Прим. ред.*

358

эффективности, и они попытаются воспользоваться этим, едва только будут задеты их собственные карьеры и интересы. Хотя от такой нашей стратегии администраторские страхи не рассеются окончательно, я все же считаю, что, если бы мы добросовестно следовали ей, это могло бы послужить толчком к формированию политического климата, менее настоятельно принуждающего к самозащите.

Вторая рекомендация адресована тем, кто защищает новые программы исходя из серьезности вопроса, а не достоверности того или иного ответа и при этом настаивает на необходимости переходить к следующим попыткам решить проблему, если не удастся первая (Campbell [16]). Шейвер и Стейнс [83] оспорили данную установку на том основании, что занять эту позицию проведения научных опытов означает для администратора сложить с себя обязанности руководителя. Убежденность, упорство, энтузиазм, вера — вот что требуется для всякого эффективного стремления изменить традиционную

институциональную практику. Допустить всего лишь временную веру в новую программу — значит гарантировать равнодушие при ее осуществлении. Проблема, однако, остается: «Ловушка сверхагитации» по-прежнему грозит сорвать работу по оценке эффективности программы. Ясно, что здесь нужны социально-психологические и организационно-теоретические подходы.

### Искажающий эффект количественных индикаторов

Исследования по оценке эффективности в США становятся признанным инструментом принятия социальных решений. Определенные социальные индикаторы, отбираемые с помощью таких социологических методов, как выборочные обследования, уже получили этот статус: таковы индексы безработицы и стоимости жизни, которыми располагает Статистическое управление министерства труда. Поскольку они составляют постоянную часть процесса принятия политических решений, полезно рассматривать их по аналогии с голосованием в ходе политических выборов (Gordon, Campbell [46]; Campbell [18]). При таком широком подходе, подтверждаемом качественными социологическими исследованиями того,

359

как создается общественная статистика, я прихожу к следующим пессимистическим законам (по крайней мере, для практики США): *чем более какой-либо количественный социальный индикатор используется для принятия социального решения, тем больше будет он подвержен искажающим влияниям, и тем более вероятно он будет извращать, и нарушать социальные процессы, контролировать которые он предназначен.* Я проиллюстрирую эти два закона некоторыми примерами, которые я принимаю всерьез, хотя они и носят по преимуществу анекдотический характер.

Воспользуемся для примера сравнением статистики голосования и данными переписи в Чикаго. Процесс голосования сопровождается детально разработанными предохранительными мерами, рассчитанными на то, чтобы обеспечить его честность; в отношении процесса переписи таковых мало, и обойти их ничего не стоит. Тем не менее, в нашем регионе статистика голосования внушает подозрения, тогда как статистика переписи пользуется неограниченным доверием (если не считать неполного учета взрослых мужчин-негров). Я нахожу такой порядок относительного доверия оправданным. Лучше всего он объясняется тем фактом, что «голоса» постоянно *используются* — это связано с реальными последствиями, поскольку дело касается работы, денег и власти, — и поэтому испытывают на себе мощное давление попыток исказить их. Данные же переписи до последнего времени не использовались у нас для выработки политических решений. (Даже конституционное требование, согласно которому следует менять избирательные округа, чтобы приводить их в соответствие с расселением после каждой переписи, игнорировалось десятилетиями.)

Другой пример. В согласии с духом научного управления, ответственной подотчетности, стратегии «PPBS» (о которой уже упоминалось) и т. п. полицейские участки в некоторых административных округах оцениваются с помощью «показателя очистки», иначе говоря, пропорцией раскрытых преступлений, и в случае, когда этот показатель низок, следствием является значительное давление со стороны администрации и общественности. Сколник [84] показывает, что это давление вы-

360

зывает как искажение самого индикатора, так и коррупцию уголовного судопроизводства. Регистрация не всех поступающих от населения жалоб или регистрация их задним числом,

уже после раскрытия преступления, — таковы простые уловки, которые трудно разоблачить, поскольку отсутствует независимая регистрация жалоб. Более сложный случай коррупции возникает при сочетании «показателей очистки» с «заключением сделки о признании». Это переговоры истца и суда с обвиняемым, завершающиеся соглашением относительно преступления, в котором обвиняемый готов признать себя виновным, и наказания, на которое он согласен. Тем самым сокращается стоимость и длительность судебного разбирательства. Хотя этот обычай и является законным лишь частично, в большинстве случаев он, вероятно, не столь уж нежелателен. Однако в сочетании с «показателями очистки» он, как обнаружил Сколник, ведет к следующему злоупотреблению в отправлении правосудия. Для грабителя, пойманного с поличным, дело может кончиться тем, что он получит тем менее суровое наказание, чем большее число ранее совершенных и нераскрытых преступлений он готов взять на себя. Заключением сделки о признании он оказывает полиции большую услугу, улучшая «показатель очистки», а та в свою очередь гарантирует более мягкое наказание. Сколник полагает, что во многих случаях грабитель сознается в преступлениях, которых он на самом деле не совершал. Цифры преступности являются в целом индикаторами, подверженными искажению. Для многих преступлений изменение числовых показателей отражает скорее изменение активности полицейских, а не изменение числа актов правонарушения (Gardiner [41]; Zeisel [105, 106]). Представляется достаточно хорошо документированным фактом то, что одна из широко известных преднамеренных попыток социального изменения — меры, принятые Никсоном против преступности, — имела своим главным следствием искажение количественных показателей преступности (Seidman, Couzens [81]; Morrissey [70]; Twigg [91]), связанное с неполной регистрацией преступлений и их неправомерной квалификацией как менее серьезных.

### 361

Имеются данные о сходных искажениях, происходящих при пользовании административными записями также и других типов (Kitsuse, Cicourel [59]; Garfinkel [42]). Блау [9] приводит целый ряд примеров того, как нормы производительности, установленные для работников правительственных учреждений, отрицательно сказываются на их деятельности, губительно влияя на эффективность программы. Оценивание служащих по числу посетителей, которые прошли «через их руки», привело в одном бюро по найму к поспешности и неэффективности, как в интервьюировании, так и в решениях об определении на работу. Количественная оценка персонала по числу трудоустроенных лиц привела к концентрации усилий на самых легких случаях и к игнорированию тех посетителей, кто более других нуждается в помощи, — тактика, известная как «снятие пенек» (Miller et al. [68]). В пессимистическом очерке Риджуэя о дисфункциональных последствиях измерений выполняемых действий (Ridgeway [72]) приводится еще ряд примеров.

Весьма красноречивую иллюстрацию этого принципа можно найти в одной экспериментальной программе компенсаторного образования. В эксперименте, проведенном в городе Тексаркана и посвященном «контракту о выполнении» (Stake [86]), дополнительное обучение менее подготовленных детей проводилось лицами, заключившими контракт, которые приходили в школы со специальными обучающими машинами и особыми инструкциями. Искажающие давления были весьма сильны, так как размер платы этим лицам зависел от оценок выполнения теста их учениками. Оказалось, что лица, заключавшие контракт, обучали ответам на конкретные вопросы теста, которые должны были использоваться при итоговом тестировании, и по результатам которых и производилась выплата вознаграждения. Хотя они приводили в свою защиту логико-позитивистский, операционально-дефиниционалистский аргумент о том, что цель, на которую они согласились, была определена как улучшение показателей по одному только

данному тесту, этот случай был расценен всеми как скандальный. Однако вопрос о том, насколько приемлемо натаскивание учащихся на сходных вопросах из других тестов, дебатруется до сих пор. С моей

362

точки зрения, тесты успеваемости вполне могут служить ценными показателями общей школьной успеваемости в условиях нормального преподавания, ставящего своей целью общую образованность. Но когда тестовые оценки становятся целью процесса обучения, они утрачивают свое значение как индикаторы образовательного статуса и в то же время извращают в нежелательном отношении процесс образования. (Использование «объективных» тестов в курсах обучения и на вступительных экзаменах сопровождается, несомненно, подобными же погрешностями.) С компенсаторным образованием в целом связаны случаи еще и других искажений процесса измерения, таких, например, как проведение предварительного тестирования в расчете на получение как можно более низких показателей, с тем чтобы итоговое тестирование обнаружило более значительное улучшение, или выбор для специального обучения только тех, кто при предварительном тестировании показал наихудшие результаты, так чтобы регрессия к среднему обнаружила явное улучшение. Стэйк [86] приводит перечень еще и других проблем. Тесты успеваемости фактически являются индикаторами, весьма подверженными искажениям.

