

---

---

## ПЕРЕВОДЫ

Ф.А. Шродт  
(США)

### СЕМЬ СМЕРТНЫХ ГРЕХОВ СОВРЕМЕННОГО КОЛИЧЕСТВЕННОГО АНАЛИЗА В ПОЛИТИЧЕСКОЙ НАУКЕ<sup>1</sup>

Технологические изменения, методологические течения и толика интеллектуальной неряшливости и нерасторопности, особенно по части философии науки, привели к тому, что современная количественная методология политической науки накопила ряд опасных привычек,

---

**Филип А. Шродт** — старший исследователь в Parus Analytical Systems, в прошлом – профессор политических наук, Университет штата Пенсильвания. E-mail: schrodt@psu.edu.

*Переводчик:* **Анна Александровна Широконова** – кандидат социологических наук, старший научный сотрудник, Лаборатория сравнительных социальных исследований, Национальный исследовательский университет «Высшая школа экономики»; доцент, Департамент социологии, Национальный исследовательский университет «Высшая школа экономики» – Санкт-Петербург. E-mail: ashirokanova@hse.ru.

Оригинал статьи размещен здесь: <http://7ds.parusanalytics.com/Schrodt.7Sins.APSA10.pdf> Сокращенная версия опубликована в: Journal of Peace Research. 2014. Vol. 51. No. 2. P. 287–300. Перевод выполнен по полной версии с незначительными сокращениями в рамках конкурса, объявленного журналом «Социология: методология, методы, математическое моделирование» в 2016 г.

<sup>1</sup> Работа подготовлена для тематической панели «Поворот в политической методологии?» на Ежегодной встрече Американской ассоциации политической науки (APSA) 2–5 сентября 2010 г., г. Вашингтон. Эта статья улучшалась в обсуждениях с Дэвидом Кольером, Патриком Брандтом и Джоном Фрименом, которые не несут ответственности ни за содержание, ни за представление. Особенно за представление. Исследование частично поддержано грантом от Национального научного фонда США (SES-1004414).

делающих значительную часть современных исследований с научной точки зрения более-менее бесполезными. Лекарство от этого – не в том, чтобы отбросить количественные методы, и, совершенно определенно, не в постмодернистском нигилистическом отрицании любого систематического метода, а в том, чтобы, обратившись к началам, рассмотреть некоторые неприятные проблемы, не пытаясь приращивать знание исключительно с помощью все более интенсивного использования быстро-сокращающихся мышц для нажатия на клавиши компьютерной мышки.

В данной работе «семь смертных грехов» определены следующим образом.

1. Модели по принципу «танцуют все!», игнорирующие эффект мультиколлинеарности.

2. Донаучное объяснение при отсутствии предсказания.

3. Бесконечный анализ одних и тех же баз данных, пока они не взвоят.

4. Использование сложных методов без понимания их исходных предположений.

5. Интерпретация частотной<sup>1</sup> статистики, как если бы она была байесовской<sup>2</sup>.

6. Линейная статистическая монокультура, поддерживаемая в ущерб альтернативам.

7. Смещение понятий статистического и экспериментального контроля.

Ответом на эти проблемы должна быть серьезная, продуманная, оригинальная работа, основанная на уважении и к теории, и к данным. Не постмодернизм. В заключительной части приводится обзор того, как мы пришли к нынешнему состоянию с точки зрения философии на-

---

<sup>1</sup> Частотный, или фреквентистский вывод (от frequency – «частота»), или частотная статистика, – один из многих способов формулировки схем статистического вывода. На нем основана методология проверки значимости статистической гипотезы и доверительных интервалов. Альтернативы ему – байесовский вывод, фидуциальный вывод. – *Прим. перев.*

<sup>2</sup> Байесовский вывод, или подход (по имени Т. Байеса) – метод статистического вывода, использующий теорему Байеса для вычисления того, как степень веры в истинность предположения о вероятности события или априорного суждения изменяется на основе имеющихся доказательств. – *Прим. перев.*

уки с XVII по XX в., и высказывается ряд предложений по изменению философского и педагогического подходов, которые могут служить для решения некоторых из изложенных проблем.

*Ключевые слова:* количественный анализ, частотная статистика, байесианство, методы анализа данных, линейные модели.

### *Постановка проблемы*

В последние годы я замечаю, что количественные исследования, которые приходят мне на рецензирование из журналов или накануне конференций, все чаще вызывают у меня неудовольствие. В редких случаях все еще попадаются самородки, поэтому, в духе афоризма С. Джонсона «второй брак – это торжество надежды над опытом», я продолжаю работать рецензентом<sup>1</sup>. Однако для типичной работы, которую я получаю на чтение, характерен ряд (или сразу весь набор) следующих раздражающих признаков:

– около дюжины коррелирующих друг с другом независимых переменных в одной линейной модели;

– новая первоклассная и невероятно сложная статистическая процедура (доступная в удобном виде в *Stata* или в *R*), которая в лучшем случае совершенно избыточна для решения поставленной задачи, в то время как обычный *t*-тест или дисперсионный анализ вполне справились бы с задачей извлечь немногие правдоподобные результаты из данных, а в худшем – она совершенно неуместна, с учетом характеристик данных и/или теории;

– в работе анализируются данные, которые были проанализированы тысячу раз до этого;

---

<sup>1</sup> Впрочем, уже меньше, что наверняка вызовет вздох облегчения у многих читателей этих заметок.

– объем работы составляет 30–40 страниц<sup>1</sup>, хотя результаты легко могли бы быть изложены на десяти (или менее) страницах – как принято в естественных науках.

Наконец, не в самой работе, но почти наверняка по факту:

– сообщаемые результаты – итог десятков (а может, и сотен) альтернативных формулировок оцениваемой модели.

Проблема подобных работ – в том, что я не верю их результатам. Но когда я вспоминаю, что у автора(-ов), вероятно, есть голодные дети, стареющие родители, маленькие пушистые собачки, а впереди их ждет академический совет, который просто подсчитывает количество опубликованных статей в рецензируемых научных журналах, то я могу предложить мало чего конструктивного. Однако, к сожалению, такие работы всего лишь отражают практику, которая стала «нормальной наукой» в количественном политическом анализе. Не помогают даже рекомендации в стиле: «Измените тему, данные, модель и интерпретацию, и тогда, возможно, это станет интересным». И это несмотря на факт, что вообще-то существует множество вещей, которые можно было бы делать иначе<sup>2</sup>, и что я полностью положительно и оптимистично смотрю на перспективы количественного политического анализа. Если избавиться от этих грехов.

Цель данной статьи, сформулированная при подготовке к тематической панели APSA «Резкий поворот в политической методологии?», – остановиться на указанных проблемах (без риска для конкретного автора) более подробно, чем допускается в отзыве рецензента. Как завзятый неомедиевист, я использую троп «семь смертных грехов» для их классификации, хотя мне и пришлось сильно постараться, чтобы сократить их количество до

---

<sup>1</sup> Если это второй круг рецензирования (R&R). Первые варианты предлагаемых статей – 50–70 страниц с припиской извинений, где значителся, что авторы понимают, что работу, возможно, придется немного сократить.

<sup>2</sup> Начиная с того, чтобы департаменты оценивали, а не просто считали статьи.

семи. Первоначально мой список грехов был ближе к двадцати, и внимательный разбор предложенной мной дискуссии позволит разглядеть, как они встроились в более широкие категории. Данная статья развивает мои более ранние тезисы [1], а в интеллектуальном смысле - значительно обязана Эйкену [2]<sup>1</sup> и Уорду (например: [4]). И, в духе полемики, пожалуй, также работе Хаффа [5], которая, если верить Википедии, была продана более чем в 1,5 млн экземпляров [6].

Если коротко обозначить мой вывод (и пояснить, что я не просто ною: как несправедлива Вселенная к социальным наукам), я бы отметил, что эти проблемы имеют разнообразные способы решения, но в целом укладываются в следующие категории.

1. Проблемы, вызванные тем, что, как мы знаем из любого вводного курса статистики, делать нельзя, но мы все равно делаем из-за своей лени: регрессии по принципу «танцуют все!»; анализ с применением комплексных методов без соблюдения необходимых исходных предположений.

2. Проблемы, связанные с изменением концепции статистического вывода во времени: объяснение, исключаящее прогнозирование; байесовская интерпретация частотных оценок; интерпретация независимых переменных как «контрольных».

3. Проблемы, обусловленные технологиями, сделавшими нас ленивыми: бесконечный повторный анализ небольшого количества баз данных с использованием еще меньшего количества методов.

4. Проблемы, которые обсуждаются со времен появления современной статистики и, возможно, неразрешимы: частотная статистика в целом; интерпретация тестов значимости на основе данных о генеральной совокупности.

---

<sup>1</sup> Эйкен делает примерно две трети из тех замечаний, которые хочу сделать я, формулирует их лучше, чем я, и уже сказал их десяток лет назад. А также он перекликается в некоторых моментах с работой Кинга [3], написанной за 15 лет до этого. И все это принесло нам много пользы... так что повторю-ка я их еще раз.

5. Просто сложные вопросы, интеллектуальная или техническая работа над которыми может принести прогресс целой области анализа: более совершенная философия статистического вывода на основе современных практик байесовского анализа и ее внедрение в преподавание методологии взамен частотной статистики и логического позитивизма середины XX в.

Наконец, хочу отметить, что в моих наблюдениях неизбежен уклон в сторону исследований, которые я чаще всего рецензирую, – количественного анализа политических конфликтов в международной сравнительной перспективе. Вопросы, которые я поднимаю, могут в меньшей степени касаться, скажем, исследований голосования в США. Хотя я в этом сомневаюсь. Читатели чувствительного склада могут посчитать эту статью по крайней мере умеренно полемической. В этом я не сомневаюсь. Но «выступление на круглом столе APSA – это когда не нужно говорить “прости”»<sup>1</sup>.

### *1. Модели по принципу «танцуют все!»<sup>2</sup> и проблема мультиколлинеарности*

Модель по принципу «танцуют все!»<sup>2</sup> (Эйкен использует термин «модель мусорного бака», garbage can models) – это анализ, включающий, по формулировке Эйкена, «длинный список независимых переменных из социальной психологии, социологии и случайных эмпирических наблюдений, которые беспорядочно

---

<sup>1</sup> Парафраз популярной цитаты из романа Э. Сигала «История любви» и одноименного фильма 1970 г.: «Любовь – это когда не нужно говорить “прости”». – *Прим. перев.*

<sup>2</sup> В оригинале – kitchen sink models (от англ. everything but the kitchen sink – «все, кроме кухонной мойки», т.е. «все, что смогли унести») – пренебрежительное название моделей, использующих длинный список независимых переменных и игнорирующих теоретическое обоснование модели в надежде получить значимые эффекты; вид охоты за статистической значимостью. – *Прим. перев.*

перемешиваются в готовых к употреблению пакетах линейных регрессий» [2, р. 424]. Эта проблема поставлена первой в списке смертных грехов не потому, что я собираюсь излагать ее в деталях, и не потому, что ее никто не касался – Эйкен [2], если считать по Google Scholar<sup>1</sup>, был процитирован уже 166 раз<sup>2</sup>, – а потому, что это самый частый корень зла и причина примерно 80% моего недоверия к современным количественным исследованиям.

Эйкен ввел краткое «правило трех» (на основе ряда методологических и технических обоснований): «Где есть более трех независимых переменных, никто не может гарантировать, что модель точна, а предпосылки соблюдены настолько, насколько утверждает исследователь... Совершенно резонно, что при тщательном анализе данных модель с тремя объяснительными переменными обычно достаточно хороша – не слишком проста и не слишком сложна – для любой статьи» [2, р. 446].

В другой своей работе [1] я уже касался вопроса, где эта проблема обычно возникает и какие серьезные последствия впоследствии, когда коррелирующие переменные используются в регрессионных или логистических моделях<sup>3</sup>. В отличие от контролируемого эксперимента (на котором выросли современные статистические методы), где необходимые переменные можно варьировать независимо друг от друга, политический аналитик обычно сталкивается с ситуацией, когда одинаково правдоподобные теории предлагают в качестве возможных причин несколько тесно связанных (и поэтому коррелирующих) переменных.

Линейные модели не справляются с подобными ситуациями. Мультиколлинеарность может вылиться в ситуацию, когда все действительные коэффициенты будут казаться незначимыми, а часто

---

<sup>1</sup> Ребята из мира журналов, может, вы будете следовать наставлениям автора, а не просто его цитировать?

<sup>2</sup> 476 раз на апрель 2017 г. – *Прим. перев.*

<sup>3</sup> Отсюда и до конца раздела – в значительной степени самозаимствование из [1].

это может привести и к противоположному знаку оценки прямого эффекта переменной. Уберите нужную переменную (проблема пропущенной переменной) – и ее объяснительная способность отразится на всех связанных с ней переменных, оставшихся в уравнении. Попробуйте включить любые категориальные переменные, помимо самых простых, – и от анализа не останется живого места от *dummy*–переменных (чи коэффициенты тоже на практике лишь изредка интерпретируются верно – по крайней мере, в статьях, которые я рецензирую).

В отсутствие сильного линейного эффекта в генеральной совокупности регрессия не изолирует, а *усиливает* влияние аномальных выборок. Какая же часть из опубликованных результатов – это всего-навсего «франкенштейны», состоящие из облака некоррелирующих наблюдений, где все значимые коэффициенты определены немногими группками выбросов? – неизвестно; потому что лишь очень немногие авторы тщательно проверяют такую возможность.

