Повторение как стадия исследовательского процесса

А. А. Фёдоров

doi: 10.38098/thry_23_0457_06

Раньше, чем ты выведешь из этого случая общий закон, произведи испытание два или три раза и посмотри, дают ли испытания одинаковые результаты.

Леонардо да Винчи

«Да» нуждается в повторении.

Жак Деррида

Повторение – мать учения.

Пословица

Библиометрическая прелюдия

Мнимый или реальный кризис психологии, споры о котором длятся уже больше полутора столетия, в последние годы часто обсуждается в контексте проблемы воспроизводимости результатов психологических исследований. Безусловно, «кризис воспроизводимости» касается не только психологии, но именно на страницах психологических изданий, если опираться на данные базы Scopus, ведутся наиболее оживленные дискуссии. Из 707 публикаций, посвященных кризису воспроизводимости и вышедших с 2012 по 2021 г., более 37% относятся к отрасли «Психология» (≈43% с учетом мультидисциплинарных изданий)[1].

1 Данные извлекались 15 апреля 2022 г. из библиографической базы данных Scopus. Поисковый запрос был записан следующим образом: TITLE-ABS-KEY (“reproducibility crisis” OR “replication crisis” OR “replicability crisis” OR “crisis of replicability” OR “crisis of replication” OR “crisis of reproducibility”) AND PUBYEAR < 2022. Учитывались статьи, опубликованные по 2021 г. включительно, нижняя граница временного диапазона не ограничивалась; поиск производился по всем типам документов в названии статьи, аннотации и/или в ключевых словах. Основной анализ был ограничен публикациями в отраслях знания «Психология» и «Междисциплинарные исследования» (N=305).

Согласно сформулированному запросу, первая публикация, в названии которой использовалось словосочетание «кризис воспроизводимости» (в варианте *«replicability crisis»*), вышла в 2012 г. (Pashler, Harris, 2012). Безусловно, воспроизводимость исследований обсуждалась в психологии с момента ее зарождения как науки, но для целей данной прелюдии, которая не претендует на полный библиометрический анализ проблемы, вполне достаточно ограничиться публикациями, прямо отсылающими к обсуждению сложившейся ситуации как кризиса. Начиная с 2012 г., число статей стабильно растет, что отражает усиливающийся интерес исследователей к данной проблеме (рисунок 1). Этот рост находит статистическое подтверждение посредством простой линейной регрессии ($F(1,8)=86{,}55$, $p<0{,}001$, $R^2_{adj}=0{,}905$).

При помощи свободно распространяемой программы VOSviewer (версия 1.6.17) была построена понятийная карта. Подсчитывалась частота совместной встречаемости авторских ключевых слов с порогом вхождения, равным 5. Нормализация проводилась при помощи линейно-логарифмического метода. Понятия, прямо указывающие на кризис воспроизводимости (*replicability crisis, replication crisis, reproducibility crisis*), исключались, так как предметная область исследования была очерчена самой логикой построения базы. Термины, представляющие собой разные написания одного понятия (например, *p value / p-value, nhst / null hypothesis significance testing / null hypothesis signi-*

Рис. 1. Динамика количества публикаций, посвященных кризису воспроизводимости (2012–2021)

ficance test) объединялись. Также были объединены различные понятия, указывающие на байесовскую статистику (например, *bayes factor* / *Bayes' theorem* / *bayesian* / *bayesian analysis* / *bayesian methods* и т. п.), понятия, связанные с проблемой доверия в науке (например, *trust in science* / *public trust in science* / *trust in scientists* и т. п.), и некоторые другие, имеющие общий смысл. Учитывая порог вхождения, в ходе проведенного отбора из 786 понятий осталось 22.