Эта серьезная методологическая проблема, возможно, является универсальной. Я убежден, что это одна из главных проблем, которые предстоит решить, если мы собираемся достичь осмысленных оценок наших усилий в осуществлении запланированных социальных изменений. Эта проблема будет стоять тем острее, чем большее распространение получат количественные оценки социальных программ. Мы должны разрабатывать средства, позволяющие избежать этой проблемы, если собираемся идти вперед. Нам следовало бы изучать социальные процессы, посредством которых обнаруживается коррупция и искажение, пытаться проектировать социальные системы, наделенные этими средствами. В исследовании по теме «Контракт о выполнении» данную проблему обнаружил «внешний оценщик». В одном из более поздних американских исследований такого рода роль «караульного» взял на себя Союз учителей Сиэтла. Мы должны искать и институционализировать подобные формы и средства, предотвращающие утрату объективности. Нам следовало бы, кроме того, изучить

363

институциональную форму таких индикаторных систем, как индексы переписи населения или стоимости жизни в США, которые представляются сравнительно невосприимчивыми к искажениям. Многие комментаторы, включая меня самого (Campbell [16]), считают, что применение многочисленных индикаторов, из которых каждый признан несовершенным, облегчит решение проблемы; Риджуэй [72], однако, сомневается в этом. Имеются еще и другие проблемы, которые можно предвидеть в будущем. Особого внимания требует группа проблем, сконцентрированных вокруг использования при оценке эффективности программ обследований общественного мнения, вопросников или измерений социальных установок. В США дело идет к тому, что в скором времени нужно будет предварительно сообщать всем, кто участвует в подобных опросах, как будут использоваться полученные данные, и высылать им копии результатов. Участники будут иметь право пользоваться этими результатами в своих собственных политических целях. (Когда дело касается обследований общественного мнения, к которым прибегает правительство США, существующего у нас статуса свободы информации было бы достаточно, чтобы установить это право уже сейчас.) Можно ожидать, что в этих условиях использование опросов для оценки локальных программ правительственных служб, если оно будет осуществляться в

сложных в политическом отношении сообществах, каковые мы находим в некоторых из наших беднейших городских кварталов, приведет к следующим новым проблемам. Будут проходить политические кампании, рассчитанные на то, чтобы участники обследования отвечали именно так, как угодно местным политическим организациям, — точно так же, как проводятся кампании, имеющие своей целью влиять на результаты голосования. Будут попытки того же порядка, что и подтасовка избирательных бюллетеней. «Ошибка интервьюера» станет проблемой в еще большей степени. Нужно ждать и эффектов «бэндвагона»<sup>1</sup>, то есть кон-

---

<sup>1</sup> Bandwagon (*амер.*) — сторона, победившая на выборах. Буквально обозначает автобус с музыкантами, сопровождающими предвыборные и послевыборные действия. Эффект «бэндвагона» связан с отношением к движению, имеющему или получившему шансы на

364

формного влияния со стороны опубликованных результатов предшествующих обследований. Появятся новые смещения вроде повышенного числа жалоб.

Мне думается, что обследования общественного мнения будут все-таки полезны, если принять соответствующие меры предосторожности. Это в большинстве своем проблемы, исследованием которых мы могли бы заниматься уже сейчас, предвосхищая будущие запросы. (Гордон и Кэмпбелл [46] приводят подробное обсуждение этих проблем на материале оценки эффективности одной программы службы социального обеспечения.)

\* \* \*

## ЗАКЛЮЧЕНИЕ

Был дан сжатый обзор некоторых проблем, возникших в практике США при измерении воздействий, оказываемых запланированным социальным изменением. Разделы этой работы, где речь шла о проблемах, имеющих отношение к политическим процессам, выглядят по преимуществу пессимистическими. Хотя здесь и имеются очень серьезные проблемы, общая картина все-таки не столь уж мрачна, как кажется. Отметим, что разделы о временных сериях и рандомизированных планах эксперимента содержат данные об успешном опыте, достойном подражания. И многие квазиэкспериментальные оценки эффективности, которые я раскритиковал, могли бы быть осуществлены лучше, если бы этого требовало сообщество методологов социальной науки. Существуют, однако, новые методологические проблемы, которые возникают, когда мы переходим от экспериментирования в лабораториях в область оценки

---

успех, с подверженностью публики авторитарному давлению. О социально-психологической специфике «эффекта бэндвагона» см., в частности: Campbell D. T. On the possibility of experimenting with the «bandwagon» effect. — In: Readings in reference group theory and research. Ed. by Hyman H., Singer E. N. Y., The Free Press, 1968, p. 452—460. — *Прим. ред.*

365

эффективности социальных программ. При решении этих проблем мы испытываем потребность в новых социально-организационных усовершенствованиях.

## ЛИТЕРАТУРА

1. Anderson J. K. Evaluation as a process in social systems. Unpubl. doct. diss. Northwestern University, 1973.
  2. Baldus D. C. Welfare as a loan: An empirical study of the recovery of public assistance payments in the United States. — *Stanford Law Review*, 1973, 25, 123—250.
  3. Barnow B. S. The effects of Head Start and socioeconomic status on cognitive development of disadvantaged children. Unpubl. doct. diss. University of Wisconsin, 1973.
  4. Bauman R. A., David M. H., Miller R. F. Working with complex data files: II. The Wisconsin assets and incomes studies archive. — In: R. L. Bisco (ed.). Data bases, computers, and the social sciences. N. Y., Wiley-Interscience, 1970, 112—136.
  5. Beck B. Cooking the welfare stew. — In: R. W. Habenstein (ed.). Pathways to data: Field methods for studying ongoing social organizations. Chicago, Aldine, 1970.
  6. Becker H. M., Geer B., Hughes E. C. Making the grade. N. Y., Wiley, 1968.
  7. Becker H. W. Sociological work: Method and substance. Chicago, Aldine, 1970.
  8. Berliner J. S. Factory and manager in the USSR. Cambridge, Mass., Harvard Univ. Press, 1957.
  9. Blau P. M. The dynamics of bureaucracy (rev. ed.). Chicago, Univ. of Chicago Press, 1963.
  10. Boruch R. F., Campbell D. T. Preserving confidentiality in evaluative social research: Intrafile and interfile data analysis. Paper presented at the meeting of the American Association for the Advancement of Science. San Francisco, February/March 1974.
  11. Box G. E. P., Tiao G. C. A change in level of non-stationary time series. *Biometrika*, 1965, 52, 181—192.
  12. Box G. E. P., Jenkins G. M. Time-series analysis: Forecasting and control. San Francisco, Holden Day, 1970.
  13. Campbell D. T. Pattern matching as an essential in distal knowing. — In: K. R. Hammon (ed.). The psychology of Egon Brunswik. N. Y., Holt, Rinehart & Winston, 1966, 81—106.
  14. Campbell D. T. Administrative experimentation, institutional records, and nonreactive measures. — In: Stanley J. C., Elam S. M. (eds.). Improving experimental design and statistical analysis. Chicago, Rand McNally, 1967, 257—291. — Reprinted in: W. M. Evan (ed.). Organizational experiments: Laboratory and field research. N. Y., Harper & Row, 1971, 169—179.
  15. Campbell D. T. A phenomenology of the other one: Corrigible, hypothetical and critical. — In: Mischel T. (ed.). Human action: Conceptual and empirical issues. N. Y., Academic Press, 1969a, 41—69.
- 366
16. Campbell D. T. Reforms as experiments. *Amer. Psychologist*, 1969b, 24, 409—429.
  17. Campbell D. T. Considering the case against experimental evaluations of social innovations. *Administrative Science Quarterly*, 1970, 15, 110—113.
  18. Campbell D. T. Methods for an experimenting society. — Paper presented to the Eastern Psychological Association. Apr. 17, 1971, and to the American Psychological Association, Sept. 5, Washington D. C., 1971.
  19. Campbell D. T. Experimentation revisited: A conversation with Donald T. Campbell (by S. Salasin). *Evaluation*, 1973, 1, 7—13.
  20. Campbell D. T. Qualitative Knowing in Action Research. — (См. перевод в данной книге.)
  21. Campbell D. T., Boruch R. F. Making the case for randomized assignment to treatments by considering the alternatives: Six ways in which quasi-experimental evaluations in compensatory education tend to underestimate effects. Lecture delivered at the conference on Central Issues in Social Program Evaluation. Washington, July 1973.
  22. Campbell D. T., Boruch R. F., Schwartz R. D., Steinberg J. Confidentiality-preserving modes of useful access to files and to interfile exchange for statistical analysis. National Research Council, Committee on Federal Agency Evaluation Research, Final Report, Appendix A., 1975.
  23. Campbell D. T., Erlebacher A. E. How regression artifacts in quasi-experimental evaluations can mistakenly make compensatory education look harmful. — In: J. Hellmuth (ed.). Compensatory education: A national debate, vol. 3. N. Y., Brunner/Mazel, 1970.
  24. Campbell D. T., Fiske D. W. Convergent and discriminant validation by the multitrate-multimethod matrix. *Psychol. Bull.*, 1959, 56, 81—105.

25. Campbell D. T., Slegman C. R., Rees M. B. Direction-of-wording effects in the relationships between scales. *Psychol. Bull.*, 1967, 68, 293—303.
26. Campbell D. T., Stanley J. C. Experimental and quasi-experimental designs for research. Chicago, Rand McNally, 1966. (См. перевод в данной книге.)
27. Caporaso J. A., Roos L. L., Jr. (eds.). Quasi-experimental approaches: Testing theory and evaluating policy. Evanston, 111., Northwestern Univ. Press, 1973.
28. Caro F. G. (ed.). Readings in evaluation research. N. Y., Russell Sage Foundations, 1971.
29. Cicirelli V. G., et al. The impact of Head Start: An evaluation of the effects of Head Start on children's cognitive and affective development. A report presented to the Office of Economic Opportunity, June 1969. Westinghouse Learning Corporation, Ohio University.
30. Conner R. F. A methodological analysis of twelve true experimental program evaluations. Unpubl. doct. diss., Northwestern University, 1974.
31. Cook T. D., Appleton H., Conner R., Shaffer A., Tamkin G., Weber S. J. Sesame Street revisited: A case study in evaluation research. N. Y., Russel Sage Foundation, 1975.

