



Научная жизнь, обзоры, рецензии

УДК 165.0

DOI:

10.15372/PS20150310

В.С. Пронских

*Объединенный институт ядерных исследований, Дубна, Россия
Национальная ускорительная лаборатория им. Э. Ферми, Батавия, США
vitali.pronskikh@jinr.ru*

НОВАЦИИ В СТАНДАРТАХ ЭКСПЕРИМЕНТИРОВАНИЯ В ФИЗИКЕ ЭЛЕМЕНТАРНЫХ ЧАСТИЦ XX ВЕКА

В работе выполнен обзор современных дискуссий об изменениях в стандартах научно-экспериментирования в физике элементарных частиц в XX веке на основе книги Аллана Франклина “Shifting Standards: Experiments in Particle Physics in the Twentieth Century” и ряда работ других авторов. Обсуждается выход на ведущие роли статистических критериев, таких, как «5 сигма», и проблемы их применения для обоснования утверждений об открытии новых явлений. Изучены уровни теоретической нагруженности эксперимента, включая триггеры, отбор и исключение данных, роль систематических погрешностей. Прослежены изменения в способах представления эксперимента в научных публикациях и коммуникации результата, стиле и авторстве публикаций большими научными коллаборациями. Анализируются сформулированные Франклином эпистемические стратегии экспериментирования и роли эксперимента в научном познании.

Ключевые слова: философия научного эксперимента, физика элементарных частиц, теоретическая нагруженность, статистические критерии, эпистемические стратегии, роли эксперимента

Pronskikh V.S.

*Объединенный институт ядерных исследований, Дубна, Россия
Национальная ускорительная лаборатория им. Э. Ферми, Батавия, США
vitali.pronskikh@jinr.ru*

NOVELTIES IN THE STANDARDS OF EXPERIMENTATION IN PARTICLE PHYSICS OF XX CENTURY

A review of contemporary discussions of shifts in the standards of scientific experimentation in particle physics in XX century is made based on the book “Shifting Standards: Experiments in Particle Physics in the Twentieth Century” by Allan Franklin and a number of other authors. It is discussed that leading roles are taken by statistical criteria, such as “5 sigma”, and the problems of their application for the substantiation of the statements on the discovery of new phenomena. The levels of theoretical loading of the experiment, including triggers, selection and exclusion of data, the role of systematic errors are studied. Changes in the ways of presentation of the experiment in scientific publications and communication of the result, style and authorship of publications by large scientific collaborations are traced. Epistemological strategies of experimentation formulated by Franklin and the roles of the experiment in scientific cognition are analyzed.

© Пронских В.С., 2015

Particle Physics in the Twentieth Century” by Franklin A. and works of other authors. A leading role of statistical criteria (5 sigma) and the issues related to its applications for substantiation of discovery claims are discussed. Various levels of theory-ladenness of experimentation such as triggers, data exclusion and selectivity, and systematic errors are analyzed. Changes in the ways of presentation of experiment in scientific publications and communication of results, style and authorship of publications by large collaborations are traced. Epistemic strategies of experimentation as proposed by Franklin as well as the roles of experiment in cognition are examined.

Keywords: philosophy of scientific experimentation, particle physics, theory-ladenness, statistical criteria, epistemic strategies, roles of experiment

Книга «Shifting Standards: Experiments in Particle Physics in the Twentieth Century» вносит заметный вклад в литературу по истории и философии науки и особенно в современную философию научного эксперимента, одним из основателей которой Аллан Франклин является. В центре внимания автора происходящий на протяжении XX века сдвиг в нормах представления экспериментальных результатов в физике элементарных частиц, а также возрастающая роль статистических критериев принятия результатов сообществом и недостатки таких критериев.

Стандартное отклонение

В начале книги автор разъясняет смысл т.н. стандартного отклонения (статистической значимости, приписываемой измерению) – критерия, лежащего в основе всех современных дискуссий об обосновании открытий в физике элементарных частиц. Согласно Франклину, стандартное отклонение – «наиболее заметный экспериментальный стандарт в данной области» [1]. В силу статистического характера экспериментальных измерений, каждому из результатов измерений может быть приписана статистическая неопределенность (обозначаемая греческой буквой σ), которая, вообще говоря, тем меньше, чем больше число повторений данного измерения. Неопределенность указывает на то, насколько удалено измеренное значение величины от ее истинного значения, а в случае экспериментов в физике частиц – насколько оно удалено от фона, то есть явлений отличной от изучаемых природы, способных мимикрировать под изучаемые. Из теории вероятностей следует, что если измеренный результат удален от фонового значения только на одно стандартное отклонение (в предположении наиболее популярного – нормального распределения), то вероятность того, что измеренный эффект не является фоном, составляет 68%. Однако если результат удален от фона на два таких отклонения (2σ), то вероятность того, что эффект

«истинный», составляет уже 95%. Таким образом, даже если число стандартных отклонений, отделяющих измеренную величину от фона (или другого условного «нуля»), очень велико (что недостижимо на практике), то вероятность гипотезы о том, что искомый эффект обнаружен, хотя и стремится к 100 %, но никогда этого значения не достигнет. Тем не менее, ученые принимают некоторое количество стандартных отклонений достаточным для того, чтобы делать утверждения об обнаружении искомого эффекта. Поэтому в центре внимания автора *Shifting Standards* то, как на протяжении XX века отношение ученых-экспериментаторов к числу σ , достаточных для заявления об открытии, претерпевало изменения и что лежало в основе этих изменений.