В общем, для многих, если не для большинства проблем, обычно встречающихся в политическом анализе, линейные модели не просто плохи, а очень и очень плохи. Сложно было бы придумать худший набор потенциальных побочных эффектов.

Вследствие этого результаты линейных регрессий, как печально известно, весьма нестабильны, так что даже небольшие изменения в модели приводят к тому, что коэффициенты пляшут, как коробка с тушканами на метамфетаминах. Это прекрасно для создания большого количества статистических исследований – возьмите хотя бы непрекращающиеся дебаты в международных отношениях о демократическом мире (ну пожалуйста), – но уже не так хорошо, если вы хотите когда-нибудь прийти к выводам. Консервативный ответ на скачущие коэффициенты – «нужно сначала разрешить эти несоответствия на теоретическом уровне». Но обычно весь смысл проведения теста как раз в том и состоит, чтобы эмпирически различить, какие из альтернативных



и равновероятных теорий заслуживают внимания! Лекарство становится равнозначно болезни, и мы рассмотрим это дальше на несоответствиях между гипотетико-дедуктивным методом и частотной статистической парадигмой, частью которой являются эти линейные модели.

## *2. Научное объяснение при отсутствии предсказания*

Недавно мне был отправлен черновик введения к тематическому выпуску журнала с моей статьей о прогнозировании. Черновик – я надеюсь, это было в конечном счете изменено, – начинался со слишком знакомого для меня утверждения: «Многие исследователи международных отношений все еще видят в прогнозировании второстепенную задачу по сравнению с объяснением».

Отрывок «Многие исследователи международных отношений все еще видят» неоспорим: это глубоко укоренившийся догмат веры среди значительной части сообщества. Не устраивает меня часть о «прогнозировании как второстепенной задаче по сравнению с объяснением».

Мое обдуманное и взвешенное отношение к подобным утверждениям таково: это в высшей степени, совершенно и полностью своекорыстный абсурд, лишенный и следа философского обоснования, и при этом невероятно вредный для всего нашего интеллектуального предприятия. Лучше всего такие утверждения прямиком посылать на свалку истории и припоминать лишь как примеры того, что *не следует* делать. Я ясно выразился?

В чем же я вижу проблему? Я постоянно наталкиваюсь на подобные высказывания, но до сих пор ни разу, *ни разу* никто не смог привести ни одной подходящей цитаты из философии науки в ее поддержку. По-видимому, это просто обыденное знание, полученное, как я подозреваю, из силлогизма:

модели, с которыми мы работаем, практически бесполезны для прогнозирования;  
мы доблестные ученые несравненного интеллекта и интуиции,  
*следовательно,*  
научные модели не должны служить для прогнозирования.

Вполне понятно человеческое стремление «не можешь забить гол – передвинь ворота», но в данном контексте оно приводит меня в смущение, поскольку создавать успешные прогнозы политического поведения – это прямое назначение моделей (недавние показательные примеры включают в себя Проект по политической стабильности (Political Instability Task Force, PITF) [7] и Интегрированную систему раннего предупреждения о возникновении кризисных ситуаций (Integrated Crisis Early Warning System, ICEWS)[8]), даже если большинство моделей этого и не делают [4]. Более того, первопроходцы в области исследований международных отношений были решительно заинтересованы в моделях, которые могли бы точно предсказывать события, поскольку их отчасти мотивировали надежды применить эти знания и таким образом снизить риски испечься в американо-советском термоядерном пожаре. С тех пор в понимании вещей произошли определенные подвижки.

Позиция современников понятна, но при этом философски пуста. И более всего раздражает, что сторонники этого утверждения – по крайней мере, на моем опыте – просто переписывают его из текста в текст, будто от этого оно станет истинным. Так что в отсутствие аргументов, с которыми стоит поспорить, можно лишь привести доказательства в пользу обратного. И даже это довольно трудно, так как здесь мы наталкиваемся на полный разрыв между подобным мнением и философией науки XX в., которая исходила из прогностических детерминистских традиций Ньютона, Лапласа, Максвелла и всего опыта естественных наук начиная с середины XVII в., так что центральное место прогноза (а не просто объяснения) в большинстве научных дискуссий принимается как должное.

Покопавшись, можно найти ряд емких аргументов. Например, название классического эссе об охватывающих законах Карла Гемпеля «Объяснение и предсказание с помощью охватывающих законов» [9] предполагает, что оба процесса (объяснение и прогнозирование) идут бок о бок в собственно научной теории. Текст этого эссе, а также название поддерживают это утверждение: Гемпель недвусмысленно рассматривает объяснение как эквивалент предсказания. Логические позитивисты, будучи достаточно (иногда до безумия) строгими<sup>1</sup> логиками, были бы, несомненно, возмущены подобным утверждением. Поскольку если две вещи эквивалентны, то одновременно одна из них не может быть слабее другой!

Гемпель и Оппенгейм (1948) пишут об этом следующее:

Мы хотели бы здесь заметить, что один и тот же формальный анализ, включающий четыре необходимых условия, применим как к объяснению, так и к научному предсказанию. Различие между ними имеет прагматический характер. Так, если дано  $E$ , т. е. если мы знаем, что явление, описанное  $E$ , имело место, и затем обеспечен набор утверждений  $C_1, C_2, \dots, C_k, L_1, L_2, \dots, L_i$ , то мы говорим об объяснении происшедшего явления. Если сначала даны последние утверждения, и  $E$  выводится из них до того, как явление, описанное им, будет иметь место, то мы говорим о предсказании. Таким образом, можно сказать, что объяснение не является полностью адекватным, если его эксплананс с течением времени не сможет послужить основой для предсказания объясненного явления. Следовательно, все, что будет сказано в этой статье относительно логических характеристик объяснения или предсказания, применимо к любому из них, даже если речь будет идти только об объяснении или только о предсказании [...]

Однако, многие известные объяснения, в особенности в *донаучном рассуждении*, лишены этого свойства предсказания. Так,

---

<sup>1</sup> Взять, например, затраченные ими (неуспешные) усилия по распутыванию такой загадки: логически желтый карандаш – доказательство утверждения «Все вороны черные», поскольку утверждение «Все вороны черные» логически неотличимо от «Все нечерные предметы не есть вороны».

тот факт, что машина перевернулась на шоссе, можно объяснить как следствие прокола шины при высокой скорости движения. Ясно, что на основании только этой информации данный случай невозможно предсказать, поскольку эксплананс не включает ни определенных общих законов, с помощью которых можно осуществить предсказание, ни адекватных утверждений об антецедентных условиях, которые необходимы для предсказания [10, с. 93–94], (курсив мой. – Ф.Ш.).

Для Гемпеля (и всех логических позитивистов) объяснение в отсутствие предсказания не считается чем-то научно превосходящим по отношению к прогнозированию – оно вообще не научно! Оно «донаучно». Современник Гемпеля Куайн, цитируемый в последнем разделе, утверждает почти то же самое.

Донаучный характер объяснения без предсказания можно проиллюстрировать на примере феномена молнии. На протяжении многих столетий широко принятым и достаточно разработанным объяснением молнии среди северных европейцев – мертвенно белых, вооруженных топорами мужчин европейского происхождения – было убеждение, что молнии посылает им бог Тор. Для тех, кто верил в Тора, это «объяснение» обладало такой же интеллектуальной сложностью и связностью, как, скажем, теория рационального выбора (и уж точно большей развлекательной ценностью). Была у него и предсказательная ценность, поскольку в качестве целей Тор предпочитал одиноко стоящие деревья и горные вершины, поэтому таких мест следовало избегать в грозу.

При всем этом теория молнии Тора не выдержала критического испытания, когда в одном из своих миссионерских путешествий св. Бонифаций срубил, по преданию, священный дуб Тора во Фритцларе (современная Германия) в 723 г., а Тор не встал на защиту дуба. В общем смысле, знание способов появления молнии требовало знания особенностей поведения Тора (как теория рационального выбора требует знания непознаваемой полезности для политических акторов), но при этом обладало ограниченной практической полезностью.

Сравним его с научным пониманием молнии, которое было развито в XVIII в. путем (очень опасных) экспериментов Франклина в Америке и Далибара и Делора во Франции. Они установили, что молния является формой электричества. По дедукции, если молния – это электричество, то она пройдет по электрическим проводникам вроде железа и меди быстрее, чем по изоляторам вроде дерева или камня. Значит, металлические молниеотводы могут защитить здание от молнии, и это – практическое и эмпирически проверенное предсказание. А также бесспорный прорыв по сравнению с угадыванием мыслей Тора или жертвоприношением ему с целью защиты дома: Свен жертвует козла Тору; сарай Свена сгорает. Хельга устанавливает новенький молниеотвод; сарай Хельги цел. Теория электричества хороша; теория Тора – не очень.

Перенесемся в XX в., к теориям и экспериментам по искусственному генерированию молнии Теслы и Ван де Граафа и растущему эмпирическому пониманию условий появления молнии – благодаря сочетанию теории, совершенствующейся за счет наблюдения (все еще опасных) экспериментов и моделирования – и увидим, что для предсказания, где вероятнее всего молния вызовет лесной пожар и где разместить ресурсы борьбы с огнем, используются математические модели. Иногда предсказания неверны: системы погоды, а уж тем более конкретные удары молнии сложно предсказать, – но, по сравнению с попытками прочесть мысли Тора, это значительное улучшение.

Эта аналогия – не просто дешевый трюк. [...] Различение научного и мифологического (или другого ненаучного, например, фрейдистского) объяснения – одна из центральных тем для логических позитивистов. В отсутствие предсказания научное объяснение невозможно.

Для донаучных обоснований, конечно, есть свое место. Астрология<sup>1</sup> дала эмпирическую основу астрономии, алхимия – химии,

---

<sup>1</sup> И это сравнение – тоже не просто дешевый трюк: астрологии свойственны практически все компоненты легитимного научного предприятия, кроме предсказательной валидности, и задача развести астрологию и классическую науку ставилась со времен Бэкона.

и даже такой научный разум, как Ньютон, посвятил алхимии немало времени<sup>1</sup>. Более того, для моего плюралистического мировоззрения научный способ создания знания – не единственный действенный путь получить новые знания о политике. Прозрение могут дать и донаучные эвристики из теории рационального выбора, беседы с жителями зоны военных действий или поиски в пыльных архивах. Но ничто из этого не станет строго научным, даже если его так назвать. Только способность к прогнозированию сделала бы их таковыми.

*3. «Безумие – это повторять одно и то же действие каждый день и ждать разных результатов»<sup>2</sup>*

Когда что-то становится очень доступным, оно или радикальным образом изменяет нашу жизнь (транспорт в 1820–1960-е гг., антибиотики, Интернет, мобильные телефоны), или позволяет нам совершать очень глупые поступки (мировые войны, телевидение, Твиттер). По этой логике, наука развивается, когда относительно рутинные процедуры, которые можно систематически внедрить среди большого количества людей, ведут к постепенному увеличению роста знания, и наука застопоривается, когда прирост знания, который достигим рутинным способом, себя исчерпывает.

Сегодня мы находимся именно в таком «ступоре»: многие простые вещи уже освоены, а рутинизированные процедуры

---

<sup>1</sup> Если верить Википедии, Джон Мейнард Кейнс, который приобрел многие из трудов Ньютона по алхимии, считал его не «первым из века разума», а «последним из магов».

<sup>2</sup> Обычно приписывается Альберту Эйнштейну и иногда – Бенджамину Франклину; на самом деле, вероятно, написана современной писательницей Ритой Мэй Браун в 1983 г. Опять же, согласно Википедии, это парафраз из «Базового текста» Анонимных наркоманов (1981).

только усиливают смятение, поскольку любой результат можно опровергнуть, слегка видоизменив анализ тех же данных, так что даже эксперт (не говоря уже о публике) с трудом сможет различить их. Именно поэтому я верю лишь малой части того, что читаю в журналах, а это плохо.

В естественных науках есть поговорка, что нужно пытаться написать или первую статью по какой-то теме, или последнюю. В политической науке все иначе: работа Раммеля о демократическом мире была интересной, как и оригинальный анализ Онила и Рассета [11] собранной ими же базы данных. Вполне возможно, что авторы упустили что-то важное, и несколько других статей могли бы восполнить пробел. Но уже сегодня есть больше 100 ссылок на эту статью<sup>1</sup>, и большинство из них дают лишь небольшие уточнения спецификации, операционализации или методологии изначально мультиколлинеарного набора данных! Что мы получим в итоге, кроме случайных вариаций в коэффициентах и стандартных ошибках?

Конечно, далеко не все указанные работы включают в себя повторный анализ данных. Предположим, что только половина из этой сотни статей заново анализируют данные и что публикуется одна из трех написанных работ, и средняя статья появляется после двадцати (а может, и ста) попыток анализа. Это означает, что данные были повторно проанализированы около 3 тысяч раз... И это еще не самая популярная база данных!

Очевидно, что в сравнении с Американским электоральным опросом (ANES), нам в международных отношениях еще, можно сказать, повезло.

Данные нельзя анализировать до белого каления, как сталь самурайских мечей. От сгибания и расплющивания они не становятся более надежными. Теоретически в каждой базе данных содержится конечное количество информации (включая модифи-

---

<sup>1</sup> Почти 600 на апрель 2017 г. – *Прим. перев.*

кации и добавление индикаторов), хотя измерить это на удивление трудно. При компетентном анализе (обычно это так, если данные не просто скачиваются, а предварительно собираются) основную информацию можно найти в первых пяти-десяти публикациях. Пусть читатель сам найдет хотя бы один пример, когда *повторный анализ* базы данных, проведенный года через два после ее публикации, давал бы надежные новые и важные результаты (кроме тех случаев, когда предпосылки первоначального анализа были ошибочными, включая временные ряды, кластеры или перекрестное влияние).