32. Cook T. D., Campbell D. T. The design and conduct of quasi-experiments and true experiments in field settings. — In: Dunnette M. D. (ed.). Handbook of industrial and organizational research. Chicago, Rand McNally, 1976.
33. Cronbach L. J. Response sets and test validity. *Educ. and Psychol. Measurement*, 1946, 6, 475—494.
34. Cronbach L. J. Further evidence on response sets and test design. *Educ. and Psychol. Measurement*, 1950, 10, 3—31.
35. David H. P. Family planning and abortion in the socialist countries of Central and Eastern Europe. N. Y., The Population Council, 1970.
36. David H. P., Wright N. H. Abortion legislation: The Romanian experience. *Studies in Family Planning*, 1971, 2, 205—210.
37. Douglas J. D. The social meanings of suicide. Princeton Univ. Press, 1967.
38. Edwards A. L. The social desirability variable in personality assessment and research. N. Y., Dryden, 1957.
39. Fairweather G. W. Methods for experimental social innovation. N. Y., Wiley, 1967.
40. Fischer J. L. The uses of Internal Revenue Service data.— In: Borus M. E. (ed.). Evaluating the impact of manpower programs. Lexington, Mass., D. C. Heath, 1972, 177—180.
41. Gardiner J. A. Traffic and the police: Variations in law-enforcement policy. Cambridge, Mass., Harvard Univ. Press, 1969.
42. Garfinkel H. «Good» organizational reasons for «bad» clinic records. — In: H. Garfinkel. Studies in ethnomethodology. Englewood Cliffs., Prentice Hall, 1967, 186—207.
43. Glaser D. Routinizing evaluation: Getting feedback on effectiveness of crime and delinquency programs. Washington, D. C., Center for Studies of Crime and Delinquency, 1973.
44. Glass G. V., Wilson V. L., Gottman J. M. Design and analysis of time-series experiments. Boulder. Colo., Univ. of Colorado, 1972.
45. Goldberger A. S. Selection bias in evaluating treatment effects: Some formal illustrations. Discussion Papers, 123—172. Madison, University of Wisconsin, 1972.
46. Gordon A. C., Campbell D. T. et al. Recommended accounting procedures for the evaluation of improvements in the delivery of state social services. Dupl. paper, Northwestern University, 1971.
47. Goslin D. A. Guidelines for the collection, maintenance, and dissemination of pupil records. N. Y. Russell Sage Foundation, 1970.
48. Guttentag M. Models and methods in evaluation research *J. for the Theory of Soc. Beh.*, 1971, 1, 75—95.
49. Guttentag M. Evaluation of social intervention programs. *Annals of the New York Academy of Sciences*, 1973, 218, 3—13.
50. Gullentag M. Subjectivity and its use in evaluation research. *Evaluation*, 1973, 1, 60—65.
51. Hatry H. P., Winnie R. E., Fisk D. M. Practical program evaluation for state and local government officials. Washington, D. C., 1973.

52. Heller R. N. The uses of social security administration data. — In: M. E. Borus (ed.). *Evaluating the impact of manpower programs*. Lexington, Mass., 1972, 197—201.

368

53. Ikeda K., Yinger J. M., Laycock F. Reforms as experiments and experiments as reforms. Paper presented to the meeting of the Ohio Valley Sociological Society, Akron, May 1970.
54. Jackson D. N., Messick S. Response styles on the MMPI: Comparison of clinical and normal samples. *J. of Abn. and Soc. Psychol.*, 1962, 65, 285—299.
55. Кепка Е. J. Model representation and the threat of instability in the interrupted time series quasi-experiment. Doct. diss. Northwestern Univ., 1971.
56. Kershaw D. N. A negative income tax experiment. *Scientific American*, 1972, 227, 19—25.
57. Kershaw D. N. The New Jersey negative income tax experiment: a summary of the design, operations and results of the first large-scale social science experiment. Dartmouth/OECD Seminar on «Social Research and Public Policies. Sept. 13—15, 1974.
58. Kershaw D. N., Fair J. Operations, surveys, and administration. Vol. IV of the Final Report of the New Jersey Graduated Work Incentive Experiment. Madison, Wisconsin, 1973.
59. Kitsuse J. K., Cicourel A. V. A note on the uses of official statistics. *Social Problems*, 1963, 11, 131—139.
60. Кун Т. Структура научных революций. М., 1977.
61. Kutchinsky B. The effect of easy availability of pornography on the incident of sex crimes: The Danish experience. *J. of Soc. Iss.*, 1973, 29, 163—181.
62. Levenson B., McDill M. S. Vocational graduates in auto mechanics: A follow-up-study of Negro and white youth. *Phylon*, 1966, 27, 347—357.
63. Lohr B. W. An historical view of the research on the factors related to the utilization of health services. Dupl. Res. Report. Bureau for Health Services Research and Evaluation. Rockville, Md., U. S. Government Printing Office, 1972.
64. Lord F. M. Large-scale covariance analysis when the control variable is fallible. *J. of the Amer. Statistical Association*, 1960, 55, 307—321.
65. Lord F. M. Statistical adjustments when comparing preexisting groups. *Psychol. Bull.*, 1969, 72, 336—337.
66. McCain L. J. Data analysis in the interrupted time series quasi-experiment. Master's thesis. Northwestern Univ., 1975.
67. Miller A. R., *The assault on privacy*. Ann Arbor, Mich., Univ. of Michigan Press, 1971.
68. Miller S. M., Roby P., Steenwijk A. A. V. Creaming the poor. *Trans-action*, 1970, 7, 38—45.
69. Morgenstern O.. *On the accuracy of economic observations*. (2nd ed.) Princeton, Princeton Univ. Press, 1963.
70. Morrissey W. R. Nixon anti-crime plan undermines crime statistics. *Justice Magazine*, 1972, 1.8—11.14.
71. Porter A, C. The effects of using fallible variables in the analysis of covariance. Unpub. doct. diss., Univ. of Wisconsin, 1967.
72. Ridgeway V. Disfunctional consequences of performance measures. *Administrative Science Quart.*, 1956, 1, 240—247.
73. Riecken H. W., Boruch R. F., Campbell D. T., Caplan N., Glennan T. K., Prall J., Rees A., Williams W.

369

*Social Experimentation: a method for planning and evaluating social intervention*. N. Y., Academic Press, 1974.

74. Rivlin A. M. *Systematic thinking for social action*. Washington, D. C., The Brookings Institution, 1971.
75. Ross H. L. Law, science and accidents. The British Road Safety Act of 1967. *J. of Legal Studies*, 1973, 2, 1—75.
76. Rossi P. H. Practice, method, and theory in evaluating social-action programs. — In: Sundquist J. L. (ed.). *On fighting poverty*. N. Y., Basic Books, 1969, 217—234.

77. Rossi P. H., Williams W. (eds.). Evaluating social programs: Theory, practice, and politics. N. Y., Seminar Press, 1972.
78. Reubhausen O. M., Brim O. G., Jr. Privacy and behavioral research. *Columbia Law Review*, 1965, 65, 1184—1211.
79. Sawyer J., Schechter H. Computers, privacy, and the National Data Center: The responsibility of social scientists. — *Amer. Psychologist*, 1968, 23, 810—818.
80. Schwartz R. D., Orleans S. On legal sanctions. *Univ. of Chicago Law Review*, 1967, 34, 274—300.
81. Seidman D., Couzens M. Crime statistics and the great American anti-crime crusade: Police misreporting of crime and political pressures. Paper presented at the meeting of the American Political Science Association. Washington, D. C., Sept. 1972.
82. Segall M. H., Campbell D. T., Herskovits M. J. The influence of culture on visual perception. Indianapolis, Ind., Bobbs-Merrill, 1966.
83. Shaver P., Staines G. Problems facing Campbell's «Experimenting Society». *Urban Affairs Quarterly*, 1971, 7, 173—186.
84. Skolnick J. H. Justice without trial: Law enforcement in democratic society. New York, Wiley, 1966.
85. Smith M. S., Bissell J. S. Report analysis: The impact of Head Start. *Harvard Educational Review*, 1970, 40, 51—104.
86. Stake R. E. Testing hazards in performance contracting. *Phi Delta Kappan*, 1971, 52, 583—588.
87. Suchman E. Evaluation research. N. Y., Russell Sage Foundation, 1967.
88. Sween J. A. The experimental regression design: An inquiry into the feasibility of non-random treatment allocation. Unpubl. doct. diss. Northwestern Univ., 1971.
89. Thompson C. W. N. Administrative experiments: The experience of fifty-eight engineers and engineering managers. *IEEE Transactions on Engineering Management*, 1974, EM — 21, p. 42—50.
90. Thorndike R. L. Regression fallacies in the matched groups experiment. *Psychometrika*, 1942, 7, 85—102.
91. Twigg R. Downgrading of crimes verified in Baltimore. *Justice Magazine*, 1972, 1, 15.
92. Watts H. W., Rees A. (eds.). Final report of the New Jersey graduated work incentive experiment. Vol. 1., Madison, Wisconsin, 1973.
93. Weber S. J., Cook T. D., Campbell D. T. The effects of school integration on the academic self-concept of public school children. Paper presented at the meeting of the Midwestern Psychological Association. Detroit, 1971.
94. Weiss C. H. (ed.). Evaluating action programs: Readings in social action and education. Boston, Allyn & Bacon, 1972a.