Становление критерия 5 σ

Книга открывается обширным и детальным 50-страничным прологом, подробно освещающим, как устанавливалось «правление сигма» на протяжении столетия, содержит 18 глав, обсуждающих отдельные физические эксперименты, а также главу «Случай исчезающих сигма», анализирующую отдельные ситуации, когда экспериментальные результаты, заявленные с высокой статистической достоверностью (многими σ), затем были опровергнуты более поздними измерениями. Франклин указывает, что экспериментаторы начали опираться на количество σ как меру достоверности результатов в начале 1960-х годов. В предшествующие годы количество σ (статистическая достоверность результатов) тоже, как правило, сообщалось в экспериментальных статьях, однако эпистемические утверждения авторов с ним непосредственно не связывались. Однако уже в ранние 1970-е появились утверждения ученых, что «эффект в менее чем три стандартных отклонения довольно несущественен» [2]; Франклин обращается к опубликованным в те годы работам, сообщающим о «наблюдении» эффекта (термин, близкий по значению к «открытию») со значимостью 4σ . Рост значимости сигма как условия открытия в экспериментальной физике частиц, названный Франклином «правлением сигма», продолжился и в последующем десятилетии [3], и в статьях, вышедших в 1980-х годах, результаты со статистической достоверностью в 4 σ считались достаточными для утверждений о «наблюдении» (observation) эффекта, хотя это условие не всегда было явно отражено. В те же годы авторы относили данные с достоверностью в 3 σ к результатам пограничной значимости.

Требование значимости результатов 5 σ хотя еще и не формулировалось как официальное условие, появилось в физике элементарных час-

тиц в 1990-х годах. К примеру, коллаборация Коллайдерного Детектора в Фермилабе (CDF) опубликовала последовательно две статьи о топ кварке, в 1994 и 1995 годах. Первая из этих статей представляла 2.8σ эффект и была названа «Свидетельство» (Evidence). Вторая статья содержала уже 4.8σ эффект и называлась «Наблюдение» (Observation). Тем не менее, даже тогда «пятисигмовый» стандарт «...определенно еще не был ни согласован, ни проведен в жизнь» (4). В качестве собственно стандарта, который редакторы ведущих физических журналов распространяли уже в качестве инструкции для авторов, он появился около 2003 года; тогда же ему последовало большинство коллабораций для публикаций статей типа «Наблюдение» (большинство статей, представляющих четыре и менее σ , публиковались с заголовком «Свидетельство»). Тот же стандарт 5σ стал вполне устоявшимся стандартом к моменту появления первых публикаций об обнаружении бозона Хиггса [5].

В Таблицах 1 и 2 приведены количества статей типа «Наблюдение» или «Свидетельство» в зависимости от числа σ [6]. Данные Франклина четко показывают, что после 2003 года отмечен рост статей как первого типа со значимостью не ниже 5σ , так и второго менее 5σ . В то же время, критерии 5σ служит очевидным разделителем между публикациями обоих типов.

Франклин поднимает вопрос о том, могут ли прежние заявления об открытиях, опирающиеся на меньше, чем 5σ , быть пересмотрены ретроспективно. Кроме того, как показывает Франклин в дальнейшем, при любом количестве сигма флюктуации теоретически возможны, и они происходят на практике.

Таблица 1

Количество статей типа "Наблюдение" в Physical Review Letters [6]

Год	2003	2004	2005	2006	2007–2010	Итого
Тип критерия						
Заявлено $\geq 5\sigma$	7	15	11	27	30	90
Заявлено $\leq 5\sigma$	1	1	0	0	0	2
Нет указаний о σ , Но явный $\geq 5\sigma$ эффект	1	1	2	4	3	11
Нет указаний о σ	1	0	1	0	0	2
Итого	10	17	14	31	33	105

Таблица 2

Количество статей типа "Свидетельство" в Physical Review Letters [6]

Год	2003	2004	2005	2006	2007–2010	Итого
Тип статьи						
Заявлено $\geq 5\sigma$	2	0	0	1	0	3
Заявлено $\leq 5\sigma$	3	5	9	4	19	40
Нет указаний о σ , Но явно $\geq 5\sigma$ эффект	0	0	0	0	0	0
Нет указаний о σ	0	0	1	0	0	1
Гибридные	1	1	2	2	1	7
Итого	6	6	12	7	20	51

Исчезающие сигма

Ситуационное исследование «исчезающих сигма» занимает в книге центральное место, ему отведена отдельная глава, а также посвящен ряд дискуссий в тексте. В 2003 году несколько экспериментальных групп объявило об обнаружении так называемого «пентакварка», адрона, состоящего из пяти кварков, со статистической достоверностью, приближавшейся к 5σ . Этим результатам предшествовали теоретические предсказания существования таких частиц (иногда отмечается, что упоминания пентакварковых адронов встречаются еще в работах Гелл-Манна и Цвейга, авторов кварковой модели строения адронов). Комбинированная статистическая достоверность обнаружения пентакварка этими экспериментами, по оценке А.Франклина [7], составила от 17 до 20σ . Чтобы получить такую оценку, Франклин с коллегами использовал вероятности, опубликованные десятью независимыми положительными пентакварковыми экспериментами (каждый из этих экспериментов заявлял о достоверности на уровне $4-5\sigma$). Учитывая опубликованные экспериментаторами погрешности, они получили для каждого из этих экспериментов вероятности того, что наблюдение пентакварка было статистической флуктуацией (см. Табл. 3). Затем эти вероятности были перемножены, чтобы получить комбинированную вероятность флуктуации в соответствии с правилами статистики, которая затем была преобразована в число σ . Несмотря на такую высокую статистическую достоверность, в 2006 году вышли три новые экспериментальные работы, опровергающие существование пентакварков. В результате публикации ряда после-

довавших работ в научном сообществе в конце 2000-х утвердилось мнение, что пентакварков (по крайней мере, данного типа) не существует, а заявления об их обнаружении было «ложной тревогой» [8]. В книге есть отсылка еще к одному эксперименту с достоверностью 20σ – обнаружению нейтрального бозона группой Маглича, которое впоследствии также не было подтверждено новыми экспериментами. Это ситуационное исследование ставит ряд вопросов об эпистемической роли σ (и вообще статистических критериев) в научном познании, и Франклин ищет ответы на эти вопросы на протяжении всей книги.