Добавление или замена индикатора на похожий обычно тоже не имеют большого смысла (если первоначальные индикаторы опять-таки не содержали систематических ошибок измерения или пропущенных переменных). При этом методы, которые могут справиться с мультиколлинеарностью (кластерный анализ или метод главных компонент), просто проигнорируют данный факт, поскольку обнаружат латентные измерения. Хрупкие методы вроде линейной или логистической регрессии в таком случае начнут выдавать сумасшедшие результаты, меняя знаки у коэффициентов от малейших взаимодействий переменных в результате обращения и ковариации матриц. Никакого отношения к реальному миру все это не имеет.

Самое трагическое в этом – то количество данных, которое в результате остается недоисследованным. Систематический сбор данных сегодня *действительно* процветает. Количество доступных документированных и собранных данных заворачивает (ничто не совершенно, в особенности сгенерированный мной набор данных); ничего подобного просто не существовало еще тридцать лет назад. Конференция APSA в Беркли осенью 2009 г. по межнациональному качеству управления выявила около 45 наборов данных, которые могут иметь отношение к проблеме; в сборнике индикаторов с открытым исходным кодом, доступных RITF, было обнаружено почти 3000 переменных. Мы уже разрабатываем методологию для работы со всем этим объемом данных.



Кроме того, сбор данных *не является* монокультурой: мы можем получать похожие результаты на основе множества индикаторов. Но обычно этого не происходит, потому что большинство исследователей занимаются повторным анализом небольшого количества канонических баз данных, даже (!) если у тех есть известные проблемы (например, промежуточные категории в измерении демократии-авторитаризма индекса *Polity*) или если коррелирующие индикаторы превращают регрессию в горячую смесь. Поскольку у нас есть необходимые методы и никто не принуждает анализировать одни и те же данные, основную проблему я вижу в сложившихся привычках (в том числе в неправильном использовании методов снижения размерности, таких как факторный анализ и моделирование структурными уравнениями), что можно легко исправить.

#### *4. Использование сложных методов без понимания исходных предпосылок*

Недавно мне прислали на рецензию статью, использовавшую ранее мне не известный метод конкурирующих моделей риска. Статья ссылалась на первоисточник с обоснованием данного теста, я его изучил и сделал необходимые поисковые запросы в Интернете, в результате чего обнаружил, что ключевое предположение для вычисления оценки состояло в том, что «механизмы неуспеха, ведущие к определенному типу провала, действуют *независимо* друг от друга, пока не наступит неуспех» (курсив мой. – Ф.Ш.). В самой статье модели не были даже отдаленно независимыми, и я написал в рецензии, что в результате нарушения этой предпосылки результаты, вероятно, не имели смысла; на этом мое общение с конкурирующими моделями риска завершилось. Но спустя некоторое время в презентации при приеме на работу одного из наших лучших аспирантов я снова о них услышал. Причем речь шла о факторах риска, которые коррелировали друг с другом теорети-

чески и, скорее всего, практически. На мой вопрос о нарушении предпосылок модели я услышал довольно обескураживающий ответ, что «все остальные так тоже делают»<sup>1</sup>.

Коротко говоря, ЭТО – не статистические исследования, а просто кликанье мышкой. Сам метод, возможно, и устойчив к нарушению предпосылки о независимости факторов, но в статье-первоисточнике результаты очень сильно зависели от этого и при нарушении предпосылки они практически не поддавались подсчету. На литературу о возможном нарушении предпосылок (вероятно, она существует) не было ссылок ни в статье, ни в презентации. Но вполне может быть, что статистические особенности, исходя из которых нам следует использовать метод конкурирующих моделей риска в первую очередь, теряются из-за нарушения допущения независимости. Нередко сложные методы, используемые с нарушением предпосылок, работают хуже, чем простые методы с выдержанными предпосылками.

В этом смысле конкурирующие модели риска не уникальны: за годы работы я видел много случаев, когда используется сложный метод, обычно созданный в области, далекой от политической науки, и автор, по-видимому, его даже не до конца понимает, но применяет в ситуациях, которые прямо нарушают необходимые предпосылки<sup>2</sup>. В половине случаев аргумент «все так делают» убеждает редакторов – в конце концов, методологи ведь такие противные и скучные

---

<sup>1</sup> Место он все равно получил.

<sup>2</sup> Другой лично меня раздражающий пример: в 1990-х гг. было очень модным использовать дополненный тест Дики–Фуллера (ADF) для оценки возможности коинтеграции в моделях временных рядов с использованием данных о событиях. Но в этом было минимум две проблемы. Во-первых, ADF печально известен низкой мощностью, приближающейся к нулю, в случае сильно коррелирующих стационарных данных ( $p \ll 1$  при  $y_{t+1} = py_t$ ), что справедливо для многих данных о событиях. Во-вторых, процесс генерирования данных для анализа событий гарантирует ограниченность (а значит, стационарность в смысле коинтегративных альтернатив для длинных временных рядов), так что для применения ADF – при всей ненадежности его результатов – не было даже оснований.

нытики, постоянно волнуются о фундаментальных предпосылках и других бесполезных штуках. Каждая статья породит каскад таких же плохих работ – ведь делать плохо обычно легче, чем хорошо, – пока кто-нибудь не заметит, что большинство итоговых оценок в литературе не согласуются между собой и с тем же успехом могли быть произведены генератором случайных чисел, а потом мы примемся публиковать статьи уже с другими ошибками.

И вновь, Эйкен [2] уже довольно тщательно прошелся по этим вопросам. Но, судя по моему опыту, довольно безуспешно. Я просто снова это повторяю, добавляя лишь несколько наблюдений.

Сложные модели не всегда неуместны; в некоторых ситуациях они имеют очевидное преимущество перед простыми моделями. Один из самых заметных примеров – использование бинарных срезовых (cross-sectional) временных рядов в международных отношениях [12]. В международных отношениях было проанализировано много бинарных срезовых данных с временными рядами, а существующие до этого методы с легкостью могли исказить стандартные ошибки вдвое и более. Новая, хотя и сложная методология, примененная в этой работе, соответствовала теории и данным и, таким образом, была на своем месте. Другой хороший пример – рост использования в последние двадцать лет иерархических линейных моделей в случае вложенных наблюдений. Сложная техника подбора контрольной группы (matching) тоже, пожалуй, подходит, – как и использование импутации, в случаях когда она согласуется с процессом генерирования данных.

Но на каждый удачный пример найдется десяток неудачных, где можно увидеть сложность ради самой сложности в надежде, что головокружительная техника анализа (доступная всего в несколько кликов на сервере CRAN<sup>1</sup>) препроводит ваш довольно

---

<sup>1</sup> Comprehensive R Network Archive – «Всесторонний архив сети R», сеть серверов по всему миру, содержащих открытый код для установки статистического пакета R, о котором автор пишет далее. – *Прим. перев.*

посредственный анализ в один из пяти (именно так!) «священных трех топовых журналов» по политике<sup>1</sup>, а позже – к вечеринке по поводу получения пожизненного контракта. На самом деле головокружительные методы часто лишь незначительно изменяют коэффициенты и стандартную ошибку, потому что они эффективны, только если мы знаем то, чего не знаем (матрицу вариаций-ковариаций ошибок в уравнениях или функцию склонности в задаче по мэтчингу, например). Другой частый случай – головокружительный метод сводится к намного более простому вопросу вроде «Сколькими способами можно написать словосочетание “неинформативные априорные вероятности” (flat priors)»?

Тем временем увлечение сложностью ради сложности (и пожизненного контракта) вытесняет более надежные методы. Зачем? Если можно аргументировать свою позицию простым тестом о равенстве средних, то я гораздо охотнее поверю результату, потому что *t*-тест надежен и не требует много вспомогательных допущений (а ключевое допущение обычно предоставляется нам бесплатно и неоспоримо благодаря Центральной предельной теореме). Зачем делать регрессию с одними независимыми dummy-переменными, когда, на самом деле, нужна модель дисперсионного анализа (ANOVA, очень робастная, хотя и редко преподаваемая в курсах по методологии политических исследований)? А если есть относительно короткий временной ряд и теоретическое обоснование автокорреляции зависимой переменной и ошибки (как часто бывает в политических исследованиях) – можно, конечно, поклоняясь в храме Бокса, Дженкинса и Тиао, свернуть ваши переменные в трансформационные узлы, которые не распутает и мастер йоги, а можно провести обычную регрессию наименьших квадратов; в обоих случаях различить два эффекта будет

---

<sup>1</sup> Вероятно, American Journal of Political Science (AJPS), American Political Science Review (APSR), Annual Review of Political Science, Political Analysis, World Politics. – *Прим. перев.*

невозможно, зато коэффициенты регрессии хотя бы можно будет интерпретировать.

Вывод: использовать самый простой статистический метод, подходящий под характеристики теории и данных. Прямо как советовал д-р Эйкен более 10 лет назад.

### ***5. Если данные говорят с вами, значит, вы – байесианец***

На уровне преподавания и мейнстримных публикаций – но уже не на высшем уровне – мы узаконили частотную методологию статистики, порой противоречащую здравому смыслу. Это почтенное наследие непростого компромисса, образовавшегося после горячих, но сегодня почти забытых философских дебатов Фишера, Неймана, Пирсона, Сэвиджа, Уолда и других в первой половине XX в. [13], целью которого было решить проблемы, довольно далекие от политической науки. По замечанию Джила, Фишер, Нейман и Пирсон образовали свой вариант «АВВА» (расшифровывающийся как «Что угодно, кроме байесовского анализа»<sup>1</sup>), который не был даже логически последовательным (возможно, именно это обстоятельство объясняет трудности студентов при изучении статистики).

Патологические последствия внедрения частотного вывода во всех областях, кроме его первоначального применения (индукция из случайных выборок), бесконечны, и сегодня они образуют значимую часть статистической литературы (например: [14; 15]). Перечислю лишь некоторые из часто встречающихся из этих последствий в политической науке:

– исследователям чрезвычайно трудно придерживаться правильной интерпретации проверки статистической значимости.

---

<sup>1</sup> Anything But Bayesian Analysis.

Уровень  $p$  говорит только о вероятности получения результата при (обычно) совершенно нереалистичных условиях верности нулевой гипотезы. А это совсем не то, что мы хотим узнать, – обычно мы хотим знать величину эффекта независимой переменной с учетом имеющихся данных. Это байесовский вопрос, а не частотный! Вместо этого мы постоянно видим, что значение  $p$  интерпретируется так, будто оно показывает силу связи, – вот он, вездесущий Мистический Культ Звездочек и Статистической Значимости<sup>1</sup>.  $P$ -значения этого не показывают и никогда не покажут. По опыту, этой ошибки почти невозможно избежать: даже тщательный аналитик, осознающий проблему, часто осекается, когда устно описывает полученные результаты. И лучше не вспоминать, сколько чернил мы извели, исправляя эту ошибку у студентов;

– частотная парадигма, в общем ключе, не касаясь внутренних противоречий (с которыми мы как-то справлялись в течение почти века), работает достаточно хорошо в двух случаях, для которых она была изначально разработана: в случайных выборках и в реальных экспериментах. Нужно признать, что эти характеристики описывают лишь *часть* важных исследований политики, прежде всего опросы. Но есть куда более обширные области политической науки, к которым они никогда не будут применимы: почти все международные отношения, сравнительные исследования (кроме опросов) и большинство национальных политических исследований политического поведения, кроме общественного мнения (т.е. почти все исследования исполнительного, законодательного и судебного поведения). В подобной ситуации ученый исследует генеральную совокупность, а не выборку, и хотя можно бесконечно верить в невероятное: в ошибку измерения, альтернативные вселенные и т.д., лишь бы оправдать использование выборочных методов на генеральной совокупности, но основания у них разные, и об этом давно известно (см.: [16]);

---

<sup>1</sup> Пришедший на смену Культуре Максимального  $R$ -квадрата и сокрушенный, как дуб Тора, Кингом [3].

– легкость работы с разведывательной компьютерной статистикой сделала традиционные частотные тесты значимости просто бессмысленными<sup>1</sup>. Сегодня можно проверять альтернативные модели всего парой кликов мышки (или, для самых сообразительных, парой строчек кода) за микросекунды вычислений. Практически все ныне публикуемые работы сообщают лишь о верхушке айсберга из десятков, если не сотен перепробованных формулировок. Уровни значимости можно было бы скорректировать с учетом этого, но на практике так не происходит. Само количество обрабатываемой в тысячах моделей информации в рамках даже одного проекта масштаба Онила и Рассета (или ANES, или Polity, или GSS, или Евробарометра, или Correlates of War<sup>2</sup> и т.д.) делают такую коррекцию невозможной<sup>3</sup>;

– наконец, существует очень серьезное несоответствие между предположениями частотной статистики и гипотетико-дедуктивным анализом, основанным на теории («микрооснованиями», по выражению Эйкена). С теорией все в порядке; теория позволяет не терять голову и не тонуть в индикаторах вроде «количество попугаев на душу населения»... Ну да, в большинстве моделей. Но если ваша модель уже основана на теории, опровержение нулевой

---

<sup>1</sup> Ситуация изменилась радикально по сравнению с тем, что было всего пятьдесят лет назад. Мой отец получил докторскую степень по образованию достаточно поздно, но написал диссертацию по теме, которую я до сих пор считаю одной из самых умных, – «Почему диссертации не дописываются?» – и я вспоминаю, как в середине 1960-х гг. он и другие слушатели обязательного курса статистики жаловались об одном домашнем задании, которое требовало нескольких дней изнурительных расчетов (которые проводились у нас дома, и делала их моя мать). Я почти уверен, что это была множественная регрессия с выборкой гораздо меньше 100 и тремя или четырьмя независимыми переменными.