95. Weiss C. H. Evaluation research. Englewood Cliffs., Prentice-Hall, 1972b.
96. Weiss R. S., Rein M. The evaluation of broad-aim programs: A cautionary case and a moral. *Annals of the American Academy of Political and Social Science*, 1969, 385, 133—142.
97. Weiss R. S., Rein M. The evaluation of broad-aim programs: Experimental design, its difficulties, and an alternative. *Administrative Science Quart.*, 1970, 15, 97—109.
98. Westin A. F. Privacy and freedom. N. Y. Atheneum, 1967.
99. Wheeler S. (ed.). Files and dossiers in American life. N. Y., Russell Sage Foundation, 1969.
100. Wholey J. S., Nay J. N., Scanlon J. W., Schmidt R. E. If you don't care where you get to, then it doesn't matter which way you go. Dartmouth/OECD Seminar on Social Research and Public Policies. Sept. 13—15, 1974.
101. Wholey J. S., Scanlon J. W., Duffy H. G., Fukumoto J., Vogt L. M. Federal evaluation policy. Washington, D. C., 1970.
102. Wilder C. S. Physician visits, volume and interval since last visit. U. S. 1969. Rockville, Md., National Center for Health Statistics, Series 10, № 75, July 1972.
103. Williams W., Evans J. W. The politics of evaluation: The case of Head Start. *The Annals*, 1969, 385, 118—132.

104. Wortman C. B., Hendricks M., Hillis J. Some reactions to the randomization process. Dupl. res. report, Northwestern Univ., 1974.
105. Zeisel H. And then there were none: The diminution of the Federal jury. *Univ. of Chicago Law Review*, 1971, 38, 710—724.
106. Zeisel H. The future of law enforcement statistics: A summary view. — In: Federal statistics: Report of the President's Commission. Washington, D. C. U. S. Government Printing Office, vol. II, 1971, 527—555.

371

## ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ<sup>1</sup>

- Административные, архивные и другие косвенные данные, использование 210, 255, 341, 351—353, 360—362, 364
- «Административный эксперимент», модель 334
- Ad hominem, исследования 358, 359
- Артефакты, таксономия 202—211, 233, 234
- Асимметричный (уницентрированный) перевод 300—302, 312
- Асимметрия сравнения единичных случаев 293
- Байесова статистика, проверка при изучении единичных случаев 290
- «Бэндвагон», эффект 364, 365
- Валидность
- внешняя 43, 45, 62, 63, 71—84, 86, 88—90, 94—99, 101—107, 113, 121, 130, 139—141, 150, 198, 200—206, см. также «эффекты взаимодействия», «осведомленность об эксперименте», «обобщение результатов», «нереактивные измерения»
  - внутренняя 41, 42, 45, 50, 63—72, 86, 105, 106, 116—121, 130, 133, 150, 200—204, см. также «главный эффект»
  - качественного подхода 271
  - конструктивная 41—43, 348
  - научного вывода 193, 194, 206, см. также «метанаучные проблемы»
  - неэкспериментального знания 43, 44
  - полевого исследования и наблюдения 264
  - социальные источники валидности науки 33, 356—359
  - статистического вывода 40, 41, см. также «проверка значимости, статистической, эффекта воздействия»
  - теории и научная революция 335
  - теории при изучении случаев 282—286, 288—295

---

<sup>1</sup> Именной и предметный указатели составлены М. И. Бобневой.

372

## Варьирование

- дополнительное 211
- качества данных 213
- систематическое (интерполирующее и экстраполирующее) 213—214

## Взаимодействие

- монотонность 97
- определение категории 95—97, см также «эффекты взаимодействия», «смешивание», «валидность»
- переменных при изучении единичных случаев 292, 293

## Воспроизведение (повторение) эксперимента, требование

- истинного эксперимента 140
- оценивания программ 265, 333
- полевого исследования 266
- при использовании переводов вопросников и тестов 310
- с использованием различных методов (гетерометодное) 105—107, 152, 215—216
- экспериментального воздействия 103, 104

## Временных серий (рядов), эксперимент по плану 114—123, 124, 144, 172, см. также «прерывных временных серий, эксперимент по плану»

## Выборки

- выделение групп сравнения при анкетировании одной аудитории 156
  - из конечной популяции 101
  - интервьюеров 180
  - источник систематической погрешности 145, 146
  - когда невозможно случайное комплектование групп 140, 142, 143
  - по ситуации, ошибка 127, 128
  - принадлежность к одной и той же популяции 136, 157, 158
  - смещения выборки 76, 77, 147
  - уравненные, в квазиэксперименте 155
  - формирование подгрупп популяции по принципу «достоинства» 161 — 172
  - формирование блоков 91
  - эквивалентные, ситуаций 123
  - эквивалентные, экспериментального материала 128—131
- см. также «Репрезентативность», «Смещение, погрешности», «Рандомизация», «Попарное уравнивание», «Панельные исследования»

## Главный (основной) эффект 71, 88, 89, 95—99, 139—140, 142, 143, 200—206, 209, 210, см. также «Валидность»

## Демаскирование эксперимента 224—226, 229, 233

## Децентрированный (симметричный) перевод 300—302, 308, 309

## Добровольное участие в эксперименте 62, 70, 71, 86, 87, 92, 136, 147, 209, 214, 215, 219

## Добровольцы-испытуемые 209, 214, 215

## Дозэкспериментальные планы 48-62

## Единичные случаи — см. «Исследование единичного случая»

## Естественное осуществление эксперимента 175

## «Естественный эксперимент» 176, 177

## Жребий 156, 354, 355

## Замаскированные эксперименты

- отказ от демаскировки 224, 225

- приемы маскировки лабораторных исследований 220
- таксономия полевых 217, 221—234
- этические проблемы 222—227, 229, 232—233, 235
- Искажающий эффект количественных индикаторов 359—364
- Информанты, метод использования 261, 266, 328
- Исследование единичного случая 48—50, 265—267, 278—297
- Истинные (подлинные) эксперименты 62—107, 140, 144, 146, 147, 171, 206

Качественное знание, качественный подход

- валидное обоснование количественных данных 256, 267, 270, 271, 327
- в социальных науках 254, 255, 320
- зависимость количественного знания от 243, 272
- качественное обобщение качественных суждений 268—270
- при оценивании программ 243, 253—257, 324—330
- полярность качественного и количественного подхода 242, 271, 272, 297
- типы качественного оценивания 257—267, 324—330, см. также «Обыденное знание», «Метанаучные проблемы»
- Каузальные связи и эксперимент — см. «Эксперимент», «Корреляционные планы», «Корреляции», «Валидность»
- Квазиэксперименты (квазиэкспериментальные планы) 107—184, см. также «Оценивание эффективности социальных программ»
- Контрольные выборки для предварительного и итогового тестирования, план 147—150, 172
- Контрольные группы
  - «двойной слепой эксперимент» 66—81, 202, 220
  - «боязнь оценки» 202, 203, 215, 219

374

- влияние ожиданий экспериментатора на результаты эксперимента 202, 203, 233
- как средство контроля внутренней валидности, история использования 63, 105, 113, 114, 200—202
- плацебо 70, 71, 105, 112, 201, 202, 219
- подвергающиеся дополнительному воздействию 203—205, 219
- подвергающиеся «ложной операции» 201, 211
- применение в естественных науках 200—202
- при оценивании программ 338, 343, 348—350
- см. также «Рандомизация», «Попарное уравнивание», «Эквивалентные группы», «планы экспериментов»

Корреляции

- выявление ложных 213
- выявление отрицательных 345
- источник погрешностей 178, 179
- обратная корреляция 175
- при предварительном испытании гипотез 174
- с третьей (сопутствующей) переменной 175, 342
- см. также «Корреляционные планы» Корреляционные планы и планы ex post facto 172—181

Латинский квадрат

- в истинном эксперименте 140, 141
- вместо контрольной группы 94, 139—143
- для построения сбалансированных планов 138—143
- «Ловушка сверхагитации», эффект 357, 359
- Лонгитюдинальный подход 154—160, 178
- «Лоскутные» планы 113, 144, 152, 160