Таблица 3

Статистическая значимость ряда экспериментов и оцененные на их основе Франклином [6] вероятности обнаружения пентакварка Θ^+

Эксперимент	Значимость	Вероятность	Вероятность, мин	Вероятность, мин
LEPS	$4.6_{-1}^{+1.2}\sigma$	$4\cdot 10^{-6}$	$3\cdot 10^{-4}$	$7\cdot 10^{-9}$
DIANA	4.4σ	$1\cdot 10^{-5}$		
CLAS-d	$5.2\pm 0.6\sigma$	$2\cdot 10^{-7}$	$4\cdot 10^{-6}$	$7\cdot 10^{-8}$
SAPHIR	4.8σ	$2\cdot 10^{-6}$		
Asratyan et al.	6.7σ	$2\cdot 10^{-11}$		
CLAS-p	$7.8\pm 1.0\sigma$	$6\cdot 10^{-15}$	$1\cdot 10^{-11}$	$1\cdot 10^{-18}$
HERMES	$5\pm 1\sigma$	$6\cdot 10^{-7}$	$6\cdot 10^{-5}$	$2\cdot 10^{-9}$
SVD	5.6σ	$2\cdot 10^{-8}$		
COSY-TOF	$5\pm 1\sigma$	$6\cdot 10^{-7}$	$6\cdot 10^{-5}$	$2\cdot 10^{-9}$
ZEUS	$4.0\pm 4.6\sigma$	$6\cdot 10^{-5}$	$6\cdot 10^{-5}$	$4\cdot 10^{-6}$
Total		$8\cdot 10^{-73}$	$2\cdot 10^{-62}$	$6\cdot 10^{-86}$

Во-первых, статистическая достоверность некоторого эффекта (скажем, пика в некотором распределении (например, спектре масс рождающихся частиц)) определяется для некоторого набора данных, так называемых катов (cuts), условий отбора, которые накладываются при анализе экспериментальных данных обработчиками на основе нестатистических (и, как правило, не обсуждаемых в публикациях) критериев. Если бы каты выбирались другим образом, то и число σ было бы другим. Это так называемый эффект неполного набора (Look Elsewhere Effect), широко обсуждаемый в англоязычной методологической литературе (например, [9, 10]). Во-вторых, помимо статистической погрешности необходимо учитывать систематическую ошибку (ниже мы обсудим

понятие предвзятости экспериментаторов, влияющей на эту ошибку). Систематические ошибки могут включать в себя как неизвестные фоны [11], так и следствия недостаточного понимания режимов функционирования экспериментальных установок и приборов; все они могут быть охарактеризованы как ошибки по незнанию. Франклин подчеркивает, что знание деталей экспериментальной процедуры или особенностей анализа данных существенно для понимания того, насколько надежны и достоверны результаты и какие утверждения могут быть сделаны на их основе. Его позиция состоит в том, что обоснование утверждений об экспериментальных результатах и их эпистемологическом статусе не может быть сведено целиком к статистическим аргументам. В качестве процедур, необходимых для обоснования таких утверждений, Франклин предлагает набор эпистемических стратегий (которые должны быть отражены в экспериментальных работах, в дополнение к статистическим критериям), таких, как воспроизведение на установке новых и ранее известных эффектов, проверка внутренней согласованности результатов, использование хорошо подтвержденных теорий как для дизайна установок, так и для интерпретации результатов.

Отбор и исключение данных

Франклин посвятил одну из глав книги прояснению понятий плохих и хороших данных, а также процедур отбора и исключения данных. То, что он относит к хорошим данным, – данные не только произведенные корректно функционирующей установкой, но и очищенные от фоновых явлений, т.е. явлений, которые могут проявляться в установке или ее части (отдельном приборе) аналогично изучаемым (т.н. «истинным») явлениям, но иметь при этом иную физическую природу (подробнее об экспериментальном фоне и связанной с ней проблеме теоретической нагруженности см. в [11]). Вообще говоря, устранение фона происходит на стадии анализа данных и опирается на определенную теорию явления в явном виде (например, чтобы определить величину фона как некоей «подложки» под пиком в распределении, экспериментаторы должны сначала идентифицировать собственно пик и его положение в распределении как искомый, теоретически понятый эффект). Однако, в соответствии с подходом Франклина, отбор (хороших данных) предваряет анализ данных в вышеописанном смысле (как поиск артефактов в распределениях) и предполагает получение качественных данных за счет дизайна и настройки установки («Эксперименты редко корректно ра-

ботают сразу по включении установки» [12]). С другой стороны, Франклин дает определение плохим данным как данным, произведенным неверно работающей аппаратурой либо неверно интерпретированным. Исключение означает устранение подобных плохих данных. Приводя в качестве примера эксперименты по измерению ядерного спина и эксперимент Милликена по определению заряда электрона, он отмечает, что в ходе эксперимента экспериментаторы зачастую способны определить или предвидеть проблемы в работе аппаратуры (по крайней мере, многие из них) и принять меры к их устранению. Однако такая практика редко находит отражение в экспериментальных публикациях. Это наблюдение Франклина о том, что эксперименты требуют дополнительных и не всегда очевидных и документируемых усилий по подстройке приборов, интерпретации результатов, отбору и исключению данных, хорошо согласуется с утверждением Питера Галисона, высказанном в [13], о том, что, «читая статью, можно было бы заключить, что эффект следует из экспериментальной установки с неотвратимостью логической импликации», тогда как только сам экспериментатор знает сильные и слабые стороны его «приборов, материалов, кол-лаборантов, интерпретаций и суждений».