Визуализация полученной понятийной сети приведена на рисунке 2. На этой карте площадь круга отражает частоту встречае-

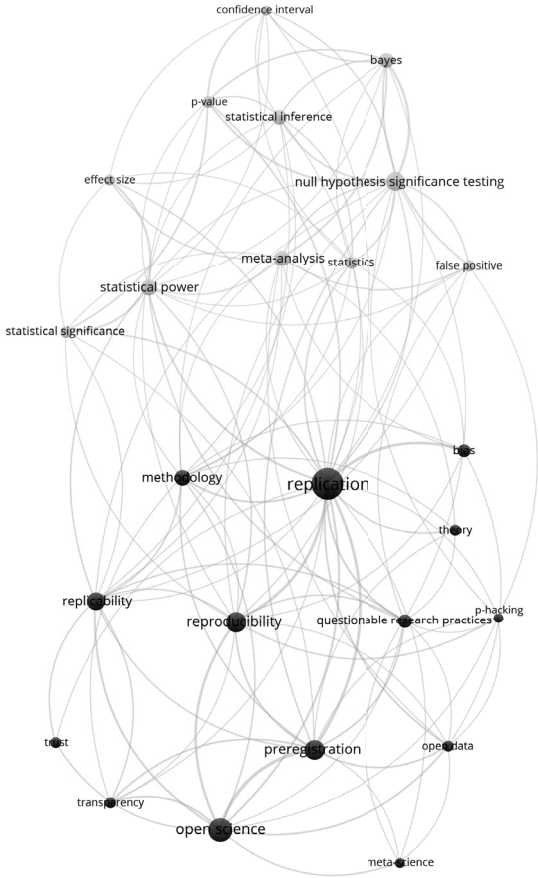

Рис. 2. Понятийная карта авторских ключевых слов психологических и междисциплинарных публикаций, посвященных кризису воспроизводимости (2012–2021)

мости термина, а расстояние между кругами является мерой связи терминов: чем оно меньше, тем сильнее связь, которая понимается как частота совместной встречаемости терминов[1].

Понятийная карта показывает, что термины образуют сеть, в которой можно выделить два тематических кластера. Первый кластер (светло-серый цвет) связан с вопросами статистического вывода. Он включает такие понятия, как *тестирование значимости нулевой гипотезы, байесовская статистика, статистическая мощность, размер эффекта* и пр. Центральная тема второго кластера (темно-серый цвет) – *открытая наука.* В него входят такие понятия, как *прозрачность, сомнительные исследовательские практики, p-хакинг, пререгистрация, доверие* и пр.

Таким образом, предварительный библиометрический анализ позволяет сказать, что в современной психологии выделяются две основных стратегии решения кризиса воспроизводимости: реформа статистического вывода и построение открытой науки. Обе стратегии предполагают изменения в сложившейся в XX в. практике проведения психологических исследований, и обе стратегии, как я попытаюсь показать далее, недостаточны. Но прежде чем обратиться к их обсуждению, следует указать, что существуют еще как минимум два варианта реагирования на проблемы с воспроизводимостью: отказ от «нарратива кризиса» и отрицание воспроизводимости как краеугольного камня науки.

Отказ от нарратива кризиса

Д. Фанелли полагает, что «нарратив кризиса», сложившийся вокруг попыток улучшить воспроизводимость результатов психологических исследований, является ошибочным, а современную ситуацию следует характеризовать в терминах позитивного «нарратива трансформации» (Fanelli, 2018). С его точки зрения, проблема, безусловно, реальна, однако нет фактов, которые однозначно бы доказывали, что ситуация ухудшается и что существующие требования, вынуждающие ученых публиковаться как можно больше, связаны с нарушениями научной этики. Более того, результаты исследований, посвященных оценке уровня воспроизводимости в психологии, как минимум неоднозначны. Оценка общего уровня воспро-

1 VOSviewer строит библиометрические сети, основанные на расстоянии (distance-based visualization), а не графах (graph-based). При такой визуализации связанность элементов передается через приблизительное – в силу частичной потери информации – расстояние.