Межкультурные сравнения как слабая форма квазиэкспериментального плана 279

Метанаучные проблемы

- влияние процесса качественного оценивания программ на методы гуманитарных наук 265
- объективность данных и пристрастность исследователей 331, 332
- при оценивании программ 324—336
- степень интеграции — унифицированности — науки 249
- эмпирический статус методологических требований 198
- см. также «Эпистемология», «Качественное знание»
- Методологические проблемы исследований
  - конформизма 130, 131, 224, 226, 227
  - межкультурных различий 211—213, 245, 278—

## 375

- 280, 282, 291—297, 301, 302, 306, 308, 310—315
  - национальных предрассудков 54, 74, 179—181
  - общественного мнения 122, 144—146, 179, 180, 199—201, 230, 231, 248, 364
  - оказания помощи и контактов в общественных местах 229—233
  - психотерапевтических 135—136, 144, 158
  - социальных установок 154, 159, 177, 178, 219, 232, 310, 314, 315, 321, 364, 365
  - убеждающих воздействий 102, 206, 225, 226, 230
- см. также «Панельные исследования», «Оценивание эффективности социальных программ»
- Методы и приемы оценивания программ
  - внешнее наблюдение 262, 330—336
  - возможности экспериментального оценивания 162, 318—330
  - выработка единого решения участниками 268, 269
  - использование информантов 328
  - использование косвенных данных 338—340, 351—353, 360—362
  - критерий отбора кандидатов 161, 168, 356
  - недостатки количественно-экспериментального подхода 324, 327
  - опыт участников 257—259
  - системный анализ 261
  - типы качественного оценивания 257—267
- см. также «Квазиэкспериментальные планы», «Наблюдение», «Эксперимент», «Качественное знание», «Оценивание эффективности социальных программ»
- Множественные серии измерений, план 150, 151, 172, 175
- Модели дисперсионного анализа — конечные, рандомизированные, фиксированные и смешанные 101
- Модель исследования с использованием многих наблюдателей и множества объектов описания 294—297

### Наблюдение, полевое

- методы 263, 264
- надежность и валидность 264
- разведение субъективного и объективного компонентов 293, 294

### Надежность

- оценки экспериментальной и контрольной группы 344
- полевого наблюдения 264
- теста при предварительном и итоговом тестировании 344

Нарушение непрерывности регрессии как признак экспери-

## 376

- ментального эффекта, план 160—172, 356
- Независимые переменные как признаки классификации 97—101
- Нереактивные измерения — 55, 210, 219, 221, 234, 255, 307, 314
- см. также «Угрозы внешней валидности»
- Нестинг, перекрестная классификация 97—101

Неэквивалентная контрольная группа, план 56, 63, 64, 68, 113, 132—138, 147, 150, 172, 175, 182, 183

Обобщение (генерализация) результатов 91, 102—106, 131, 198, 205—207, 209, 218, 268, см. также «Валидность», «Угрозы внешней валидности», «Эффекты взаимодействия»

Обозначения, используемые при описании планов 47, 48

Обработка административных данных с учетом требования конфиденциальности, методика 351—353

Обратный перевод вопросников и тестов, методика 302—305

Обыденное знание, здравый смысл

— в социальных науках 278, 279

— и научное знание 245—247, 283, 326

— погрешности 287—289

— сомнение и доверие 279 см. также «Качественное знание»

Осведомленность об эксперименте 217, 221—231, 235, см. также «Замаскированные эксперименты», «Угрозы внешней валидности», «Этические проблемы исследований»

Оценивание эффективности социальных программ

— в антропологических и этнографических исследованиях 261—265, 328, 329

— влияние политической ситуации 32, 322, 323, 356—365

— информантами 328

— практические проблемы организации процесса 330—333

— представителями различных социальных дисциплин 271, 319—322

— специальными подразделениями 330—331

— участие социальных психологов 320, 321

— участниками программы 267—271, 364, 365

— экспертами (качественное оценивание) 258

Ошибка в независимых переменных 342

Панельные исследования 178—181, 199—201, см. также «Методологические проблемы исследования общественного мнения»

Перекрестная валидизация

— в экспериментальном исследовании 38, 82, 122

377

— данных тестов социальных установок данными невербальных тестов 314

— наук 29

— науки здравым смыслом 252—253

— количественных оценок качественным знанием 267, 270

— при оценивании программ 259, 260, 334, 335

— эффекта экспериментального воздействия с помощью различных методов 105—107, 152, 215, 216

Перекрестные планы 138

План эксперимента

— в традиции Фишера 34, 35, 63, 90

— и предварительное тестирование 89, 198—200

— и способ статистической обработки 35, 36, 84, 85, 260

— понятие 28, 34

— эмпирический продукт науки 200, 201

План Соломона для четырех групп 88, 89

Планы, не предусматривающие предварительного тестирования 198—200

Планы с переключением 138

Погрешности — см. «Смещения»

Полевые исследования, методологические проблемы

— воспроизводимость 266

— перенос лабораторных исследований в полевые условия 108—114, 138, 146, 151, 177

— стандарты регистрации данных 266

- стратегии 152—160, 310
- таксономия замаскированных 221—234
- см. также «Методы и приемы оценивания программ»
- этические проблемы 222—227, 229, 232, 233, 235, 331—333
- Попарное уравнивание групп 35, 68, 135, 181—183
- Поэтапные срезы, сравнительное изучение 154—160
- Правдоподобные конкурентные гипотезы 50, 66, 111, 112, 115, 140, 152, 174, 176, 197— 213, 215, 234, 256, 310
- Предварительное и итоговое тестирование 50, 56—58, 71, 74, 75, 255, 256, 311, 342, 362, 363, см. также «планы экспериментов и квазиэкспериментов»
- Предварительное и итоговое тестирование и контрольная группа, план 63—88, 201
- Предварительное и итоговое тестирование на одной группе, план 50—61, 201
- Предварительное и итоговое тестирование на различных выборках, план 143—147, 172
- Прерывные временные серии (ряды), план 336—341
- Применение контрольной группы и тестирования, осуществляемого только после воздействия, план 89—93

### 378

- Проверка значимости в качественном исследовании 288—291
- Проверка статистической значимости эффекта воздействия 84—87, 88, 89, 92, 93, 94, 123, 124, 127, 128, 131, 142, 151, 166—171, 339—342, 344—356
- Проверка (испытание) теории
  - валидный выбор 91
  - как процесс отбрасывания несостоятельных гипотез 109, 110, 194—197
  - с помощью «степеней свободы» при изучении единичных случаев 282, 283, 285, 286, 288—295
- Рандомизация
  - единицы 81, 82, 348—350
  - и обеспечение равенства 162
  - как средство достижения эквивалентности групп 35, 68, 86, 90, 135, 143, 207, 355
  - коррекция (подгонка) регрессии как замещение 341—345
  - при интерпретации данных 166, 167
  - при компенсаторном обучении 344
  - распределения испытуемых по группам для тестирования 143
  - распределения кандидатов для участия в программе 161, 162, 165, 167
  - режима эксперимента для целых групп 85, 147
  - уровни, на которых проводится 348—350
  - уровни эквивалентности, достигаемые 150, 153
- Регрессионные артефакты при оценивании программ 340, 342, 343
- Регрессия
  - дифференциальная 182
  - множественная 171, 342
  - подгонка в плане с прерывными временными сериями 337
- см. также «Угрозы внутренней валидности»
- Рекуррентный институциональный цикл («Лоскутный план»), план 113, 144, 152—160
- Рекуррентная форма эксперимента с одной группой, план 123
- Репрезентативность
  - выборки 146, см. также «Выборка»
  - временных интервалов измерения и воздействия 124
  - в естественных науках 207, 208
  - в социологии и психологии 207, 208
  - и фактор времени 207
  - контрольной группы при оценивании программ 355
  - объекта исследования 75
  - опроса всей намеченной выборки 171, 172

- опроса и получения сопоставимых данных 171, 172
- результатов точных наук и воспроизводимость результатов 74
- результатов исследования в отношении более широкой популяции 86
- см. также «Выборка», «Валидность», «Угрозы внешней валидности», «Смещения»
- Ретроспективное предварительное тестирование 176—178

Сбалансированные планы 138—143

Смешивание 124, 209—211, 325

Смещения в результатах, погрешности

- в рандомизированном эксперименте при оценивании программ 354, 355
- в результате отсева при оценивании программ 349, 350, 351
- измерительных приборов 193
- и систематический эффект 154, 342
- как процесс спонтанной ремиссии 144
- коррекции между воздействием и результатом итогового тестирования 178, 179
- «переменная» и «постоянная» ошибка 320
- при уницентрированном переводе тестов и вопросников 305, 306, 310—314
- таксономия погрешностей, связанных с главными эффектами и эффектами взаимодействия 202—211
- см. также «Выборка», «Репрезентативность», «Угрозы внутренней валидности», «Угрозы внешней валидности»

Снижение качества

- валидности и надежности теста для получения интерпретируемых данных 307
- данных
- — как средство выявления ложных корреляций 213
- — как средство контроля конкурентных гипотез 211—213