Обсуждая работу Милликена, Франклин замечает, что тот измерил данные для 175 капель масла, причем опубликовал только 58 из них, а для определения заряда электрона использовал только 23. Это указывает на тот факт, что уже Милликен активно применял отбор и исключение данных. По сравнению с этим экспериментом количество данных в последнее время значительно возросло (для сравнения, в 1999 году эксперимент КTeV измерил 2607274 события только распада K_1^0 мезона на 2 пиона) [14]. Такой рост объема данных привел и к росту требуемых компьютерных мощностей. Однако, поскольку мощности компьютеров и систем сбора данных все равно оказывается недостаточно по сравнению с количеством разнообразных явлений, происходящих в установке, возникла необходимость отбрасывать значительную часть данных еще на этапе их накопления за счет систем, называемых триггерами.

Пристрастность экспериментаторов

Франклин отводит особое место обсуждению пристрастности экспериментаторов (bias) в ситуациях, когда результаты измерений согласовались с предсказаниями теории высокого уровня, исходными допущениями экспериментаторов или предшествовавшими измерениями, отме-

чая важную роль ошибок нестатистической природы для понимания подобной согласованности. Для восемнадцати экспериментов, подробно описанных в книге, он приводит подробные описания использовавшейся аппаратуры и процедур анализа данных, опираясь на оригинальные публикации в ведущих научных журналах. Такой подход может встретить определенную критику, связанную с тем, что, как неоднократно отмечалось в работах П. Галисона, Э. Пикеринга и ряда других авторов, экспериментальные публикации нередко содержат в значительной мере рационально реконструированное изложение процедур измерения и анализа данных, оставляя за скобками множество существенных аспектов. Тем не менее, подход Франклина представляется оправданным по ряду причин. Во-первых, благодаря его богатому опыту в экспериментальной физике, трудно усомниться в его способности «читать между строк» анализируемые публикации и заполнять недосказанность пониманием вопроса, почерпнутым в собственной практике. Во-вторых, наблюдение конструктивистов о реконструкции экспериментаторами логики нарратива экспериментальной публикации обычно относилось к результатам Большой Науки, тогда как книга охватывает эксперименты вплоть до работ начала XIX века (таких, как статья Милликена, датированная 1911 годом), к которым подобная критика обычно не относилась. Впрочем, работа Милликена также подвергается Франклином определенной критике, связанной с тем, что тот также использовал отбор и исключение данных, основываясь на не вполне понятных критериях.

Франклин пишет: «Использование критериев отбора открывает возможность для экспериментатора подобрать каты таким образом, чтобы произвести желаемый эффект, если связь катов с конечным результатом ему известна. При этом результат может быть в согласии с существующей теорией, допущениями экспериментатора или предшествовавшими результатами» [15]. В частности, при обсуждении постановки эксперимента по определению температуры красного накаливания меди, в работе Кеннелли и Фессендена отмечается, что «...систематическая вариация между наблюдениями при смене наблюдателя составляла около двух градусов Цельсия, и, так как критерий осязательной видимости чисто психологический, его, по-видимому, невозможно точно определить» [16]. В описании экспериментов Комптона по изучению процесса рассеяния электромагнитного излучения на электронах авторы также уделяют внимание проблеме пристрастности, указывая, что не исключают, что «наблюдаемые совпадения – результат неосознанной склонности ложно оценивать угол, делая согласующиеся ошибки в результате» [17].

В экспериментах XX века для уменьшения влияния пристрастности экспериментаторов стало распространенным применение т.н. «слепого анализа», когда при обработке данных экспериментаторы до окончания анализа не видят, в какую область графика попадают обрабатываемые ими данные. С появлением этого типа анализа в научной литературе обсуждение пристрастности стало весьма редким, хотя, очевидно, «слепой анализ» не может полностью снять данную проблему. Во-первых, уже в современных дискуссиях экспериментов по поиску темной материи авторы указывают [18], что фактически на практике имеет место также «пост-слепой анализ», когда отдельные данные (три события-кандидата в эксперименте коллаборации XENON100) отбрасывались уже после завершения «слепого анализа» и открытия графиков. Во-вторых, если допустить, как изначально предполагал Франклин, что обработчики данных могут знать или предполагать, как наложение тех или иных катов может влиять на результат, то даже последующее скрытие результата в ходе «слепого анализа» не сможет предотвратить возможности (неосознанного) выбора ими «нужных» катов.

Триггеры и теоретическая нагруженность

Триггеры – это электронные системы, основанные на логических схемах, используемые в ходе эксперимента, чтобы отобрать для хранения и последующей обработки только некоторую часть из данных, производимых установкой. Благодаря этому значительная часть данных эксперимента вообще не записывается. Та же часть из них, которая записывается, отбирается логическими схемами, по образному замечанию Франклина, в предположении того, что «новая физика», на поиск которой нацелены экспериментальные приборы, повторяет старую. Это происходит потому, что практически все модели за пределами Стандартной Модели предсказывают рождение новых частиц, распадающихся на частицы с большими поперечными импульсами и струи [19]. Франклин выделяет две стадии исключения данных из анализа: меню триггера и отбрасывание плохих данных (описанное выше в разделе про отбор и исключение). Меню триггера представляет собой список условий совместного появления определенных сигналов в установке (срабатывания определенных датчиков), которые следуют из теории высокого уровня, проверяемой в эксперименте (см. Табл. 3.). Только сигналы, удовлетворяющие условиям, записанным в меню, попадают в системы сбора данных и могут попасть на дальнейшую обработку (остальная часть данных просто пропадает).

Как указывает Франклин, «на БАК это исключительно серьезная проблема. На БАК столкновения пучков происходят каждые 25 наносекунд и в каждом столкновении происходит 20 протон-протонных взаимодействий. Это означает, что события происходят с частотой 800 мегагерц. (Это, естественно, зависит от светимости пучка). В то же время системы сбора данных могут обработать только скорость набора событий около ста циклов в секунду. Таким образом, число записываемых событий меньше числа происходящих в столкновении пучков событий примерно в миллион раз» [19]. Например, триггер эксперимента CMS на БАК содержит два уровня: Level-1, который реализован аппаратно и осуществляет назкоуровневый первичный анализ данных, и триггер верхнего уровня HLT, который реализован как множество программных фильтров, проверяющих вышеописанные условия, заложенные в меню. Только данные, прошедшие оба этих уровня триггеров, поступают на дальнейший анализ [19]. «Большая часть произведенных данных никогда не записывается», заключает Франклин.