<sup>2</sup> Данные Oneal-Russett о демократическом мире, Американский электоральный опрос, проект Polity IV о политических режимах, Общее социальное обследование (GSS) в США, Корреляты войны (COW) – проект о военных конфликтах. Все данные находятся в открытом доступе. – *Прим. перев.*

<sup>3</sup> Подробнее об этом: [1].

гипотезы не скажет вам ничего нового: теория говорит, что одна переменная должна иметь *хоть какое-то* влияние на другую, иначе бы мы не включали ее в модель, так что опровержение нулевой гипотезы лишь подтверждает это. Конечно, если работать в строгой парадигме фальсификации и как-то обойти проблемы измерения и мультиколлинеарности, неточность теста статистической значимости при оценке множественных тесно связанных моделей и т.п. – и при этом доверять результатам больше, чем собственным инстинктам (снова байесовский подход!), и оценивать очередную альтернативную модель – вот тогда принятие нулевой гипотезы становится полезным. А если при множестве альтернативных формулировок переменная все же не значима, то, наверное, это достаточно хорошее подтверждение, что ее нужно отбросить, и в этом и будет заключаться вклад данной работы. Но есть здесь еще один момент. Во многих работах показано – и это стало одной из отправных точек для Куна, – что научное исследование, хотя и принимает *принцип* фальсификации, но редко продвигается в соответствии со строгими нормами фальсификации. Общая тенденция такова, что сначала проводится значительная разведывательная работа; а парадигмы сменяются, только когда доступна более совершенная альтернатива (Лакатос; это очень близко тому, что обозначается тезисом Куайна–Дюгема). Когда мы работаем не в строгой фальсификационистской парадигме – а в стохастическом царстве социального поведения, это почти всегда так, – невозможность отвергнуть нулевую гипотезу в одном-единственном случае (а именно так работает частотный подход) фактически ни о чем нам не говорит.

Итак, какова альтернатива? Очевидно: байесианский подход. На уровне элит он уже широко принят; по крайней мере, пятнадцать лет назад какой-то из августейших членов Общества политической методологии<sup>1</sup> произнес фразу: «Мы все теперь байесианцы». Как

---

<sup>1</sup> Возможно, это был Ларри Бартельс, известный другими коронными фразами.



я покажу далее, к сожалению, этот тезис еще не просочился ни в нашу педагогическую практику, ни в практику самой дисциплины. Если говорить о журналах, байесовские подходы характерны для топового журнала (по числу ссылок согласно Web of Science) *Political Analysis*, но статьи в более мейнстримных журналах *APSR* и *AJPS* – это все еще преимущественно частотная статистика.

Я верю, что в каждом аспиранте или кандидате наук, стремящемся выяснить, зачем проверять нулевую гипотезу на генеральной совокупности (подсказка: незачем), – в каждом из них пытаются вырваться наружу байесианец. Освободите ваших внутренних байесианцев!

## ***6. Хватит уже линейных моделей!***

Даже поверхностный взгляд на количественные исследования в мейнстримных журналах за последние двадцать лет<sup>1</sup> покажет, что, за исключением недавних приложений байесовского метода, политическая наука превратилась в статистическую монокультуру: практически каждый представленный анализ основывается на линейной регрессии или логит-анализе.

С линейных моделей очень хорошо начинать: они быстро вычисляются на компьютере, легко интерпретируются, асимптоты критериев оценки обладают удобными свойствами и, используя разложение в ряд Тейлора, линейная функция дает достойное первое приближение почти любых данных. У любого преподавателя количественных методов найдется папка диаграмм рассеяния, демонстрирующих, как реальные данные выстраиваются в линию, возможно – с несколькими интересными выбросами. Элитные

---

<sup>1</sup> Уточнение: до того можно было встретить достаточное количество непараметрических тестов для таблиц сопряженности, иногда – дисперсионный анализ, варианты факторного анализа и других моделей латентных переменных. Впрочем, и ANOVA, и факторный анализ, и моделирование структурными уравнениями, – линейны.

медиа вроде «The Economist» даже иногда включают в диаграммы значения коэффициентов корреляции.

Но, как подтвердит любой учитель фермерской школы<sup>1</sup>, монокультуры всегда плохо кончают: развиваются болезни и паразитизм<sup>2</sup>, а затем следует коллапс (это еще одна формулировка главной идеи данной статьи).

Проблемы данной монокультуры мы рассмотрели выше; сейчас я хочу показать, что существуют альтернативы. В соответствии с моей метафорой монокультуры, статистика в социальных науках раньше была намного богаче и креативнее с методологической точки зрения, приспособливая тесты, основанные на теории вероятности, для конкретных проблем, теорий и данных (см. например: [17; 18; 19]). Можно также утверждать, что мы отстаем от коммерческого анализа данных (см.: [20; 21]). Как бедный городской ребенок, который никогда не видел нетепличных овощей, многие исследователи в политической науке сегодня полагают, что «статистика» означает «регрессия» и, как следствие, считают, например, что статистический вывод невозможен, если число потенциальных объяснительных переменных превышает количество наблюдений. На деле почти всякий человеческий вывод происходит в подобной ситуации, и это обстоятельство становится ограничением лишь в мире линейного вывода.

---

<sup>1</sup> Подсказка: в их названии обычно есть слово «государственный».

<sup>2</sup> Паразитизм в данном контексте – это индивид, год за годом выдающий статьи следующим образом: загружает массив данных, прогоняет различные варианты моделей, пока рано или поздно не найдет какой-нибудь умеренно интересный набор значимых коэффициентов, стряпает за выходные статейку и затем через сеть придерживающихся тех же убеждений рецензентов и измотав редакторов журнала, в конце концов это публикует. При этом считает, что это и есть легитимная «нормальная наука» и не скрывает мнения, что любой, кто поступает иначе, просто наивный дурачок. Дорогой читатель, не кажется ли тебе знакомым это описание? Не знает ли каждый из нас хотя бы одного человека, подпадающего под него? Вот еще один из грязных секретов частотного анализа.

Поражает количество методов, которые мы *не* используем. В 2000 г. американские методологи политической науки посетили встречу методологической секции Международной социологической ассоциации в Кельне. Во-первых, мы были удивлены тем, насколько она огромна – социология развилась несколько иначе в землях Маркса, Дюркгейма и Вебера, чем на нашем берегу, а во-вторых – тем, как много было представлено исследований со сложной методологией, в частности с использованием анализа соответствий (СА), почти невиданного по эту сторону Атлантики, хотя СА такой же замысловатый метод сокращения размерности, что и регрессия, может исходить из разных допущений и существует в тысяче вариантов.

Другой пример – метод опорных векторов (SVM). Это – рабочая лошадка современного классификационного анализа, понятная, надежная, готовая (уже сегодня есть как минимум четыре варианта реализации в *R* и открытый код почти на всех современных языках программирования) и при этом почти отсутствующая в политическом анализе, кроме применения к обработке естественных языков.

Наконец, в чисто качественной области машинного обучения – обычно *terra incognita* для регрессионно-ориентированных исследователей – уже разработано множество алгоритмов классификационных деревьев для категориальных данных, начиная с методов ID3 и C4.5 и заканчивая CART и CHAID, – опять-таки, надежных, понятных, уже готовых и написанных в открытом коде, но фактически невидимых в политических исследованиях.

И это лишь верхушка айсберга. Беглый обзор всего трех современных текстов по компьютерному распознаванию образов [22; 23; 24] прибавляет к списку следующие методы:

- множественные вариации нейронных сетей;
- множественные вариации Фурье-анализа;
- множественные вариации метода главных компонент;
- скрытые марковские модели;

– алгоритмы последовательного поиска, функциональные, топологические алгоритмы, алгоритмы иерархической кластеризации;

– множественные вариации моделей с латентными переменными;

– генетические алгоритмы и методы имитации отжига.

Вопрос не в том, что нелинейные модели «хороши только потому, что они инновационны», нет – это было бы так же непрактично, как и «сложность ради сложности», которую я критикую выше. Вопрос в том, что все эти приемы как минимум используют нелинейные пути установления закономерностей в данных (ведь если многие вещи изменяются линейно, это еще *не* значит, что линейно *все*) и во многих случаях имеют дело с вопросами, часто встречающимися в данных о политике. Метод опорных векторов и методы деревьев решений, например, прекрасно работают, когда количество независимых переменных больше, чем количество случаев, а большинство алгоритмов кластеризации не выдвигают требований к независимости переменных. Многие из этих методов могут использовать пропущенные значения в качестве возможного классификатора, что очень пригодилось бы там, где данные не проходят тест на случайность пропусков (а это – большинство случаев для международных данных).

В этом вопросе – первый раз за статью – я все-таки оптимистичен; и большую часть этого оптимизма можно выразить одной буквой – *R*. Статистический пакет *R* стал *lingua franca* систематического анализа данных во всех сферах (включая даже анализ текстов!), – и это сломало прежние дисциплинарные барьеры, по понятным причинам созданные специализированными коммерческими статистическими пакетами. Основанный на открытом коде *R* способствовал быстрому распространению новых методов. Скоро мы достигнем точки, когда каждый новый опубликованный статистический метод будет отражен в *R*: по меньшей мере – самим автором, а если получит признание, то, возможно, и в CRAN. И наконец, последняя версия *R* находится на расстоянии в несколь-

ко кликов, и здесь нет проблем с тем, что на компьютере вашего подразделения установлена версия 8, тогда как вам необходима версия 11<sup>1</sup>.

Доступность метода, конечно, еще не гарантирует его правильного использования. Думаю, нам предстоит изменить некоторые принципы преподавания. Пока что мы пытаемся обучать методологов на пределе возможностей линейной модели, используя продвинутую эконометрику Гуджарати или Мэддала. Однако те же студенты совершенно незнакомы даже с базовыми приложениями CA, SVM или CART<sup>2</sup>. Думаю, со временем – хотя это и потребует много экспериментирования – нам придется принять подход к обучению, который будет включать больше простых и надежных методов из разных сфер (даже однофакторный дисперсионный анализ) вместо последовательного и глубокого изучения одного методологического подхода (как сегодня).

---

<sup>1</sup> Недостатком *R* является его сложность и некоторая необычность: я не единожды слышал в разных заведениях, обычно от тех, кто использует, но не разрабатывает статистические методы, следующее: «Наши аспиранты больше не изучают статистику – они тратят все время, изучая программирование в *R*».

Могу их понять: как *программист*, я считаю *R* значительным шагом вперед по сравнению с ранними альтернативами, такими как SAS и Stata. Но *R* требует мыслить в достаточно своеобразной манере, которая на практике ближе к программированию, чем к статистике. SAS, Stata и, в более раннюю эпоху, SPSS, были намного ближе к методу, и для быстрого анализа я, как многие другие пользователи *R*, скорее использую Stata.

В долгосрочной перспективе я думаю, что *R* все равно победит по той причине, что это открытый код и скоро мы получим дружественный для пользователя интерфейс, так же как Macintosh OS-X и Ubuntu Linux создали удобные для пользователя интерфейсы на базе Unix. Сам себя так определивший «опытный пользователь» будет использовать для работы командную строку; обычный пользователь даже не будет знать, что командная строка существует. Но, несмотря на различные усилия (в виде облегчающей работу оболочки RStudio, выпущенной в 2016 г. – *Прим.перев.*), мы пока еще идем к этому.

<sup>2</sup> После изучения продвинутой эконометрики студент сможет разобраться с этими методами, но не увидев их на занятиях и не поэкспериментировав с ними.

## 7. Смещение статистического и экспериментального контроля

Однажды мне пришлось участвовать в увлекательном методологическом опровержении [25; 26] статьи из *Journal of Conflict Resolution*, целью которой было заявлено установить эффективность трансцендентальной медитации на расстоянии в снижении уровня политического насилия [27]<sup>1</sup>. Хотя в статье мною было обнаружено много проблем (ровно как Марковски и Фейлсом), главной была интерпретация дополнительных независимых переменных как «контрольных».

Авторы той работы были не уникальны, просто они попались на горячем. За исключением тщательно отобранных случайных выборок – и уж точно не в генеральных совокупностях целиком – при использовании статистически независимых переменных (что в социальных науках возможно разве что в экспериментах) статистический «контроль» служит лишь для того, чтобы разбросать объясненную вариацию по (часто) случайным изменениям в оцениваемом значении параметра. Он ни в коем случае не равносителен экспериментальному контролю. Так или иначе, часто эти «контрольные переменные» вбрасываются в модель с таким видом, будто это не приносит вреда и даже поможет уберечь исследование от ошибочных выводов. Нельзя быть более далекими от истины.

Здесь мы, возможно, опять могли постепенно и неосознанно прийти к способу выражения, который, несмотря на удобство (ведь что может быть лучше рандомизированных экспериментов для установления причинных связей) и корни (значительная часть

---

<sup>1</sup> Поскольку далее я упоминаю Церковь сайентологии, новое религиозное движение, печально известное в том числе негативной реакцией на критику, отмечу, что за двадцать лет последовательного взаимодействия с сообществом сторонников трансцендентальной медитации они все без исключения проявляли себя доброжелательно, вежливо и профессионально.