Сомнение и доверие в процессах научного исследования 247—252, 279

Социальные индикаторы при оценивании программ 28, 340, 359—361

Социальные программы и нововведения

- дистрибутивные 343, 344
- критерии отбора кандидатов для участия 161—172
- компенсаторные 265, 268, 269, 333, 343-345, 363
- помощи нуждающимся 162, 345—347
- примеры программ, проходящих оценивание

333, 334, 337, 338, 345, 346, 360, 361

Сравнение при интерпретации данных

- возможность сравнения в квазиэкспериментах 334—335
- и снижение качества данных 307, 308
- неинтерпретируемость сравнений двух единичных случаев 173, 280
- при оценивании программ 341, 343
- см. также «Эквивалентные группы»
- «Степени свободы» в нестатистической ситуации 282—289, 291—293

Тесты и вопросники, методологические проблемы разработки и применения

- вербальные и невербальные 314, 315
- использование эквивалентных форм теста 311
- исследование процедуры тестирования 91
- надежность при предварительном и итоговом тестировании 344
- пересмотр версии теста на языке оригинала 307-309
- повторное применение того же самого теста 92
- погрешность 192

- построение вопросников 310
  - разработка децентрированного теста 306—315
  - смещения при использовании уницентрированного перевода 310—314
  - «стандартизированная форма» 306, 307
  - установка на ответы 209
- см. также «Предварительное и итоговое тестирование»
- Триангуляция данных, методика 293—295

Угрозы внешней валидности (факторы, угрожающие)

- взаимная интерференция экспериментальных воздействий 47, 122, 125, 126, 137, 138, 292, 293
  - реактивный эффект (эффект взаимодействия тестирования и экспериментального воздействия) 47, 74—76, 88, 90, 121, 137, 147, 151, 172, 210, 211
  - реакция испытуемых на эксперимент 47, 80—84, 121, 125, 131, 137, 155, 210, 217—221, 231, 235
  - эффект взаимодействия фактора отбора с экспериментальным воздействием 47, 76—78, 88, 122, 125, 137, 138, 146, 150, 172, 325
- см. также «Эффекты взаимодействия»
- Угрозы внутренней валидности (факторы, угрожающие)
- взаимодействие фактора отбора с естественным развитием 47, 133,

134, 151, 154, 160, 172, 203, 204

- другие факторы 47, 202—204
  - естественное развитие 46, 51, 53, 54, 64, 66, 79, 88, 120, 125, 132, 133, 134, 140, 144, 151, 154, 156, 171, 203, 204
  - инструментальная погрешность (нестабильность измерительного инструмента) 46, 55, 56, 66—67, 78, 88, 120—121, 125, 132, 133 отбор испытуемых (фактор состава групп) 46, 60, 61, 62, 68, 69, 121, 125, 133, 151, 154, 158, 171, 206
  - отсеб (выбывание) 46; 62, 69, 78, 125, 154, 159, 171, 172, 183, 200—204, 349—354
  - статистическая регрессия 46, 56—61, 67, 68, 121, 134, 146, 151, 154, 171, 203, 204
  - и качественное знание 256
  - фон (история) 46, 50, 64-66, 79, 88, 116-120, 125, 132, 133, 140, 142, 143, 154, 171, 205—206, 256
  - эффект тестирования 46, 54, 55, 66, 71, 74, 75, 88, 90, 125, 132, 140, 141, 145, 146, 154, 203, 204, 219
- см. также «Главный эффект», «Оценивание эффективности социальных программ»
- Учет сезонных тенденций при оценивании программ 337, 339, 340

Факторные планы 93—101

Фокальная депривация 355, 356

Хоуторнские исследования, эффект 42, 346, 347

Эквивалентные временные выборки, план 123—128

Эквивалентные группы

- контрольные, для сравнения при оценивании программ 343
- неэквивалентные группы до введения программы 341
- неэквивалентные группы при нарушении регрессии 171
- перед экспериментом 61, 62, 87, 123, 132, 135
- при сравнении 281
- уравнивание групп, когда невозможно их случайное комплектование 140, 142, 143
- частные характеристики и тенденции 147

Эксперимент

- естественный 176, 177
- истинный 62—107, 140, 144, 146, 147, 171, 206
- как средство проверки каузальных гипотез 39—43, 198—211, 325

- квазиэксперимент 107—184
- многомерность 39
- несовершенство 45, 108—113, 324, 327

382

- ориентация психологов на 320
- полевой 221—234
- рекуррентный 123
- ротационный 36, 138 см. также «План эксперимента», «Методы и приемы оценивания программ», планы
- Эксперимент *ex post facto*, план 181—183, 343
- Экспериментальная изоляция 51, 119, 120
- Экспериментальные приемы использования переводов тестов и вопросников 310
- Эпистемология
  - дескриптивная 30—32, 200, 204, 244, 245, 326
  - качественного и количественного подхода 243
  - качественного распознавания образа 252, 326
  - эволюционная 109—113, 244, 248
- Этические проблемы исследования
  - анонимность респондента 92
  - добровольное участие в исследовании 62, 70, 71, 86, 87, 92, 136, 219
  - конфиденциальность данных в исследовании 351—353
  - запрет на исследования «Ad hominem» 358, 359
  - обман и демаскировка 224—235
  - согласие на участие в эксперименте 223, 348—350
  - устранение вторжения в личную жизнь 223
- Эффекты взаимодействия 71, 88, 89, 95—97, 139—141, 198, 204, 205

383

## ИМЕННОЙ УКАЗАТЕЛЬ

- Абельсон Р. П. 231
- Адамс Дж. С. 228
- Анастаси А. 54
- Андервуд Б. Дж. 38, 92, 102, 104, 113, 115, 126, 131, 138
- Андерсон Дж. К. 341
- Андерсон Н. Г. 75
- Андерсон Р. Б. У. 311
- Андреева Г. М. 15, 16, 26
- Апхофф Г. Ф. 66
- Аргайл М. 16, 224, 229
- Аронсон Э. 219, 221, 224
- Аш С. Э. 222, 224
  
- Байрен Дж. 221
- Бакленд У. Р. 36

Барноу Б. С. 333  
Бауман Р. А. 351  
Бауэр Р. Т. 306  
Барч А. М. 176  
Бейкер К. Г. 66, 145  
Бейкер Ф. 261  
Бек Б. 255, 320  
Беккер Г. 255, 283, 287, 320  
Беннет Дж. У. 306  
Беннис У. Дж. 259  
Бенсмен Дж. 255  
Бенц В. А. 310, 311  
Беркан М. 229  
Беррин Ф. К. 308, 313, 314  
Бимен Э. 107  
Бин Дж. Р. 229  
Бине А. 192  
Бисселл Г. У. 180  
Бисселл Дж. С. 333  
Биттермен М. 213, 214  
Блау П. М. 255, 362  
Блахович Дж. А. 251  
Блейк Р. Р. 227, 231  
Блок Дж. 210  
Богомолова Н. Н. 16  
Бокс Дж. Э. П. 123, 339  
Болдус Д. 334, 337, 339  
Боринг Э. Г. 49, 63, 200  
Борук Р. 32, 135, 160, 319, 345, 348, 350, 351, 356  
Боханнен П. Дж. 301  
Бош Э. Э. 279  
Брайан Дж. 227, 229  
Браун Р. 220  
Браунли К. А. 35  
Брим О. Дж. 176, 351  
Бройер М. Б. 306  
Брольер С. Р. 137  
Брок Т. К. 231  
Брукс Р. 115  
Брунсвик Э. 104, 124, 233, 240  
Бэкон Ф. 287, 288  
Бялек Г. М. 229

Вебер С. Дж. 333, 341, 344  
Вейсс Р. С. 259, 261, 324, 325  
Вейсс С. Г. 319  
Верба С. 306  
Вернер О. 267, 300  
Верной О. 311  
Верч Дж. 34

Виатт Д. 288  
Виатт С. 125  
Видич А. 255  
Винтер Б. Дж. 35  
Витгенштейн Л. 249, 250  
Вудвортс Р. С. 137  
Вудьярд Э. 137  
Вулф Р. Л. 269

Галилей Г. 290  
Гардинер Дж. Э. 361  
Гартманн Дж. У. 225, 226  
Гарфинкель Г. 255, 320, 362  
Гау Г. Дж. 312  
Гвишиани Д. М. 330  
Гейер Ф. М. 286, 290  
Гецкоу Г. 104, 106  
Гласе Дж. В. 33, 84, 336, 337, 339  
Гленнэн Т. К. 3, 319, 348, 350, 356  
Гликман С. Е. 113  
Глок К. И. 75, 179, 180, 199  
Глэзер Д. 319  
Гнанадезикан Р. 39  
Голдбергер Э. С. 165, 356  
Гор П. М. 231  
Гордон Л. В. 312, 313  
Гордон А. 270, 341, 359, 365  
Гослин Д. А. 351  
Госнелл Г. Ф. 225, 226  
Готтмен Дж. М. 336, 339  
Грант Д. А. 100  
Гржегоржевская И. А. 15  
Грин В. Ф. 84, 99  
Гринвуд Э. 181, 183  
Гринстоун Дж. Д. 265, 266  
Гросс А. Э. 221, 227, 228  
Гуд К. В. 37  
Гудмен Н. 244  
Гутгентаг М. 259, 324, 325