Проверка селективности и устойчивости результатов при анализе данных включает в себя процедуру варьирования выборки данных (катов), используемых для поиска новых явлений (пиков в распределениях), с тем, чтобы убедиться, что результат устойчиво воспроизводится. Однако, как отмечает К.Карака, чьи исследования обсуждает Франклин, этот подход неприменим (по крайней мере, не вполне надежен и недостаточен), когда отбор происходит уже на этапе набора данных (как в случае триггера). К тому же, для конструирования триггера непосредственно применяются теоретические модели, что создает ситуацию прямой теоретической нагруженности результатов эксперимента.

Проблемы анализа данных

Если Милликен обрабатывал свои данные вручную и наиболее продвинутым средством расчетов была таблица логарифмов [20], то уже в середине 1960-х годов компьютеры использовались весьма активно, а события, происходившие в искровых камерах под действием пучков ускорителя, фотографировались, оцифровывались вручную, и данные записывались на магнитные ленты, которые читались компьютером в ходе последующего анализа. С появлением проволочных и дрейфовых камер, кремниевых стриповых детекторов данные стало возможно считывать только в цифровом виде. Это потребовало увеличения мощности компьютеров. По сравнению с 1960-ми, когда типичный объем памяти

компьютера составлял 32 килобайта, в 2010-х такой объем был уже на уровне 4 гигабайт, а триггер эксперимента CMS включает 1000 таких компьютеров. Собственно триггер и новации, которые он принес в анализ данных, обсуждаются ниже. Таким образом, гигантский рост количества набираемых данных и их принципиально цифровой характер, делающий использование мощных компьютеров необходимым, – одна из особенностей современного этапа развития физики элементарных частиц.

Наряду с тем, что анализ данных в современном эксперименте существенно усложнился [21], методы как анализа, так и расчета погрешности не всегда описаны в работах. В частности, при сравнении данных с предсказаниями теоретической модели различие в 1σ (одна погрешность) считается хорошим согласием, а 5σ – опровержением, тогда как определение погрешности – это не просто применение формулы (одной из множества возможных), а процедура, требующая применения фоновых (дополнительных) знаний и суждений. Например, вероятности того, что сигнал (пик над фоном в некотором распределении) будет флюктуировать вниз (в сторону приближения к фону) и того, что фон будет флюктуировать вверх (в сторону пика), различны. Кроме того, чтобы определить величину фона непосредственно под пиком в распределении, требуется его интерполировать в область под пиком, для чего фон необходимо «фитировать» – приблизить некоторой функцией. От выбора фитирующей функции и собственно способа фитирования (подгонки) результат анализа также может зависеть. Таким образом, компьютеризация анализа данных привела к росту числа допущений и неопределенностей в самой процедуре анализа.

Изменения в авторстве публикаций

Франклин указывает, что смысл авторства экспериментальной публикации так же, как и число авторов таких работ, претерпели серьезные изменения, от одного в экспериментах Милликена до двух тысяч в эксперименте Компактный Мюонный Соленоид (CMS) в ЦЕРН (недавно появилось сообщение, что вышли публикации ЦЕРН со списком авторов, превышающим пять тысяч человек). Он рассказывает историю из конца 1950-х годов о молодом ученом, который внес заметный вклад на различных этапах эксперимента, но не был включен в число авторов публикации старшими коллегами, ведущими авторами работы, поскольку он «не имел достаточных знаний, чтобы выступить с докладом об

эксперименте» [22]. В последующие годы стандарты изменились, и в авторы стали включать множество сотрудников, которые внесли в работу тот или иной вклад, но тем не менее выступать с докладами общего характера от имени эксперимента (коллаборации) по-прежнему получают право лишь немногие его участники. Согласно Франклину, в современных экспериментальных коллаборациях только от пяти до двадцати человек имеют доступ к анализу (обработке) данных эксперимента, который непосредственно ведет к получению экспериментальных результатов. Соответственно, эта небольшая группа экспериментаторов и знакома с анализом данных в достаточной мере, чтобы делать общие доклады о результатах. Чередуясь, они оттачивают свое ораторское и полемическое мастерство на регулярных конференциях и совещаниях. Другие авторы, а это научные сотрудники и инженеры, которые вносят вклад в дизайн установок, дежурят на измерениях, разрабатывают компьютерные программы, составляют более «бессловесную» часть коллаборации. Она, хотя и не имеет доступа к анализу данных и не допущена представлять коллаборацию с общими докладами во внешних организациях и конференциях, как правило, может представлять свои более частные, технические результаты. Таким образом, Франклин выявляет колоссальные изменения в политике авторства между серединой двадцатого и началом двадцать первого века по причине увеличения списков авторов и их эпистемической стратификации. Эта контроверза, хотя и не очень объемная часть книги, тем не менее, представляется знаковой.

Также, указывает Франклин, по сравнению с началом прошлого столетия, кардинально изменился стиль научных публикаций с личного на безличный [23]. В частности, авторы ранних публикаций нередко содержали критические замечания в адрес предшественников личного характера и, в том числе, основывали на подобной критике критику их научных результатов [24]. Например, обсуждая эксперименты Гука, где тот определил смещение падающих тел в южном направлении, Эдвин Холл замечал в 1903 году в отношении самого Гука, что «в человеке с такой репутацией пристрастность трудно не заметить» [24]. Статьи индивидуальных авторов, как замечает Франклин, обычно бывают менее выдержанными, чем коллективные. Теперь личные комментарии по поводу научных результатов и их авторов также появляются, но уже не в публикациях, а в интервью, твиттере и т.д. Также в современных публикациях, по сравнению с работами прошлого, значительно выросла плотность цитирования.