оригинальных работ о частотном выводе появилась в контексте контролируемых экспериментов), просто ошибочен в современном контексте, когда мы пытаемся оценить линейные коэффициенты на основе набора коррелирующих независимых переменных, измеренных в неоднородных генеральных совокупностях. Причем мы знаем, что этот язык ошибочен, но забываем об этом, также как забывает о том, что  $p$ -значение – это не показатель силы влияния независимой переменной.

На протяжении ряда лет первым заданием, которое я давал своим студентам в курсе многомерных методов (и я не советую делать это на вводном курсе), было создать на основе базы данных международного исследования какую-нибудь совершенно нелепую модель, чтобы получились значимые коэффициенты на бессмысленном наборе независимых переменных в результате мультиколлинеарности или ложной корреляции. Ни у кого с этим проблем не было! К счастью, ни один студент не попытался опубликовать свою модель, но у меня сложилось впечатление, будто редколлегия многих журналов каждый день сталкиваются с чем-то подобным.

Другая сторона этой медали – хотя ее тоже можно было бы отнести к патологиям частотного вывода<sup>1</sup> – предположение, что статистическая значимость что-то говорит о причинно-следственных отношениях. К счастью, за последние двадцать лет мы сильно продвинулись в понимании этого вопроса – как это выразилось, например, во внимании к причинно-следственным связям на летней встрече Общества политической методологии 2009 г. в Йельском университете, однако эта ошибка все еще проникает в профессиональные дискуссии. В контролируемом и рандомизированном эксперименте сильное влияние переменной обычно переходит в предсказанное влияние на зависимую переменную (если нет ложной корреляции или неучтенных переменных), – но

---

<sup>1</sup> Как я и говорил, на самом деле в работе около двадцати пунктов критики, а не семь.

это не верно для уравнения, составленного на основе не самых стройных данных из генеральной совокупности!

Последствия этой ошибки губительны. Например, значительная часть ранних работ PITF зашла в тупик, потому что статистически значимые переменные не обеспечивали лучший уровень предсказания – и эта проблема характерна для многих исследований политических конфликтов [4]. Лишь когда методология PITF стала специально учитывать предсказательную валидность (например, проверка на разных частях выборки и классификационные матрицы), их модели смогли преодолеть проблему. Проект ICEWS, вероятно, научившись на опыте PITF, использовал предсказательную оценку в качестве критерия с самого начала.

А именно предсказание, а не объяснение, определяет статус исследования как научного.

### *Что делать?*

Несмотря на длинный список претензий, я категорически заявляю, что не предлагаю отказаться от научного метода и обратиться к беспорядочному или интуитивному познанию «Я узнаю его, когда увижу (ну, может быть, наверное)» или, что еще хуже, к постмодернистскому нарциссическому нигилизму, отрицающему саму возможность объективной реальности<sup>1</sup>. Учитывая количе-

---

<sup>1</sup> Развивая эту мысль: меня беспокоит, что эта статья будет цитироваться вне контекста, чтобы подкрепить ровным счетом противоположную моей точку зрения. Вроде «Шротт – сам Шротт, бывший президент Общества политической методологии, вторя бывшему президенту Общества политической методологии Эйкену, призывает выбросить вон большую часть количественной методологии! Стряхните пыль с Фуко, Деррида и Оксфордской энциклопедии сарказмов, выбросьте компьютер из окна и используйте вещества посылнее в качестве главного метода исследования, потому что постмодернизм снова в моде, e-e-e!» Нет, я как раз этого не имею в виду.

Если говорить серьезнее, то несколько раз меня ругали и увещевали, что я не должен говорить «всякой ерунды», потому что «они» все перевернут. Но «они» все



ство хорошо изученных патологий интуитивного человеческого размышления [28; 29], даже среди экспертов, нам нужна любая помощь, чтобы разобраться с изучением политического поведения. Я предлагаю серьезно отнестись к этой критике как стимулу для развития новой, более совершенной философии статистического вывода, которая была бы создана специально для политического анализа, а не просто перенимала то, что хорошо работало в отделе контроля качества пивоварни «Гиннес» в 1908 г.

Как было отмечено в первом разделе, ряд моих замечаний направлен на то, чтобы просто избавиться от вредных привычек, о которых мы знаем из вводного курса по методологии, как-то: модели, созданные по принципу «танцуют все!», не имеют смысла; статистическая значимость говорит нам то, что нам и не нужно было узнать; не стоит использовать методы, не подходящие к вашей теории и данным. В некоторых случаях (дрейф в донаучное объяснение в ущерб прогнозированию и толерантность к бесконечному анализу одних и тех же данных) нам нужно вернуться к основам и навести там порядок (в том числе в ленивых академических советах<sup>1</sup>). Наконец, те же технологии, которые стали проклятием в некоторых вопросах, могут помочь перейти от частотных оценок к байесовским, а также открывают перед нами многообразие новых

---

равно сделают все, что «им» хочется, независимо от того, что мы говорим и чего не говорим. «Они» не будут и читать того, что мы говорим, разве что если у этого будет умный заголовок вроде «Семь смертных грехов» и оно будет доступно на сервере APSA, а позволять всем, кто мог бы провести анализ на более высоком уровне, прятаться в болоте из страха, что наше предприятие, мол, столь хрупко, что нападение на любую его часть – это нападение на него целиком (что не верно), я не считаю особенно конструктивной стратегией. Таково мое скромное мнение.

<sup>1</sup> Имеется важное движение, долгое время подкрепляемое рекомендациями Национального научного фонда, в сторону того, что решение о бессрочных позициях должны основываться на ваших лучших N статьях, где N, как правило, находится в диапазоне от 5 до 10, что, безусловно, значительно сократит стимулы для распространения таких анализов, которые воспринимаются всеми, включая автора, как полностью излишние и в лучшем случае вносящее незначительный вклад.

техник анализа, многие из которых уже используются в прикладных проектах. Ну и, наконец, есть ряд общих вопросов, у которых до сих пор нет четкого решения, оно будет непростым и – вдохните – решение не состоит в том, чтобы ускоренно кликать мышкой<sup>1</sup>. Некоторую часть этих вопросов я сейчас попытаюсь обрисовать.

Взглянем «с высоты птичьего полета» на то, что с позиции философии науки лежит в основе количественного анализа в надежде на понимание, как мы здесь оказались, которое поможет указать путь вперед. На мой взгляд, сегодня мы завязли в незавершенной философской рамке, унаследованной вместе со многими полезными идеями от логических позитивистов, в сочетании с философски несвязным подходом частотной статистики. Выход из ситуации – в обновлении интереса к вопросам логического позитивизма с учетом обновленного к XXI в. понимания стохастических подходов, а также с фокусом на социальные науки в целом. Многое уже было сделано в последнее десятилетие в сообществах практиков качественных и смешанных методов – но, что удивительно, не в сообществе количественников. Несмотря на это, в количественном анализе есть непосредственная альтернатива частотному анализу в виде байесовского подхода, который может решить большинство противоречий частотного вывода, на которые мы по привычке закрываем глаза, растолковывая студентам, что его противоречия вовсе не противоречия. Необходимо также систематически ввести байесовский подход в преподавание. Иначе говоря, может быть, мы и застряли в болоте сегодня, но выход из него достаточно понятен.

### *Как мы оказались в таком положении*

Снова взглянем с высоты птичьего полета: после целого тысячелетия гипердедуктивного интеллектуального оцепенения, вершиной которого стала схоластика, было изобретено нечто похожее

---

<sup>1</sup> Если только не по ссылке: <http://plato.stanford.edu/>.

на современный научный метод. В эпоху позднего Возрождения разработку достаточно согласованной системы научного метода начал Бэкон около 1600 г., а более-менее завершил ее Декарт около 1640 г. Этого было достаточно, чтобы проложить путь наиболее прославленному достижению XVII в. – ньютоновской механике – а также Просвещению. На протяжении XVIII в., за исключением, пожалуй, вызова радикального скептицизма Юма, вполне релевантного и в наши дни, все развивалось в этом направлении (я рассказываю историю с позиции победителя). К XIX в. начались относительно современные попытки приложить эти методы к социальным наукам (Милль, Бентам, Маркс, Парето), хотя намного больший успех сопутствовал приложению философских критериев (Джеймс, Мах) и экспериментального метода (Вундт, Пирс, Дьюи) в науках о поведении (современная психология).

Вот в такой обстановке в начале XX в. появился Венский кружок логических позитивистов. На этом этапе полезно вспомнить, к чему логические позитивисты имели доступ, а к чему – не имели, а также их место в науке и философии. Прежде всего, общая цель логических позитивистов заключалась в том, чтобы систематизировать основания научного исследования на уровне, сравнимом с математикой (Кантор, Гильберт и, конечно же, Рассел и Уайтхед) и с обобщением физических законов (уравнения Максвелла). Кроме того, в ответ на прославление науки со времен Просвещения и еще больше – со времен Промышленной революции, Венский кружок занимался попытками различения между «классической» наукой и вызовами, прежде всего со стороны фрейдизма в психологии и марксизма – в политике<sup>1</sup>.

---

<sup>1</sup> Эти два вызова привлекли наибольшее внимание Венского кружка, хотя они были далеко не единственными претендентами носить мантию «науки». Один из наиболее успешных примеров: в конце XIX в. Мэри Бейкер Эдди решила назвать свое успешное движение по исцелению верой «христианской наукой». Нужно ли напоминать, что Американская ассоциация политической науки была основана в то же время? До появления действительно научных работ по политике Чарльза Мерриاما и его коллег из Университета Чикаго в 1920-х гг. оставалось еще два десятка лет.

Многое из того, что систематизировали логические позитивисты (что-то из ранних источников, таких как работы Бэкона и Декарта, что-то – из более поздних), останется в науке – бихевиористские концепции измерения, концепты экспериментирования, фальсификации, различие между дедуктивным и индуктивным и т.д. – смотрите первые три главы из учебника [30]. Но не менее важно отметить две критически важные вещи, которых, с нашей позиции из XXI в., логическим позитивистам *не хватало*.

Во-первых, и это главное, им не хватало развитого понимания вероятностных механизмов и рассуждения. Статистика была к тому времени относительно новой и маргинальной наукой, поскольку она считалась прикладной дисциплиной (как считали во многих университетах вплоть до конца XX в.). Венский кружок работал в рамках детерминистских систем Ньютона и Максвелла, а недетерминистские системы, как показывает тот дискомфорт, который вызывала в начале XX в. квантовая механика, в то время доверия не вызывали. Детерминизм позволил им, вслед за Расселом и Уайтхедом в математике, развернуть всю мощь логики по отношению к науке.

Во-вторых, *первоначальная* цель позитивистов зашла в тупик (или достигла своих логических пределов) в 1950-х гг., когда Карнап, Куайн, Дюгем и другие установили, что различие синтеза и анализа не может быть основано лишь на логическом основании в связи с проблемой бесконечного регресса вспомогательных предположений (*ancillary assumptions*) в теории<sup>1</sup>. Косвенным об-

---

Тенденция продолжается: успешную маркетинговую стратегию Эдди взял на вооружение в 1950-х гг. начинающий писатель-фантаст Л. Рон Хаббард, создавший впоследствии смесь двухтысячелетней гностической теологии и фрейдистского психоанализа для космического века и назвавший все это... (открывает конверт в лучах софитов) «сайентологией».

<sup>1</sup> Что отличается, но похоже на более ранние рассуждения Юма о бесконечном регрессе в установлении эмпирической валидности индукции – как в анекдоте: – «Если Земля стоит на черепахе, то на чем стоит черепаха? – А там черепахи до самого низа».

разом это привело к сегодняшним неясностям с предсказанием и объяснением. Интересно замечание Куайна:

Как эмпирик, я продолжаю считать концептуальную схему науки инструментом для предсказания будущего опыта в свете прошлого опыта. Физические объекты концептуально вводятся в ситуацию как удобные посредники, причем не путем их объяснения в терминах опыта, но просто как несводимые постулируемые сущности, эпистемологически сопоставимые с богами Гомера. Что касается меня, то я, как правоверный физик, верю в физические объекты, а не в гомеровских богов, поскольку было бы научной ошибкой думать иначе. Но с точки зрения эпистемологии физические объекты и боги Гомера отличаются только по степени, а не в принципе. Оба типа сущностей входят в наше познание только как культурные постулируемые сущности [31, с. 365]

Обратите внимание, что Куайн не говорит, что «научно» читать разум богов: Куайн как *ученый* верит в прогностические модели. Но, в связи с проблемой вспомогательных предположений, Куайн как *логик* не может доказать это. Схожая проблема возникла в математике за двадцать лет до того, когда теорема Гёделя о неполноте застопорила работу над фундаментальными подходами к математике, а к 1950-м гг. квантовая механика стала признанной альтернативой детерминизму Ньютона и Лапласа. Так что, хотя об этом можно спорить, сообщество философов науки было готово к не слишком счастливому финалу первоначальной повестки дня.

Случайность ли, но как раз в то время, когда логические позитивисты исчезают со сцены, начинают развиваться современные исследования политического поведения, что во многом было обусловлено прогрессом в программировании и вливанием денег в образование в США в годы «холодной войны». В результате вся «философия науки» до конца XX в. приняла социолого-исторический подход и начиная от Куна и Лакатоса прошла через постмодернистские войны научных культур в 1990-х гг. [32], очутившись в относительном вакууме дня сегодняшнего. В начале XXI в. мы находимся в ситуации, сравнимой

с кульминацией фильма-погони: постмодернистская машина сорвалась со скалы и теперь лежит на дне ущелья в языках пламени; все «плохие парни» (философски) мертвы<sup>1</sup>, и теперь остается лишь подняться обратно на вершину и связать все свободные концы сценария, которые остались незавершенными.