Дайкер Г. К. 306  
Даймонд Р. 71  
Даннетт М. 33, 40  
Дарли Дж. М. 220  
Даффи Г. Дж. 319  
Дери С. 241  
Джастроу Дж. 249, 250  
Джекобсон Ю. Г. 306  
Джексон Д. Н. 94, 320  
Джексон Р. У. Б. 35, 94  
Дженкинс Дж. М. 339  
Джианнота Ф. Дж. 262, 263

Джилльберт Н. 259  
Джиллиленд А. Р. 75, 146, 156  
Джир Б. 320  
Джэнис А. Л. 102, 218  
Джонс Г. 240  
Джонс Р. У. 261  
Джонсон П. О. 35  
Диксон У. Дж. 42, 218  
Диннерстайн Д. 241  
Дойч М. 143, 177, 178, 199, 241  
Дуб Л. У. 315  
Дуб Э. Н. 227  
Дуглас Дж. Д. 255, 320  
Дункан К. П. 75, 146, 156  
Дункер К. 248, 249  
Дэвид Г. П. 336  
Дэвид М. Г. 351  
Дэвис Л. 75, 146, 156  
Д'Эпине Р. Л. 306  
Дюбуа С. А. 287

Жуков Ю. М. 15

385

Зеллтиц К. 143, 199, 241, 243  
Зигель А. 81  
Зигель С. 81  
Зигмен К. Р. 209, 320  
Зингер Э. 365

Ивахара 308  
Икеда К. 353, 354  
Инско К. 231

Йингер Дж. М. 353, 354  
Йост А. 129, 130, 131

Казенс М. 255, 361  
Кант И. 196  
Кантор Дж. Н. 85  
Каплан М. 259  
Каплан Н. 32, 319, 348, 350, 356  
Капорезо Дж. 319  
Кардинер А. 287  
Карлсмит Дж. М. 219, 220, 221, 224  
Каро Ф. Дж. 160, 319, 336  
Катнер Б. 232  
Кауз П. 245  
Кац Д. 174  
Кейзер Г. Ф. 84  
Кейн В. Р. 54

Келли Г. Г. 102, 218  
Келли Э. Л. 104, 106  
Келман Г. С. 220, 222  
Кемпторн О. 87, 88, 94, 126, 128, 138, 143  
Кендалл М. 36  
Кениата Дж. 278  
Кеннеди Дж. Л. 66  
Кении Д. Э. 307  
Кепка Э. Дж. 340  
Керн Р. П. 229  
Керр У. А. 124, 125, 126, 217, 228  
Киддер Л. 219, 278  
Кикучи А. 312, 313  
Кинг Д. Дж. 75  
Киршоу Д. Н. 345  
Китсьюз Дж. К. 255, 320, 362  
Клэйтон К. Н. 179, 181  
Кокс Г. 138  
Кокс Д. Р. 35, 68, 127, 128, 138  
Коллинз Б. 221  
Коллинз М. 177, 178  
Колльер Р. М. 51  
Колльер Р. О. 128  
Кордиг К. Р. 251  
Коннер Р. Ф. 333, 344, 354  
Конфуций 286  
Корсини Р. Дж. 210  
Кохрен У. Дж. 138  
Коэн Р. 300, 328  
Крачфилд Р. 241  
Кречевский И. 283  
Крёбер А. Л. 266, 284  
Крёбер Т. 284  
Кристи Р. 210  
Кровец М. Л. 229  
Кронбах Л. Дж. 209, 210, 320  
Крук М. Н. 51  
Крэно У. Д. 306, 307  
Куайн У. Б. 30, 193, 244, 245, 247, 248, 300  
Кудер 311  
Куин С. А. 181  
Кук С. У. 143, 199, 228, 241, 243  
Кук Т. Д. 33, 40, 47, 229, 231, 255, 333, 337, 341, 343, 344, 346, 356  
Кун Т. 9, 31, 193, 197, 244, 248, 249, 250, 251, 335  
Кучинский Б. 336  
Кьялдини Р. Б. 229  
Кэмпбелл Д. Т. 6—15, 18—20, 27, 29, 32—34, 39, 40, 45, 47, 54, 55, 75, 107, 109, 111,

288, 297, 300, 305—307, 310, 314, 318—321, 324, 326, 328, 329, 334, 336, 337, 339, 340—343,  
344—346, 348, 350, 351, 355—357, 359, 364, 365  
Кэнтрил Х. 306, 314

Лайонс Дж. М. 318  
Лазарсфельд П. Ф. 217  
Лакатош И. 244  
Ламздейн А. А. 75, 102, 156  
Лана Р. Э. 75, 198, 199  
Лапьер Р. Т. 232  
Лассуэл Г. Д. 286  
Латане Б. 220  
Левайн М. 269  
Левайн Р. А. 262, 267, 305, 328  
Левенсон Б. 351  
Леви Б. Г. 231  
Левин К. 6, 240, 241, 243  
Лейкок Ф. 353, 354  
Лемен Э. У. 255  
Лефкович М. 227  
Лиджфарт А. 279, 285, 291  
Линдсей Дж. 315  
Линдквист Э. Ф. 35, 68, 81, 85, 94, 140  
Липпитт Р. 243  
Лоннер У. Дж. 310, 311  
Лорд Ф. М. 61, 292, 343  
Лови 266  
Любин А. 97, 143

Майерс Дж. Л. 100  
Мак-Гайр В. 16, 202, 215, 219, 222, 224  
Мак-Дилл М. С. 351  
Мак-Дэвид Дж. 17  
Мак-Кэйн Л. Дж. 340  
Мак-Кичи У. Дж. 104, 106  
Маккоби Н. 174  
Мак-Кол У. Э. 35, 36, 39, 63, 68, 138  
Мак-Кормак Т. 153, 156  
Мак-Лафлин К. П. 261  
Мак-Нимар К. 35, 59, 61  
Максуэлл Э. 115, 127, 138  
Малиновский Б. 261, 262, 278, 280  
Манн Л. 228  
Мерритт К. Б. 228  
Мессик С. 320  
Мецей Л. 228  
Мил П. Э. 290  
Милгрэм С. 220, 224, 228, 229  
Миллер А. Р. 351  
Миллер Дж. К. 231  
Миллер Н. 221, 231  
Миллер Р. Ф. 351

Миллер С. М. 362  
Милль Дж. С. 73, 204  
Мичел Р. 306  
Монро У. С. 37  
Морз Н. 174  
Моррисси У. Р. 255, 361  
Мур Г. Т. 130  
Мутон Дж. С. 227, 231  
Мэррей Д. К. 75, 146, 156  
Мэрроу А. Дж. 243  
Мэрчисон К. 240  
Мюллер Г. Э. 129, 130  
Мюллер-Лайер 245, 246, 248

Найда Э. 301  
Наннэлли Дж. 84  
Ней Дж. Н. 319  
Нейман Дж. 79  
Нейрат О., фон 247

387

Нейролл Р. 212, 213, 283, 300, 328  
Нелсон Г. 262, 263  
Новик М. Р. 292  
Ньюкомб Т. М. 178  
Ньютон И. 194, 196, 197  
Нэнгл Дж. 176

О'Брайен Р. Б. 75, 146, 156  
Олмонд Дж. 306  
Олпорт Г. 311  
Олпорт Ф. Г. 124, 125, 127  
Орлеане С. 227, 351  
Орн М. 15, 17, 202, 219, 220  
О'Шонесси Дж. 346

Панферов В. Н. 15  
Павлов И. П. 119  
Пейдж Э. Б. 83, 227  
Пепитоне А. 241  
Петерсон П. 265, 266  
Петри Г. Дж. 198  
Петровская Л. А. 16  
Пирс Э. 75  
Пирсон Г. С. 63  
Поляни М. 193, 244  
Поппер К. 32, 110, 193, 195, 197, 244, 245  
Портер А. 343  
Пратт Дж. 32, 319, 348, 350, 356  
Прошански Г. 241  
Пуанкаре А. 289

Радке М. 241, 243  
Райен Т. 291  
Райт Н. Г. 336  
Рей А. А. 229  
Рейзер Дж. Р. 282, 286  
Реймер Э. 174  
Рейн М. 259, 261, 324, 325  
Рейнолдс Дж, 262, 263  
Ривлин А. М. 319  
Рид Дж. К. 71  
Риджудей В. 255, 362, 364  
Рикен Г. У. 32, 160, 319, 345, 348, 350, 356  
Рис А. 32, 319, 345, 348, 350, 356  
Рис М. Б. 209, 320  
Ричардсон Дж. 131, 354  
Роби П. 362  
Робхаузен О. М. 351  
Роджерс К. Р. 71  
Розбум У. У. 84  
Розенберг М. 17, 202, 203, 215, 219  
Розенблат П. 221  
Розенталь Д. 202, 217  
Розенталь Р. 15, 66, 192, 202, 203, 209  
Рой С. Н. 39  
Рокич М. 224, 228  
Роос Л. 319  
Рорер Л. Дж. 209  
Росноу Р. Л. 192, 209  
Росс Х. Л. 336—338  
Росси П. Г. 319, 357  
Ротлисбергер Ф. Дж. 42, 218  
Роттер Дж. Б. 231  
Рэнкин Р. Э. 54, 220