Один из вопросов, возникающих в связи с коллективным авторством публикаций большими научными коллаборациями, – достаточно ли указывать в выходных данных статьи только название коллаборации или необходимо перечисление всех участвовавших авторов. Франклин считает, что недостаточно, поскольку коллаборация не является чем-то неизменным во времени, ее периодически пополняют новые или покидают прежние участники, поэтому неуказание конкретных авторов не позволяет впоследствии установить, кто был ответственным за те или иные результаты. Недостатком такого подхода является то, что в современных экспериментальных публикациях список авторов зачастую превосходит по размерам содержательную часть. Например, в приводимой Франклином в качестве примера публикации коллаборации CMS в ЦЕРН на 20 страниц текста приходится 17 страниц со списком авторов, и только 3 приходятся на изложение сути работы.

Изменения в способе представления эксперимента

Одно из важных изменений в том, как эксперимент представляется в научных публикациях и докладах, – изменения в способе изображения экспериментальной установки. Публикация всегда представляет некоторый идеализированный эксперимент, и в современных публикациях Франклин отмечает тенденцию от более реалистичного к максимально идеализированному его представлению. Различение между реальным и идеализированным экспериментом встречается еще у П. Дюгема, и Франклин определяет их как различие между экспериментом, который выполняет человек или группа людей, и экспериментом, из которого удалены все субъективные элементы [25]. В качестве примера Франклин приводит части экспериментальной установки – приборы, которые теперь изображаются в виде схем, а не фотографий. При этом он замечает, что фотографии также широко доступны в других источниках, помимо научных публикаций. В статьях не приводятся и подробные описания приборов, только их отдельных частей, наиболее критичных, по мнению авторов. Таким образом, ранее публикации содержали достаточно деталей для воспроизведения эксперимента, тогда как современные таких деталей не содержат. Для сравнения, Франклин приводит в качестве примера публикацию, содержащую 14 страниц только обсуждений процедур отбора данных, реконструкции и результатов, тогда как часть, посвященная описанию приборов, значительно скромнее. Помимо этого, описание *modus operandi*, собственно экспери-

ментальных процедур, из современных статей практически полностью исчезло.

Обращаясь к истории изображений установок в публикациях, Франклин указывает, что еще Милликен различал фотографии установок и их схематические изображения, причем именно последние «должны делать постижимыми необходимые операции» [25]. Сравнивая изображения установки, при помощи которой Милликен экспериментировал с масляными каплями, Франклин отмечает, что фотография мало что сообщает об установке, тогда как ее схематическое изображение, сопровождаемое пояснительным текстом, является гораздо более информативным. Однако если такие рисованные изображения установок в первой половине XX века были вполне реалистичны и достаточно детальны для того, чтобы дать представление о принципах действия установок, то во второй половине XX века становится заметной тенденция увеличения идеализации из представления. Из современных публикаций изображения установок практически полностью исчезли, но при этом они остались широко доступны в технических документах, докладах и других источниках. При этом Франклин замечает, что текст, описывающий такие изображения, практически исчез.

Если, например, Холл, описывая в 1903 году достаточно простой по постановке эксперимент по падению тел, отвел описанию техники эксперимента семь из десяти страниц, то в публикации 1967 года результатов эксперимента по изучению K_{e2}^+ распада (несоизмеримо более сложный в постановке) техническим деталям было отведено около трех строк. Описания как самих установок, так и техники выполнения эксперимента в работах начала XX века были настолько детальны, что позволяли полностью повторить эксперимент, тогда как более поздние и современные эксперименты такой возможности не предоставляют. Напротив, описания обработки результатов стали занимать значительную часть современных публикаций; например, в недавних публикациях эксперимента CMS [26] описанию установки отводится от одного абзаца до одной страницы, притом что описание отбора событий для анализа и алгоритмов их реконструкции занимает в отдельных случаях 11 и 14 страниц. Несмотря на это, изложение процедур анализа также довольно обобщенное. Франклин называет такой способ представления экспериментов в современных публикациях эссенциализмом.

Роли эксперимента и эпистемические стратегии

В своей книге Франклин суммирует высказанные им в более ранних работах так называемые эпистемические стратегии эксперимента, то есть подходы и методы, которые экспериментаторы используют в своих работах для того, чтобы обосновать корректность результатов. К таким стратегиям он относит:

1. Вмешательство (описанное также Я. Хакингом), то есть тот факт, что в эксперименте имеет место искусственное приготовление состояний измеряемых систем.
2. Независимое подтверждение (т.е. независимое воспроизведение результата другими экспериментами).
3. Экспериментальные проверки и калибровки, в которых экспериментальная установка воспроизводит ранее известные феномены.
4. Воспроизведение явлений и артефактов, о присутствии которых известно заранее.
5. Устранение возможных источников ошибок с поиском альтернативных объяснений (стратегия Шерлока Холмса).
6. Использование самих результатов для доказательства их корректности (т.е. внутренняя согласованность результатов).
7. Использование независимо хорошо подтвержденной теории явления для объяснения результатов.
8. Использование установок, чей принцип работы основан на хорошо подтвержденных теориях.
9. Использование статистических аргументов.
10. Использование «слепого анализа», когда критерии отбора данных выбираются независимо от конечного результата.

Все эти стратегии и аргументы, как заключает Франклин, в явном виде присутствовали в публикациях начала XX века, но впоследствии, к концу столетия, в связи с усложнением установок и техник их использования, практически перестали включаться в статьи, хотя, явно или неявно, следы использования многих из этих стратегий присутствуют во второстепенных источниках.