Мы уже сегодня можем отказаться от частотной статистики, поскольку – какой бы ни была ее практическая полезность – она никогда не представляла никакого логического смысла (что, возможно, может хотя бы частично объяснить пренебрежение Венского кружка к статистике или почему мы бьемся головой об одну и ту же стену каждый семестр, пытаясь убедить аспирантов, что частотный анализ имеет смысл, хотя смысла в нем нет).

В применении к социальным наукам частотный вывод страдает от трех недостатков, каждый из которых потенциально фатален. Во-первых, уже упоминавшийся компромисс Фишера, Неймана и Пирсона АВВА («Что угодно, кроме Байеса»). Там, где дело касается доверительных интервалов в случайных выборках, это совершенно обоснованно (и не надо извиняться перед аспирантами, что вы учите их доверительным интервалам, распределениям, Центральной предельной теореме, теории вероятности или тому, что линейные комбинации нормально распределенных переменных также линейны, – не выплескивайте пресловутого ребенка с водой). Но частотная статистика *как целое* не имеет логического смысла, и хотя нам удавалось уходить от этого факта полстолетия, дальше идти некуда: впереди – тупик.

Во-вторых, – что, вероятно, даже более важно для нас как политологов – просто невозможно совместить дедуктивно-гипотетический метод и частотный подход нулевой гипотезы<sup>2</sup>. Если

---

<sup>1</sup> Если они когда-либо были (философски) живы: вставьте обязательные ссылки на зомби и вампиров...

<sup>2</sup> Это одно из ряда не вполне обоснованных смелых обобщений, к сожалению, встречающихся в статье (преимущественно в контексте поддержки иногда выходящей за рамки собственно научной дискуссии борьбы яростных сторонников байесовского

у нас есть теория, определяющая наши модели, тогда *tabula rasa* нулевой гипотезы – это и интеллектуальное лицемерие, поскольку мы утверждаем, что начинаем с мысли, которой у нас, безусловно, нет, и информационная пустышка.

В-третьих, в огромном количестве случаев, где мы применяем частотные тесты статистической значимости, мы имеем дело не с выборками, а с генеральными совокупностями. В некоторых проектах (например, MID или ACLED по военным конфликтам) напрямую ставится цель описать 100% случаев и эта всеохватность считается важной чертой всего подхода, так что исследователям приходится изрядно попотеть, собирая последние 10%. В других случаях (анализ законодательства или постановлений Верховного суда) генеральная совокупность известна и, благодаря Интернету, легкодоступна. В таком случае без философской гимнастики становится сложно понять, как здесь можно использовать статистическую значимость, предназначенную для выборки. Возможно, поэтому столь многие и ошибаются, стремясь расценивать уровень значимости как силу причинной связи, а не вероятность того, что такой же результат будет получен в выборке из генеральной совокупности при условии истинности нулевой гипотезы.

### *Два открытых вопроса: какое предсказание и какое байесовство?*

Обозначив, в какой *тупик* ведет нас прежняя дорога, стоит сказать, куда нам стоит идти, и этим мы очень скоро займемся. Но предвзвешенно стоит разобраться с двумя вопросами, по которым я пока

---

подхода против традиционных подходов к статистическому выводу, основанных на проверке нулевой гипотезы). В частности, К. Гемпель, вслед за Р. Карнапом, специально обсуждает структурно сходную с дедуктивно-номологической модель индуктивно-статистического научного объяснения, лежащую в основании вероятностного статистического вывода (см.: [10, с. 147–215]). – *Прим. ред.*

выражался достаточно туманно – суть предсказания и нашего великого спасителя, байесианства, так что сначала нужно дать некоторые пояснения (относительно первого вопроса они будут намного яснее).

### *Прогнозирование чего нас интересует?*

Для логических позитивистов первостепенная важность предсказания была очевидной, но, как я отметил, они жили в мире детерминизма, а мы – нет. Предсказания в детерминистском мире могут быть очень, очень точными. Пример: аномалия «Пионеров»:

когда все известные силы, действующие на космический корабль, приняты во внимание, остается небольшая, но необъяснимая сила, которая вызывает постоянное ускорение обоих кораблей в сторону Солнца величиной  $(8,74 \pm 1,33) \times 10^{-10} \text{ м/с}^2$  [33].

Измеримая ошибка порядка  $10^{-10}$  после 30 лет наблюдений – это детерминизм<sup>1</sup>.

В социальных науках мы не в такой ситуации и никогда не будем в ней. Наши модели всегда будут содержать ошибки хотя бы по следующим причинам:

– ошибка спецификации модели – ни одна модель не содержит всех задействованных переменных;

– ошибка измерения – за очень редким исключением<sup>2</sup>, переменные будут включать в себя ошибку измерения, даже если

---

<sup>1</sup> В специфическом смысле, используемым автором. В различных разделах физики при проверке гипотез используются многократно более жёсткие критерии и стандартные уровни оценки статистической значимости, в сравнении с социальными и поведенческими науками (например, более пяти стандартных отклонений для нормального распределения). – *Прим. ред.*

<sup>2</sup> Исключения *есть*: например, я сомневаюсь, что в проекте «Корреляты войн» (Correlates of War) есть хоть какая-то ошибка, разве что опечатка. База данных THOMAS о частоте употребления слов в законодательстве, загруженном из Библиотеки Конгресса, также будет свободна от ошибки измерения, если не произойдет маловероятных технических сбоев. Но подобные ситуации необычны.



существует согласие по поводу «правильного» измерения в идеальной ситуации;

– свобода воли – как говорят коллеги из наук о поведении, изучающих крыс, в общем случае «генетически стандартизированное экспериментальное животное, подвергнутое точно контролируемым стимулам в стандартизированном лабораторном окружении, будет вести себя как ему вздумается». Безусловно, это касается и людей и может вылиться во всякие «эффекты колеи»;

– квазислучайная структурная ошибка – мы имеем дело с комплексными и хаотическими системами, по крайней мере при некоторых комбинациях параметров.

Таким образом, поскольку мы не можем зависеть от детерминистского определения прогноза, а утверждение о том, что наблюдаемый уровень «значимо отличается» для многих ситуаций тоже не годится, что нам остается? На мой взгляд, есть минимум три возможности.

1. Использовать норму, возникшую в метеорологии – естественной науке, которой свойственны ошибки спецификации модели, измерения и структурные ошибки, но отсутствует проблема свободы воли (если, конечно, не верить в теорию Тора), а именно: предсказанная вероятность и наблюдаемое проявление (в случайных наборах данных или вне выборки) на графике примерно совпадают с линией 45 градусов. Это значит, что гроза, предсказанная с вероятностью 20%, будет наблюдаться 20% всего времени. Этот подход применяется в политическом исследовании RITF и других (см.: [34]) и требует логит-моделей, которые на выходе дают вероятность.

2. Предсказанное влияние изменения значений переменных должно соответствовать распределению изменения, предсказанному по модели. Это можно сделать таким простым способом как оценить последствия стандартной ошибки оценки в частотной модели (она будет распределена нормально), или сложно: через апостериорную вероятность байесовской оценки. В идеале это можно оценить в

естественных экспериментах вне выборки, но в их отсутствие – с помощью подвыборок из генеральной совокупности.

3. Во многих классификационных моделях критерием становится ROC-кривая<sup>1</sup> (этот метод использован в PITF). В частности, этот подход очень хорош для определения, насколько наша модель лучше случайной модели (например, этой проверки астрология не проходит).

Этот список, конечно, не полон, но смысл в том, что мы не получили работающую вероятностную модель объяснения ни от логических позитивистов, ни от сторонников частотной статистики. И мы знаем больше, чем они. Например, теория хаоса показала, что даже очень простая нелинейная детерминистская система будет вести себя внешне случайным образом при некоторых комбинациях параметров. Логические позитивисты не знали о теории хаоса: Пункаре получил первые результаты в 1880-х гг.; впервые они были включены в теорию вероятности в первой половине XX в.; признание применимости полученных моделей в реальном мире пришло лишь с поздними работами Лоренца и Мандельброта в 1960-х гг., а изучением темы начали заниматься только в 1980-х, т. е. намного позже зенита славы логических позитивистов. То же можно сказать и о развивавшейся в 1980-е гг. теории сложности. Иными словами, даже если бы социальные системы были детерминистскими (как считали логические позитивисты), поведение, которое мы наблюдаем, все равно *казалось бы* неотличимым от поведения стохастической системы.

---

<sup>1</sup> Кривая характеристики приёмника, используемая для оценки качества классификации, описания процессов обнаружения сигнала в психофизике, в машинном обучении, для сравнения предсказаний разных моделей и т.п. Площадь под кривой в целом характеризует качество получаемой классификации по отношению к случайному угадыванию. Автор избегает детального обсуждения проблем, с которыми сталкивается этот подход, в частности, в силу чувствительности к шуму. – *Прим. ред.*

## *Народное байесианство или WinBUGS-байесианство?*

Байесианство решает многие проблемы: это логически последовательная идея, и, следовательно, она *может* служить основой для настоящей теории научного исследования; она решает вопрос интеграции теории и данных и соответствует тому, как люди мыслят (а это не такое уж и маленькое преимущество, если вы пытаетесь разработать модели поведения мыслящих людей); она решает парадокс логического позитивизма о желтых карандашах и черных воронах, а также успешно справляется с ситуацией, когда исследователи считают одни случаи более интересными, чем другие; наконец, она дает понятный метод интеграции неформальной априорной информации с систематическим анализом данных.

Недостатком байесианских подходов является сложность математики и вычислений. Последняя имеет чисто технологическое решение, хотя по правде говоря перспектива ждать 48 часов результатов от программы WinBUGS вместо того, чтобы выполнить одну регрессию, не очень привлекает. Кроме того, хотя многие количественники говорят о байесианстве, пока что информативные априорные вероятности мало используются.

Вопрос стоит так: нужно ли нам строгое байесианство или нам достаточно «народного байесианства» [35], где отбрасываются только самые проблематичные стороны частотной статистики: тирания (и бесполезность) теста проверки значимости в лицемерном мире нулевых гипотез и несочетаемость основанных на выборке предположений с анализом генеральной совокупности, характерном для многих разделов политической науки. Этих ограничений можно избежать, даже если учесть хотя бы некоторые из следствий теоремы Байеса.

С этим связана задача разработки видов вывода и интерпретации для широкого класса закономерностей в данных, которые можно обнаружить с помощью частотного или байесовского анализа. По

сути, мы уже делаем это при моделировании на данных генеральной совокупности, например, на данных проекта «Корреляты войн», поскольку частотные интерпретации этих выводов логически не последовательны (но популярны) и рассчитывают описательную, а не предсказательную статистику<sup>1</sup>. Но пока что эти интерпретации не включены в компьютерные методы распознавания закономерностей. Например, одна из причин, по которым в США редко используется анализ соответствий, – это неопределенность по поводу того, касается он описания или предсказания. При этом многие из методов «data mining», или «добычи данных»<sup>2</sup>, имеют намного большую предсказательную способность, чем господствующие сегодня частотные подходы.

### *Что дальше?*

На мой взгляд, нужно рассмотреть еще два взаимосвязанных, но до сих пор еще рассматриваемых отдельно вопроса. Эти события должны происходить одновременно, хотя на сегодняшний день они, похоже, происходят по отдельности.

1. Если частотная статистика мертва или, в лучшем случае, это тупик, то давайте ее похороним. Байесовская альтернатива – у

---

<sup>1</sup> Если только не делать выводов о невошедшем в выборку будущем; но количественники из международных отношений, кажется, не любят предсказаний?

<sup>2</sup> Это понятие имеет уничижительный оттенок в политической науке, но остается простым описанием во многих других науках [20]. Разница частично заключается в том, что коммерческие методы, использующие всесторонний анализ данных – например, рекомендательные системы Amazon, Netflix, – постоянно валидируются на внешних наблюдениях. В фиксированной совокупности наблюдений data mining напоминал бы старую острогу Джона фон Неймана о том, что «с четырьмя параметрами я могу создать модель слона, а с пятью – заставить его даже покачивать хоботом», тогда как при использовании данных, не вошедших в изначальную выборку, введение дополнительного параметра – это не преимущество, а определенное *обязательство*: растет риск создать сверхточную (overfitting) модель, что резко ограничивает простую аппроксимацию данных.

всех на виду, в последние двадцать лет именно ей была посвящена большая часть статистических исследований, и на уровне элит ее преимущество полностью признается.

Проблема в том, что мы до сих пор продолжаем *обучать* статистике исходя из компромисса АВВА: байесовский подход выиграл высоты, но не окружающие поля, а поля – это места, где большинство людей, будь то студенты или практикующие, все еще живут. Так мы приближаемся к классическому различению знания для простолюдинов и для посвященных. Мы учим частотному анализу, но хотим современной науки?! Забудьте о частотах, мы же байесианцы. Это нездорово. Более того, сами байезианцы, увлеченные чисто техническими достижениями – вы-вы, Общество политической методологии, – почти ничего не сделали, чтобы изменить эту ситуацию: все главные учебники для бакалавриата сегодня частотные<sup>1</sup>.