Саймон Г. А. 176  
Санфорд Ф. Г. 133  
Сегалл М. Г. 211, 245, 248, 310  
Сейдмен Д. 255, 361  
Секрест Л. 55, 210, 219, 221, 234, 255, 307, 314  
Селвин Г. С. 291  
Сепир Э. 300  
Сикурел А. В. 255, 320, 362  
Сильвермен Дж. 15, 17  
Сильвермен Л. Г. 210  
Скейтс Д. 37

Скиннер Б. Ф. 283  
Сколник Дж. Г. 227, 360, 361  
Скэнлон Дж. У. 319

Смит Г. Л. 145  
Смит М. С. 333  
Собол М. Дж. 75  
Сойер Дж. 351  
Соломон Р. Л. 63, 64, 74, 80, 88, 112, 156  
Соммер Р. 230  
Сорокин П. А. 124, 125, 127  
Стар Ш. 143  
Стауффер С. А. 19, 49, 179  
Стейк Р. 362, 363  
Стейнберг Дж. 351  
Стейнс Дж. 359  
Стерн Э. 306  
Стинвийк А. А. В. 362  
Сток Ф. Дж. 125  
Стокфорд Л. 180  
Столлак Дж. Э. 224  
Стэнли Дж. 15, 33—35, 39, 79, 84, 99, 101, 107, 143, 200, 206, 255, 279, 339, 343, 346  
Стэнтон Ф. 66, 145  
Стюарт В. М. 248  
Стюарт А. 291  
Суин Дж. А. 166, 356  
Сэлэзин С. 242  
Сэлмон У. 195

Твигг Р. 255, 361  
Термен Л. 192  
Тест М. А. 227, 229  
Тистлтуэйт Д. Л. 165  
Толмен Э. 6, 240, 283  
Томпсон С. У. Н. 334  
Торндайк Э. Л. 36, 137  
Торндайк Р. Л. 61, 182, 343  
Трамбо Д. 176  
Трусов В. П. 15  
Тулмин С. 193, 196, 244, 245  
Тьюки Дж. У. 84, 99, 101  
Тьяо Дж. 123, 339  
Тэмкин Дж. 333, 344  
Тэшфел А. 16

Уайли Д. Е. 39  
Уаттс Г. У. 345  
Уилер С. 351  
Уилк М. Б. 87, 88, 94, 143  
Уилкинс К. 232  
Уиллсон В. Л. 336, 339  
Уилсон Э. Б. 193  
Уильяме У. 32, 319, 330, 348, 350, 356  
Уиндл К. 51, 54, 85  
Уинни Р. Э. 334  
Уолд Г. 176

Уолкотт Г. Ф. 262  
Уорф Б. 300  
Уортмен К. Б. 355  
Уотсон Р. И. 112  
Уэбб Ю. Дж. 55, 210, 219, 221, 234, 255, 307, 314  
Уэстин А. Ф. 351

Фармер Э. 115  
Фаулер Р. Дж. 228  
Фейерабенд П. К. 244, 248, 249, 251  
Фейуэдер Дж. У. 319  
Фелдман Р. Е. 229, 232  
Фелдмен Г. 228  
Фелдт Л. С. 68, 85  
Фергюсон Дж. А. 35, 101  
Фестингер Л. 220  
Фёгелин С. Ф. 301  
Фёгелин Ф. 301  
Филлипс Г. П. 303, 306  
Фиск Д. М. 334  
Фиске Д. У. 107, 210, 306, 307, 314, 320  
Фишер Дж. Л. 351  
Фишер Р. А. 34—36, 39, 85, 90  
Фогт Л. М. 319

389

Франк Дж. О. 217  
Фрей П. У. 268  
Фрейд З. 280, 284, 285  
Френч Дж. 220  
Френч Л. Р. 262, 263  
Фридмен Дж. Л. 231  
Фрийда Н. 310, 314  
Фримен Г. 259  
Фрэзер Дж. А. 125  
Фрэзер С. К. 231  
Фрэнзен Р. 232  
Фукумото Дж. 319

Хаймен Г. 145, 365  
Хардинг Дж. 241, 243  
Хартер С. 228  
Хейм А. У. 54  
Хейн Дж. Д. 231  
Хеллер Р. Н. 351  
Хелсон Г. 231, 288  
Хемфилл Дж. К. 133, 134  
Хемпел К. Дж. 195  
Хендрикс М. 355  
Херсковиц М. 211, 245, 248, 310  
Хикс Ч. 84, 97

Хиллис Дж. 355  
Ховленд С. Дж. 75, 102, 156, 218  
Холи Дж. С. 319  
Хэммонд К. 104, 233  
Хэнсон Н. Р. 110, 244, 248, 249, 250, 251  
Хэртари Г. 17  
Хэтри Г. П. 334  
Хьюз Э. 143  
Хьюз Э. К. 320

Цейзель Г. 75, 334, 336, 361

Чейн А. 241, 243  
Чепмэн Дж. С. 36  
Чичирелли В. Дж. 333, 344  
Чэдвик Р. У. 282  
Чэпин Ф. С. 181, 182, 183

Шамберс Э. Дж. 115  
Шапс Э. 232  
Шаффер А. 333, 344  
Шварц Р. Д. 55, 210, 219, 221, 227, 234, 255, 307, 314, 351  
Швейцер А. 7  
Шейвер П. 359  
Шелдон А. 261  
Шервуд К. 259  
Шериф К. У. 29  
Шериф М. 29  
Шеффе Г. 101, 291  
Шеффилд Ф. Д. 156, 351  
Шехтер С. 306, 310  
Шехтер Х. 351  
Шилпп П. А. 32  
Шимони А. 30, 245  
Шпехт Г. 259  
Шу Ф. Л. К. 262, 280  
Шульберг Э. 261  
Шухман Э. 319, 320

Эббингауз Г. 126  
Эванс Дж. У. 330  
Эверхарт Р. Б. 262  
Эдвардс А. Л. 35, 94, 210, 308,  
313, 320  
Эйнштейн А. 195—197  
Эйлер Л. 194—196  
Экенсбергер Л. Г. 279  
Эпплтон Г. 333, 344  
Эрвин С. М. 306, 311  
Эриксон Э. Г. 284—287, 290, 292  
Эрлебекер А. 265, 343, 344  
Этциони А. 255

Юм Д. 72, 193, 196, 197, 245  
Юнг А. Ф. 232

Яги К. 229  
Ярроу П. Р. 232  
Яхода Г. 310, 314  
Яхода М. 143, 19

## ОГЛАВЛЕНИЕ

Методологические проблемы экспериментального метода в социальных науках ( <i>Вступительная статья</i> ).....	5
1. Несколько замечаний относительно теории науки и методологии эксперимента в социальной психологии ( <i>перевод А. У. Хараша</i> ).....	27
2. Эксперименты и квазиэксперименты ( <i>перевод Ю. М. Жукова и П. Б. Шошина</i> ).....	34
3. Научный вывод, артефакты и контроль ( <i>перевод А. У. Хараша</i> ).....	192
4. Качественное знание в исследованиях действия ( <i>перевод А. У. Хараша</i> ).....	240
5. «Степени свободы» и изучение отдельных случаев ( <i>перевод А. У. Хараша</i> ).....	278
6. Децентрированный перевод вопросников и тестов ( <i>перевод А. У. Хараша</i> ).....	300
7. Измерение воздействий, оказываемых запланированным социальным изменением ( <i>перевод А. У. Хараша</i> ).....	318
Предметный указатель.....	372
Именной указатель.....	384

**Д. Кэмпбелл**

**МОДЕЛИ ЭКСПЕРИМЕНТОВ  
В СОЦИАЛЬНОЙ ПСИХОЛОГИИ  
И ПРИКЛАДНЫХ ИССЛЕДОВАНИЯХ**

ИБ № 6099

Редактор *Пчелкина Э. М.*  
Художественный редактор *Суима А. Д.*  
Технические редакторы *Никитина В. Ю., Черкасова О. И.*  
Корректор *Шарганова Н. И.*

Сдано в набор 14.03.80. Подписано в печать 6.11.80. Формат 84×108<sup>1/32</sup> Бумага типографская № 1. Гарнитура литературная. Печать высокая. Условн. печ. л. 19,58. Уч.-изд. л. 21,72. Тираж 11 000 экз.  
Заказ № 599. Цена 1 р. 60 к. Изд. № 29314.

Издательство «Прогресс» Государственного комитета СССР  
по делам издательств, полиграфии и книжной торговли  
Москва, 119021, Зубовский бульвар, 17

Ленинградская типография № 2 головное предприятие ордена Трудового Красного Знамени Ленинградского объединения «Техническая книга» им. Евгении Соколовой Союзполиграфпрома при Государственном комитете СССР по делам издательств, полиграфии и книжной торговли. 198052, г. Ленинград, Л-52, Измайловский проспект, 29