В заключение Франклин, продолжая дискуссию, начатую в начале книги, перечисляет роли, которые, по его мнению, играет эксперимент в познании [27]. В дополнение к распространенному в философии науки представлению о том, что эксперимент служит для проверки теорий, он

може стимулювати створення нових теорій, або опровергає існуючі теорії, або виявляючи новий феномен, що потребує пояснення. Також експеримент може служити джерелом ідей про математичну структуру теорій і свідченням на користь онтологій цих теорій. Також, вважає Франклін, можливі пошукові експерименти, в яких предмет інтересу досліджується з метою спроби сформулювати теорію. Крім того, експеримент може вимірювати величини, які або важливі теоретично, або представляють практичну значимість. Крім цього, експеримент може мати «власне життя», коли експериментатори вимірюють певне явище, яке викликало інтерес.

Експеримент може служити для вираження теорії, наприклад, в випадках, коли його точність недостатня для її перевірки і результати просто узгодяться з її прогнозами в раніше перевіреній області. До цього класу Франклін відносить деякі експерименти на перевірку Стандартної Моделі, а також описані в книзі (і обговорені вище) експерименти на виявлення пентакварка, який хоча і не був передбачений теорією, але не суперечив їй, в зв'язі з чим також віднесений Франкліном до експериментів, що виражають теорію. Кожен експеримент може грати більше однієї ролі. Автор приходить до висновку, що, незважаючи на драматичні зміни, що відбулися в експериментальній фізиці з початку ХХ століття і докладно обговорені в книзі, ролі, що виконуються експериментом у пізнанні, не зазнали суттєвих змін, і їх можна вважати робастними характеристиками експерименту.

Аллан Франклін і конструктивісти

Окремий інтерес представляє продовження Франкліном в цю книгу дискусії з конструктивістами. У своїй відомій книзі [28], Е.Пікерінг, аналізуючи експерименти в фізиці частини, прийшов до висновку, що фізики приймають існування певних явищ, оскільки вони узгодяться з їх груповими переконаннями і прийнятими і підтримуваними в спільноті теоріями, і схильні відкидати дані, які суперечать їх поглядам. Інтересно, що ряд експериментів, розглянутих в книзі Франкліна, може бути інтерпретований як підтримуючий тезис Пікерінга. Наприклад, експерименти про подвійне розсіяння електронів підтвердили диракову теорію електрона, а відсутність експериментальних свідчень на користь

пятой силы подтвердили теорию Ньютона [29]. Однако Франклин не согласен с таким объяснением. Он указывает на эксперименты, которые не привели к подтверждению какой либо теории, либо такие, которые были поставлены в отсутствие какой-либо общепринятой теории. Представляется, что неподтверждение ни одной теории экспериментом можно также объяснить в рамках концепции Пикеринга тем, что ни одна из них не имела достаточной поддержки в сообществе. Контрпримером Пикерингу было бы скорее указание на экспериментальное опровержение теории, поддерживаемой сообществом.

Уже без особенной полемики обращается Франклин к вопросам, поднятым в работах Гарри Коллинза – социолога, долгие годы исследовавшего работу физиков в экспериментах с гравитационными волнами. Один из таких вопросов – что именно можно считать экспериментом с точки зрения анализа данных. Этот вопрос был поднят Коллинзом в связи с дискуссией, развернувшейся по поводу перехода от критерия «3 сигма» к «5 сигма». В примере Коллинза, ученые применяли статистический критерий к данным (графикам), накопленным за год измерений; иное количество графиков привело бы к другим результатам. Год измерений, замечает Коллинз [30], – весьма произвольный критерий. К примеру, было бы рассмотрение всех вместе графиков тем же самым экспериментом или другим? С другой стороны, если к данным применяются критерии отбора (каты), причем каждый раз разные, будут ли это разные эксперименты или один и тот же? Франклин неявно соглашается с Коллинзом в том, что это законные вопросы и ответ на них опирается на определенные произвольные допущения (см. также обсуждение эффекта ограниченной выборки (the Look Elsewhere Effect) выше). Как вспоминает Франклин [31], многолетние дискуссии с Коллинзом послужили для него одним из стимулов к написанию данной книги.

Заключение

В книге А.Франклина «Сдвиг стандартов» сделан ряд важных наблюдений об изменениях в методологии и эпистемологии физического эксперимента, произошедших во второй половине – конце XX века. В частности, прослежено вхождение в научную аргументацию критерия «сигма», практически не использовавшегося до 1960-х годов, а затем ставшего неотъемлемой частью обоснования научных открытий в физике элементарных частиц. Франклином изучены изменения в применении этого критерия и в отношении к количеству «сигма», требуемых для ут-

верждений о «наблюдении» феноменов от простого указания статистической значимости результата до требования опубликовать результаты со значимостью не менее 5 сигма. Он ставит вопрос о том, как в свете новых требований должны рассматриваться открытия прошлого, когда такого критерия не существовало. Подробно обсуждается случай, когда десять независимых экспериментов обнаружили новую частицу Θ^+ (гипотетический пентакварк), причем каждый из них со значимостью, превышающей 5 сигма (что при суммировании, выполненном собственноручно Франклином, составляет 17–20 сигма), а затем несколько последующих экспериментов этой частицы не обнаружили.

В связи с этим Франклин отмечает, что статистические критерии хотя и несомненно важны, тем не менее не являются исчерпывающими. Поясняя это утверждение, Франклин указывает на множество возможных источников систематической ошибки в эксперименте. Среди них отбор и исключение данных, опирающиеся на неявные представления экспериментаторов о том, как должна функционировать установка; пристрастность экспериментаторов, связанная с их исходными предположениями и допущениями; триггеры, отбрасывающие большую часть данных еще на этапе накопления и основанные на теориях явления; неопределенности анализа данных, такие, как отбор только части данных для анализа (каты) и произвольность выбора подгруппочных (фитирующих) функций. Для уменьшения роли систематических ошибок Франклин предлагает набор эпистемических стратегий, таких, как воспроизводимость результата, калибровка приборов или использование проверенных теорий в эксперименте.