И этот педагогический вызов весьма серьезен. Сегодняшний выпускник, обученный по Кингу, Кеохейну и Вербе и по учебникам по эконометрике Мэддала, считается достаточно хорошо подготовлен для использования практически всех статистических методов, встречающихся в основных политических науках. Просто добавить к этому Джеринг для качественных методов, Кинга, Кеохейна и Вербу, кроме частотной части, Джилл – для байесовских методов [37], Мэддала – для продвинутых линейных моделей, а еще Теодорис и Кутрумбас – для компьютерного распознавания

---

<sup>1</sup> См.: [36]. Мой вывод из их результатов – список популярных учебников слишком короток: пять авторских коллективов (Pollock; Johnson, Reynolds and Mycoff; Le Roy and Corbett; Carlson and Hyde; Babbie) отвечают за 50% из 168 ответов. Авторы исследования – противоположного мнения, они отмечают: «Хотя опрос охватил относительно небольшую выборку курсов, результаты предполагают, что преподаватели базовых (score and methods) курсов считают подходящими для своих студентов широкий круг учебников. Хотя можно было дать более одного ответа, 106 респондентов образовали список из 71 различных названий. Это разнообразие подсказывает, что преподаватели нашли множество подходящих инструментов для передачи содержимого курса».

паттернов, ну и, конечно, один-два спецкурса по временным рядам и иерархическим моделям (не говоря об дисперсионном анализе и анализе соответствий) – это нереально. Лучше использовать избирательный подход с упором на основы, а не на конкретные приемы анализа.

Проблема, к сожалению, действительно сложнее, чем «добавить тему здесь, убрать там». Если вы так хотите поступить, то почему не отказаться и от Мэддала как сторонника частотного метода. Пока многие из нас продолжают совершать странные действия с частотными моделями – и отсутствие ответа на критику в литературе [2; 3] подмывает сказать: «Как жаль, мы предупреждали, но вы не обращали внимания, и теперь нам придется забрать ваши частотные игрушки», – методы частотного вывода можно будет и дальше применять хотя бы для разведывательного анализа генеральных совокупностей, в опросах и экспериментах, для которых они подходят. Учитывая, что большая часть публичного облика политической науки как науки – это проведение опросов, нам необходимо сохранить некоторые элементы частотности. Но я не уверен, что частотная статистика нам нужна там, где дело касается причин и прогнозов: здесь нам, пожалуй, целиком придется перенять байесианство и его методы. Но более важно, что выбор должен быть сделан, хотя ответы не всегда очевидны.

В худшем случае, как отметил Макс Планк в своей научной автобиографии, «новая научная истина завоевывает сердца не когда оппоненты после долгих споров прозревают, а когда они в конце концов умирают, а новое поколение растет уже с новым знанием». Хотелось бы думать, что прогресс может быть достигнут в моей жизни, но по крайней мере нам нужно начать предоставлять материалы, чтобы новое поколение могло работать с ними<sup>1</sup>.

---

<sup>1</sup> Новый финансируемый Национальным научным фондом портал социальных наук в области методологии OPOSSEM может в конечном счете дать такую возможность. Учитывая те ледниковые темпы, с которыми издатели готовы рискнуть

2. Нужно возобновить работу в области философии науки с того места, где остановились в 1950-х гг. логические позитивисты. Нам нужна *связная теория* стохастического вывода (почти наверняка основанная на байесианстве, не на частотах) относительно социального поведения, которая была бы свободна от обломочных пород, накопленных в попытках построить теорию на основе логического детерминизма.

Многое *уже* делается в этом ключе – например, Геррингом, Кольером, Беннетом, Джорджем, Геддесом, Элманом и другими, хотя в основном в *качественных* исследованиях. Эти усилия по большей части не инкорпорировали байесовские идеи, хотя, например, многие аргументы, которые я приводил здесь в отношении байесовских интерпретаций доказательств, можно найти в десятилетней давности критике МакКеона [35] в адрес Кинга, Кеохейна и Вербы. Проблема в том, что байесовская логика сегодня находится в количественной области, где философская мысль практически отсутствовала в последние тридцать лет, а там, где мы ее находим, она очень слаба (почти все без исключения наиболее слабые доклады на Летних встречах по политической методологии – «концептуальные»). Как и многие другие, я сильно увлекся новыми техническими приемами, которые появляются каждый год, и теперь сложно сесть за Куайна и Поппера (не говоря уже о Бэконе и Декарте). Но я сам признаю правоту теоретиков «качественников» в философской безосновательности разделения дисциплины по принципу «уровня измерений»: нам нужна *единая* научная философия исследования, которая бы включала как количественные, так и качественные переменные.

---

коммерческой разработкой новых текстов по методам для рынка политологии, вероятность того, что издатель возьмет на себя финансовый риск разработки байесовского учебного плана, который мы вряд ли примем с первого, а то и с пятого раза, очень близка к нулю. Но подход OPOSSEM с открытым исходным кодом/открытым доступом может преуспеть.

В теоретическом плане многое еще предстоит сделать, чтобы опровергнуть и радикальный скептицизм Юма (и несвязный скептицизм постмодернистов), и настолько же ограничивающие требования частотной статистики относительно больших выборок (в том числе у Кинга, Кеохейна и Вербы). Ведь люди получают информацию постепенно, из небольших выборок (с которыми напрямую работает байесовский вывод) и из ситуаций, где число потенциальных объяснительных переменных намного больше числа наблюдаемых случаев (и где корреляция между индикаторами может быть и преимуществом, и ограничением). Отвергать данные факты – значит принудительно ограничивать себя небольшим кругом линейных аналитических моделей и моделей с контролируемой случайной выборкой.

Не исключено, что мы можем быть ближе к работоспособным решениям, чем может показаться, – возможно, сегодня мы просто переживаем период «предраассветной темноты». В течение периода, когда мы отбивались от натиска зомби-нигилистов, философия науки продвигалась в направлении, вполне совместимом с нашими идеями (и хорошо бы включить это в наше преподавание, а не делать вид, что Кун – это «последняя из новинок»). В частности, в рамках «научного реализма» были введены элементы прагматизма, которые прибавляют гибкости к строгой логике Венского кружка, и сегодня социальные науки уже намного ближе к науке, поэтому выше вероятность того, что будут рассмотрены и специфические для социальных наук проблемы.

### *Почему это важно*

В заключение подниму еще два вопроса: почему это важно и нужно ли серьезно относиться к критике (и более ранней критике Эйкена по близким пунктам) из уст ученого предпенсионного возраста?

В чем вообще проблема со статусом «науки»? Почему, вслед за примером Мэри Бейкер Эдди, если движение трансцендентальной



медитации стремится утвердить свою научную валидность, не разрешить ему сделать это? Здесь я оказываюсь в такой же ситуации, как члены Венского кружка, атакующие ненаучные заявления фрейдистской психологии и политического марксизма, или современного скептического движения (Саган, Азимов, Гарднер, Рэнди, Куртц и др.) в популярной культуре, выступающие против, скажем, различных культов НЛЮ. Отличительно научный подход, история которого продолжается с XVII в., дал нам мир, которым я скорее доволен – включая политическую науку, несмотря на ее немного преждевременное принятие своего титула, – и нет, извините, но далеко не все вписывается в категорию «науки» и не все позволено.

При этом я все же плюралист в том, что касается *знания* о мире политики. Научный метод – один из многих методов постижения мира, часто не самый эффективный. Политическая жизнь требует понимания (и, кстати, в меру точного предсказания), и, как показывают некоторые исследования эволюционных психологов о разнице между нашими способностями и другими приматами, развитие политических структур в значительной степени связано с развитием умственных способностей. Что-то в политическом поведении бессознательно – мы делаем вещи, не полностью осознавая причины (в этом Фрейд был прав), но, как отметил социолог Уильям Ф. Уайт, многие вещи мы делаем по тем причинам, по которым мы и думаем, что мы их делаем. Например, за последние десять лет было много исследований о важности норм справедливости и склонности людей к «иррационально» точной мести, когда эти нормы попираются. Впрочем, еще у Гомера<sup>1</sup> эта тема тоже раскрыта достаточно полно.

До тех пор, пока кто-то может меня убедить, что может хорошо описать эмпирическую реальность (и признает, что эмпирическая реальность существует), имеет какой-то достаточно системати-

---

<sup>1</sup> У греческого поэта Гомера, не у Гомера Симпсона. Но если подумать, Симпсон тоже подойдет.

ческий способ понимания этого мира и может рассказать о нем настолько связно, чтобы убедить экспертов в своей области, я буду прислушиваться. Иными словами, я бы принял в свой мир Джона Миршаймера, атакующего математические модели, даже если он не принял бы меня в свой.

Однако это не делает все знание научным, а плюралистическое мировоззрение этого и не требует: в конце концов, мы же различаем физиков и инженеров, биологов и докторов. В то же время научный подход к исследованию политики может показать нам то, чего другие подходы не могут, и показать прозрачно, с воспроизводимыми результатами, да так, что это было бы не под силу ни одному индивидуальному гению. Аналитические системы PITF и ICEWS достигли устойчивой внешней предсказательной точности, намного превышающей точность конкретных экспертов. Создать их было непросто; тридцать лет назад их нельзя было создать в принципе, даже слаженными усилиями, но благодаря систематическому сбору данных и применению статистических методов сегодня такие системы возможны.

Вероятно, нам нужно признать, что в словосочетании «политическая наука» второе слово было скорее маркетинговой уловкой столетней давности<sup>1</sup> и стоит вернуться к узкому определению науки, которое давало бы четкое и связное определение научного метода (которое, повторяю, не основано на делении по уровням измерения)

---

<sup>2</sup> И престиж этой вывески, вероятно, поблек не только из-за Рона Хаббарда: технологический прогресс последней четверти XX в., более всего сказавшийся на повседневности в развитых странах – личные компьютеры и компьютерные сети, – практически полностью был результатом прикладной инженерии и креативного искусства, а не прогресса в фундаментальной науке (примеры: компьютерное программирование в целом и чисто концептуальные прорывы, такие как графический интерфейс, Интернет, коммутация пакетов и iPod). В противоположность этому, в конце XIX в. движущей силой революций в электромагнитической коммуникации, промышленной химии и гибридных агрокультурах была фундаментальная наука.

и которое не было бы столь широким, что охватило бы всех, кто преподает на кафедре политических наук. В таком случае интерпретативистские методы – хотя они и помогают понять политическое поведение – не научны, потому что не могут обеспечить систематической и воспроизводимой методологии. Теория рационального выбора, несмотря на математический аппарат<sup>1</sup> и эвристическую ценность в дилемме заключенного или играх типа «Ультиматум», в большинстве случаев тоже не достигает прогностической валидности. Узкое определение позволило бы нам улучшить уровень исследований и преподавания, а также отделить действительно научные зерна от плевел. Именно такой подход часто подразумевается, например, при финансировании исследований Национальным научным фондом (по крайней мере на моем немалом опыте)<sup>2</sup>.

Ну и, наконец, стоит ли верить еще одной смеси манифеста и диатрибы от кого-то, кого (as a registered old fart) в традиционной культуре уже давно спустили бы на воду в каноэ с дырой в борту, отчасти повторяющей уже сказанное Эйкеном [2]? Три мысли на этот счет. Во-первых, я был бы очень рад, если бы вместо того чтобы Эйкен или я призывали других отказаться от моделей по принципу «танцуют все!», я стал бы свидетелем восстания молодого поколения под лозунгом «Долой линейные модели с 12 переменными!» В какой-то степени именно это означает отрицание некоторыми

---

<sup>1</sup> Вы уж простите, но простого использования математики недостаточно: астрология это тоже умеет. И это не банальность.

<sup>2</sup> Здесь есть интересная параллель... с Никейским собором 325 г. Несмотря на его теологическую важность в долгосрочной перспективе – никейский Символ веры и его производные до сих пор произносятся в христианских конгрегациях по всему миру – Собор был созван политической властью для решения политической проблемы: император Константин решил оказать денежную поддержку христианским организациям и для этого должен был определить, кто подходит под определение христианина, а кто – нет; а когда имеешь дело с налоговым инспектором Римской империи, просто назваться «христианин» недостаточно. Так вот, научный фонд исполняет схожую роль. Это не означает, что я предлагаю ученым писать какой-то свой Символ веры.

молодыми исследователями распространенных количественных подходов, но, к сожалению, они довольно часто выплескивают с водой не только ребенка, но заодно и ванночку.

Во-вторых, как рецензент и оппонент, я действительно часто сталкиваюсь с этими проблемами и чего только не перевидал на своем веку. Разум стоит беречь, а не тратить его на 3001-й анализ одной и той же базы данных. Возможные параллели со схоластикой просто пугают: начиная с 1650 г. европейское общество переживало невиданный политический, культурный и технологический расцвет, а университеты до конца XIX в. продолжали жить, как в Средневековье... Я совершенно уверен в способности академического мира противиться изменениям, но я не считаю, что это всегда хорошая черта.

В конце концов, существует первый закон Кларка, который гласит: если заслуженный, но престарелый ученый говорит, что нечто возможно, он почти наверняка прав; если же он говорит, что нечто невозможно, он почти определенно ошибается. Хотя сегодня не все гладко, но возможный выход достаточно очевиден. Более того, мы уже существенно продвинулись в его направлении. Поэтому я утверждаю, что решение поставленной проблемы возможно.