Также Франклин делает ряд важных наблюдений о социологии эксперимента и презентации результатов. Авторские коллективы в середине XX века включали только единицы тех экспериментаторов, которые были подготовлены в качестве докладчиков «для внешнего мира». Теперь же они включают тысячи человек, внесших вклад в отдельные этапы эксперимента, но не имеющих возможности или компетенций представлять эксперимент широкой публике. Одна из контроверз современного авторства заключается в том, что при перечислении всех авторов их список составляет большую часть статьи, тогда как простое указание названия коллаборации (коллектива эксперимента) вместо этого не решает проблему, поскольку состав участников меняется во времени и становится неясным, кто был ответственным за тот или иной результат. В отношении презентации эксперимента Франклин обнаруживает, что если эксперименты начала XX века содержали подробное описание *modus*

operandi и сопровождалась фотографиями установок, то современные экспериментальные публикации включают только схемы отдельных узлов и выборочное описание отдельных процедур. Таким образом, по его наблюдению, современные публикации больших экспериментов (например, на БАК) не содержат достаточно информации для независимого воспроизведения эксперимента. То же замечание относится к комментариям, касающимся персон экспериментаторов, которые ранее включались в текст публикаций и могли пролить свет на непроясненные иным способом особенности экспериментов.

Несмотря на то что ведущая роль теории и допущений экспериментаторов различной природы Франклином подробно рассмотрена (триггер, анализ данных, отбор и исключение данных), несколько противоречивым выглядит, например, его заключение о том, что роли эксперимента (включая стимулирование создания новых теорий и исследование новых феноменов) остались неизменными с начала прошлого века. Взгляды Франклина, таким образом, можно охарактеризовать как позитивистские и эмпиристские, что иногда приходит в противоречие с его собственными наблюдениями и выводами. При этом книга, представляя собой труд вполне энциклопедического характера, несомненно, может послужить важным источником сведений о физическом эксперименте.

Литература

1. Franklin A., *Shifting Standards: Experiments in Particle Physics in the Twentieth Century*. – University of Pittsburgh Press, 2013, – P. ix.
2. Franklin A., *Shifting Standards: Experiments in Particle Physics in the Twentieth Century*. – P. xi.
3. Franklin A., *Shifting Standards: Experiments in Particle Physics in the Twentieth Century*. – P. xxii.
4. Franklin A., *Shifting Standards: Experiments in Particle Physics in the Twentieth Century*. – P. xxxii.
5. Chatrchyan S., Khachatryan V., Sirunyan A.M., Tumasyan A., Adam W. et al., Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC // *Phys. Lett. B.* – 2012. – V. 716. – P. 30.
6. Franklin A., *The Rise of Sigmas*// In: *Biennial meeting of the Philosophy of Science Association (PSA)*.- Chicago.- 2014.
7. Franklin A., *Shifting Standards: Experiments in Particle Physics in the Twentieth Century*. – P. 220.
8. Franklin A., *Shifting Standards: Experiments in Particle Physics in the Twentieth Century*. – P. 219.
9. Staley K., *The Evidence for the Top Quark. Objectivity and Bias in Collaborative Experimentation*. – The University of Cambridge Press. – 2004. – PP.360.

10. Dawid R., Higgs Discovery and the Look Elsewhere Effect // *Philosophy of Science*.- 2015.- V.82(1).- P.76.
11. Пронских В.С., Эпистемическая роль экспериментального фона в философии научного эксперимента // *Философия Науки*. – №2.- V.65.- P. 41.
12. Franklin A., *Shifting Standards: Experiments in Particle Physics in the Twentieth Century*. – P. 127.
13. *Galison P.L.* How experiments end. – The University of Chicago Press, 1987. – P. 244.
14. Franklin A., *Shifting Standards: Experiments in Particle Physics in the Twentieth Century*. – P. 222.
15. Franklin A., *Shifting Standards: Experiments in Particle Physics in the Twentieth Century*. – P. 3.
16. Franklin A., *Shifting Standards: Experiments in Particle Physics in the Twentieth Century*. – P. 14.
17. Franklin A., *Shifting Standards: Experiments in Particle Physics in the Twentieth Century*. – P. 44.
18. Franklin A., *Shifting Standards: Experiments in Particle Physics in the Twentieth Century*. – P. 248.
19. Franklin A., *Shifting Standards: Experiments in Particle Physics in the Twentieth Century*. – P. 224.
20. Franklin A., *Shifting Standards: Experiments in Particle Physics in the Twentieth Century*. – P. 222.
21. Franklin A., *Shifting Standards: Experiments in Particle Physics in the Twentieth Century*. – P. 227.
22. Franklin A., *Shifting Standards: Experiments in Particle Physics in the Twentieth Century*. – P. 4.
23. Franklin A., *Shifting Standards: Experiments in Particle Physics in the Twentieth Century*. – P. 232.
24. Franklin A., *Shifting Standards: Experiments in Particle Physics in the Twentieth Century*. – P. 8.
25. Franklin A., *Shifting Standards: Experiments in Particle Physics in the Twentieth Century*. – P. 238.
26. Franklin A., *Shifting Standards: Experiments in Particle Physics in the Twentieth Century*. – P. 240.
27. Franklin A., *Shifting Standards: Experiments in Particle Physics in the Twentieth Century*. – P. 249.
28. Pickering A., *Against Putting the Phenomena First: The Discovery of the Weak Neutral Current* // *Studies of History and Philosophy of Science*.- V15. – №2.- P. 85.
29. Franklin A., *Shifting Standards: Experiments in Particle Physics in the Twentieth Century*. – P. 148.
30. Franklin A., *Shifting Standards: Experiments in Particle Physics in the Twentieth Century*. – P. liv.
31. Franklin A., доклад на Fermilab Colloquium, июнь 2015.

Дата поступления 17.09.2015