#### ЛИТЕРАТУРА

1. *Schrodt Ph.* Beyond the Linear Frequentist Orthodoxy // *Political Analysis*. 2006. Vol. 14. No. 3. P. 335–339.
2. *Achen Ch.* Toward a New Political Methodology: Microfoundations and ART // *Annual Review of Political Science*. 2002. Vol. 5. P. 423–450.
3. *King G.* How Not to Lie with Statistics: Avoiding Common Mistakes in Quantitative Political Science // *American Journal of Political Science*. 1986. Vol. 30. No. 3. P. 666–687.
4. *Ward M, Greenhill B., Bakke K.* The Perils of Policy by P-Value: Predicting Civil Conflicts // *Journal of Peace Research*. 2010. Vol. 47. No. 5. P. 1–13.
5. *Хаффа Д.* Как лгать при помощи статистики. М.: Альпина Паблшер, 2015.
6. How to Lie with Statistics. URL: [https://en.wikipedia.org/wiki/How\\_to\\_Lie\\_with\\_Statistics](https://en.wikipedia.org/wiki/How_to_Lie_with_Statistics) (date of access: 15.05.2017).
7. *Goldstone J., Bates R., Epstein D., Gurr T., Lustik M., Marshall M., Ulfelder J., Woodward M.* A Global Model for Forecasting Political Instability // *American Journal of Political Science* 2010. Vol. 54. No. 1. P. 190–208.

8. *O'Brien S.* Crisis Early Warning and Decision Support: Contemporary Approaches and Thoughts on Future Research // *International Studies Review*. 2010. Vol. 12. No. 1. P. 87–104.
9. *Hempel C.* Explanation and Prediction by Covering Laws // *Hempel C.* The philosophy of Carl G. Hempel: studies in science, explanation, and rationality / Ed. J. Fetzer. Oxford: Oxford Univ. Press, 2001. P. 69–86.
10. *Гемпель К.Г.* Логика объяснения. М.: Дом интеллектуальной книги; Русское феноменологическое общество, 1998.
11. *Oneal J., Russett B.* Assessing the liberal peace with alternative specifications: Trade still reduces conflict // *Journal of Peace Research*. 1999. Vol. 36. No. 4. P. 423–442.
12. *Beck N., Katz J., Tucker R.* Taking Time Seriously: Time-Series-Cross-Section Analysis with a Binary Dependent Variable // *American Journal of Political Science*. 1998. Vol. 42. P. 1260–1288.
13. *Gill J.* The Insignificance of Null Hypothesis Significance Testing // *Political Research Quarterly*. 1999. Vol. 52. No. 3. P. 647–674.
14. *Freedman D.* *Statistical Models: Theory and Practice*. Cambridge: Cambridge Univ. Press, 2005.
15. *Statistical Models and Causal Inference: A Dialogue with the Social Sciences* / Ed. D. Freedman, D. Collier, J. Sekhon, Ph. Stark. Cambridge: Cambridge Univ. Press, 2009.
16. *The Significance Test Controversy: A Reader* / Ed. D. Morrison, R. Henkel. New Brunswick, NJ: Transaction Publishers, 1970.
17. *Anderson T. W.* *The Statistical Analysis of Time-Series*. New York: Wiley, 1958.
18. *Lazarfeld P.* Some Remarks on Typological Procedures in Social Research // *Zeitschrift fuer Sozialforschung*. 1937. Vol. 6. P. 119–139.
19. *Richardson L.* *Statistics of Deadly Quarrels*. Chicago: Quadrangle, 1960.
20. *Data, Data Everywhere: A Special Report on Managing Information* // *The Economist*. 27 February 2010.
21. *Schrodt Ph.* Reflections on the State of Political Methodology // *The Political Methodologist*. 2009. Vol. 17. No. 1. P. 2–4.
22. *Duda R., Hart P., Stork D.* *Pattern Classification*, 2nd ed. New York: Wiley, 2001.
23. *Bishop Ch.* *Pattern Recognition and Machine Learning*. New York: Springer, 2006.
24. *Theodoridis S., Koutroumbas K.* *Pattern Recognition*, 4th ed. London: Springer, 2009.
25. *Schrodt Ph.* A Methodological Critique of a Test of the Effects of the Maharishi Technology of the Unified Field // *Journal of Conflict Resolution*. 1990. Vol. 34. No. 4. P. 745–755.

26. *Markovsky B., Fales E.* Evaluating Heterodox Theories // *Social Forces*. 1997. Vol. 76. No. 2. P. 511–525.

27. *Orme-Johnson D., Alexander C., Davies J., Chandler H., Larimore W.* International peace project in the Middle East: The effects of the Maharishi Technology of the Unified Field // *Journal of Conflict Resolution*. 1988. Vol. 32. P. 776–812.

28. *Vertzberger Y.* The World in their Minds: Information Processing, Cognition and Perception in Foreign Policy Decision Making. Stanford: Stanford Univ. Press, 1990.

29. *Tetlock Ph.* Expert Political Judgement. Princeton: Princeton Univ. Press, 2005.

30. *King G., Keohane R., Verba S.* Designing Social Inquiry. Princeton: Princeton Univ. Press, 1994.

31. *Куайн В. О.* Две догмы эмпиризма // Куайн В. О. Слово и объект. М.: Логос, 2000. С. 342–368.

32. *Sokal A.* Fashionable Nonsense: Postmodern Intellectuals' Abuse of Science. New York: Picador, 1998.

33. Pioneer anomaly. URL: [https://en.wikipedia.org/wiki/Pioneer\\_anomaly](https://en.wikipedia.org/wiki/Pioneer_anomaly) (date of access: 15.05.2017).

34. *King G., Zeng L.* Improving Forecasts of State Failure // *World Politics*. 2001. Vol. 53. No. 4. P. 623–658.

35. *McKeown T.* Case Studies and the Statistical Worldview: Review of King, Keohane, and Verba's Designing Social Inquiry: Scientific Inference in Qualitative Research // *International Organization*. 1999. Vol. 53. No. 1. P. 161–190.

36. *Turner Ch., Thies C.* What We Mean by Scope and Methods: A Survey of Undergraduate Scope and Methods Courses // *PS: Political Science and Politics*. 2009. Vol. 42. No. 2. P. 367–373.

37. *Gill J.* Bayesian methods: a social and behavioral sciences approach. Boca Raton, FL: Chapman and Hall, 2003.

***Schrodt Philip***

Parus Analytical Systems, Charlottesville

Translated by:

***Shirokanova Anna***

National Research University Higher School of Economics (NRU HSE),  
Moscow, Saint Petersburg, ashirokanova@hse.ru

**Seven deadly sins of contemporary quantitative political analysis**

A combination of technological change, methodological drift and a certain degree of intellectual sloth and sloppiness, particularly with respect to philosophy of science, has allowed contemporary quantitative political methodology to accumulate a series of highly dysfunctional habits that have rendered a great deal of contemporary research more or less scientifically useless. The cure for this is not to reject quantitative methods – and the cure is most certainly not a postmodernist nihilistic rejection of all systematic method – but rather to return to some fundamentals, and take on some hard problems rather than expecting to advance knowledge solely through the ever-increasing application of fast-twitch muscle fibers to computer mice.

In this paper, these “seven deadly sins” are identified as

1. Kitchen sink models that ignore the effects of collinearity;
2. Pre-scientific explanation in the absence of prediction;
3. Reanalyzing the same data sets until they scream;
4. Using complex methods without understanding the underlying assumptions;
5. Interpreting frequentist statistics as if they were Bayesian;
6. Linear statistical monoculture at the expense of alternative structures;
7. Confusing statistical controls and experimental controls.

The answer to these problems is solid, thoughtful, original work driven by an appreciation of both theory and data. Not postmodernism. The paper closes with a review of how we got to this point from the perspective of 17th through 20th century philosophy of science, and provides suggestions for changes in philosophical and pedagogical approaches that might serve to correct some of these problems.

*Key words:* Bayesian approaches, political analysis, quantitative methods

## References

1. Schrodth Ph. "Beyond the Linear Frequentist Orthodoxy", *Political Analysis*, 2006, 14 (3), 335–339.
2. Achen Ch. "Toward a New Political Methodology: Microfoundations and ART", *Annual Review of Political Science*, 2002, 5, 423–450.
3. King G. "How Not to Lie with Statistics: Avoiding Common Mistakes in Quantitative Political Science", *American Journal of Political Science*, 1986, 30 (3), 666–687.
4. Ward M, Greenhill B., Bakke K. "The Perils of Policy by P-Value: Predicting Civil Conflicts", *Journal of Peace Research*, 2010, 47 (5), 1–13.
5. Huff D. *How to lie with statistics* (transl., in Russian).M.: Alpina Publisher, 2015.
6. How to Lie with Statistics. URL: [https://en.wikipedia.org/wiki/How\\_to\\_Lie\\_with\\_Statistics](https://en.wikipedia.org/wiki/How_to_Lie_with_Statistics) (date of access: 15.05.2017)
7. Goldstone J., Bates R., Epstein D., Gurr T., Lustik M., Marshall M., Ulfelder J., Woodward M. A "Global Model for Forecasting Political Instability", *American Journal of Political Science*, 2010, 54 (1), 190–208.
8. O'Brien S. "Crisis Early Warning and Decision Support: Contemporary Approaches and Thoughts on Future Research", *International Studies Review*, 2010, 12 (1), 87–104.
9. Hempel C. "Explanation and Prediction by Covering Laws", in: Hempel C. *The philosophy of Carl G. Hempel: studies in science, explanation, and rationality*. Oxford: Oxford Univ. Press, 2001, 69–86.
10. Hempel C., Oppenheim P. "Studies in the Logic of Explanation", *Philosophy of Science*, 15, 2: 135-175.
11. Oneal J., Russett B. "Assessing the liberal peace with alternative specifications: Trade still reduces conflict", *Journal of Peace Research*, 1999, 36 (4), 423–442.
12. Beck N., Katz J., Tucker R. "Taking Time Seriously: Time-Series-Cross-Section Analysis with a Binary Dependent Variable", *American Journal of Political Science*, 1998, 42, 1260–1288.
13. Gill J. "The Insignificance of Null Hypothesis Significance Testing", *Political Research Quarterly*, 1999, 52 (3), 647–674.
14. Freedman D. *Statistical Models: Theory and Practice*. Cambridge: Cambridge Univ. Press, 2005.
15. Freedman D., Collier D., Sekhon J., Stark Ph. (eds.) *Statistical Models and Causal Inference: A Dialogue with the Social Sciences*. Cambridge: Cambridge Univ. Press, 2009.
16. Morrison D., Henkel R. (eds.) *The Significance Test Controversy: A Reader*. New Brunswick, NJ: Transaction Publishers, 1970.
17. Anderson T. W. *The Statistical Analysis of Time-Series*. New York: Wiley, 1958.



18. Lazarfeld P. Some Remarks on Typological Procedures in Social Research // *Zeitschrift Fuer Sozialforschung*. 1937, 6, 119–139.
19. Richardson L. *Statistics of Deadly Quarrels*. Chicago: Quadrangle, 1960.
20. “Data, Data Everywhere: A Special Report on Managing Information”, *The Economist*. 27 February 2010.
21. Schrodth Ph. “Reflections on the State of Political Methodology”, *The Political Methodologist*, 2009, 17 (1), 2–4.
22. Duda R., Hart P., Stork D. *Pattern Classification*, 2nd ed. New York: Wiley, 2001.
23. Bishop Ch. *Pattern Recognition and Machine Learning*. New York: Springer, 2006.
24. Theodoridis S., Koutroumbas K. *Pattern Recognition, 4th ed*. London: Springer, 2009.
25. Schrodth Ph. “A Methodological Critique of a Test of the Effects of the Maharishi Technology of the Unified Field”, in: *Journal of Conflict Resolution*, 1990, 34 (4), 745–755.
26. Markovsky B., Fales E. “Evaluating Heterodox Theories”, *Social Forces*, 1997, 76 (2), 511–525.
27. Orme-Johnson D., Alexander C., Davies J., Chandler H., Larimore W. “International peace project in the Middle East: The effects of the Maharishi Technology of the Unified Field”, *Journal of Conflict Resolution*, 1988, 32, 776–812.
28. Vertzberger Y. *The World in their Minds: Information Processing, Cognition and Perception in Foreign Policy Decision Making*. Stanford: Stanford Univ. Press, 1990.
29. Tetlock Ph. *Expert Political Judgement*. Princeton: Princeton Univ. Press, 2005.
30. King G., Keohane R., Verba S. *Designing Social Inquiry*. Princeton: Princeton Univ. Press, 1994.
31. Quine W. Two Dogmas of Empiricism”, Quine W. *Word and object*, M.: Logos, 2000.
32. Sokal A. *Fashionable Nonsense: Postmodern Intellectuals’ Abuse of Science*. New York: Picador, 1998.
33. *Pioneer anomaly*. URL: [https://en.wikipedia.org/wiki/Pioneer\\_anomaly](https://en.wikipedia.org/wiki/Pioneer_anomaly) (date of access: 15.05.2017)
34. King G., Zeng L. “Improving Forecasts of State Failure”, *World Politics*, 2001, 53 (4), 623–658.
35. McKeown T. “Case Studies and the Statistical Worldview: Review of King, Keohane, and Verba’s *Designing Social Inquiry: Scientific Inference in Qualitative Research*”, *International Organization*, 1999, 53 (1), 161–190.
36. Turner Ch., Thies C. “What We Mean by Scope and Methods: A Survey of Undergraduate Scope and Methods Courses”, *PS: Political Science and Politics*, 2009, 42 (2), 367–373.
37. Gill J. *Bayesian methods: a social and behavioral sciences approach*. Boca Raton, FL: Chapman and Hall, 2003.