изводимости психологических исследований в 36%, полученная Open Science Collaboration, является, вероятно, излишне пессимистичной (Open Science Collaboration, 2015). Это обусловлено разными причинами, включая существование связи между уровнем квалификации членов репликационной группы и полученным в репликации размером эффекта (Bench et al., 2017). Репликационные группы с бóльшим уровнем квалификации, объективной мерой которой служило число публикаций, получали бóльшие размеры эффекта, в том числе потому, что сами выбирали для репликации исследования с бóльшим размером эффекта, чем группы с меньшим уровнем квалификации. Учитывая, что в Open Science Collaboration 11% репликационных групп вообще не имели изданных статей, а само распределение публикаций характеризуется выраженной левосторонней асимметрией, это может объяснять достаточно низкий итоговый уровень воспроизводимости. Кроме того, уровень воспроизводимости достаточно сильно варьировал для разных областей психологии: от 23% (Journal of Personality and Social Psychology) или 29% (Psychological Science) для социальной психологии до 48% (Journal of Experimental Psychology: Learning, Memory and Cognition) или 53% (Psychological Science) для когнитивной психологии.

В целом можно сказать, что процент успешных репликаций действительно обычно выше 36% (таблица 1).

Значительное влияние на итоговый результат оказывает отбор исследований, которые планируется воспроизвести. Так, уже было сказано, что выбирая исследования с большим размером эффекта, можно увеличить число успешных репликаций. Если при этом ограничиться небольшим числом исследований, то, как видно из приведенных данных, процент успеха может достигнуть 100%.

Безусловно, наука не требует того, чтобы все исследования были полностью воспроизводимы. Ошибки неизбежны, и достичь стопроцентной воспроизводимости можно только ценой практически полного отказа от научного прогресса (Nosek et al., 2022). Тем не менее, открытым остается вопрос, какой процент воспроизводимости мы вправе ожидать от научных исследований. Вероятно, 36%, если исходить из данных Open Science Collaboration, мало, но достаточно ли, скажем, 65 или 75%? Каковы критерии воспроизводимости? И как добиться объективной оценки воспроизводимости, если на нее влияет множество факторов? Как бы то ни было, сторонники отказа от нарратива кризиса утверждают, что воспроизводимость в психологической науке выше пессимистичных оценок и предлагают описывать ситуацию в терминах новых возможностей и даже «революции».

Таблица 1

Процент успешных репликаций в психологии и смежных науках

Авторы, год	Научная область	Процент успешных репликаций на основе p-values (n из N)
Klein et al., 2014	Социальная психология и психология принятия решений	77% (10 из 13)
Open Science Collaboration, 2015	Различные области психологии, с выделением социальной и когнитивной психологии	36% (35 из 97)
Camerer et al., 2016	Экспериментальная экономика	61% (11 из 18)
Ebersole et al., 2016	Социальная психология и психология личности	30% (3 из 10)
Camerer et al., 2018	Экспериментальная социальная психология	62% (13 из 21)
Klein et al., 2018	Социальная и когнитивная психология	50% (14 из 28)
Zwaan et al., 2018b	Известные и устойчивые эффекты из области когнитивной психологии	100% (9 из 9)
Soto, 2019	Связь черт Большой пятерки с жизненными результатами (life outcome)	87% (66 связей из 76)
Cova et al., 2021 (online – в 2018)	Экспериментальная философия, включая психологию морали	78% (29 из 37)

Отказ от воспроизводимости

Другая стратегия представлена в работах сторонников постнеклассической психологии и связана с субъектом познания, задающим критерии научности. В соответствии с ней разрешение кризиса заключается в реинтерпретации, предполагающей выведение воспроизводимости из списка важнейших критериев научного знания. Как пишут сторонники этой стратегии, «проблема воспроизводимости психологических экспериментов в контексте постнеклассической методологии может быть рассмотрена с опорой на новые критерии научности в психологии... Воспроизводимость экспериментов если и войдет в перечень этих критериев, то не станет, вопреки мнению Носека, важнейшим признаком достоверности научных результатов» (Вачков, Вачкова, 2016, с. 100). В. Ф. Петренко говорит о необходимости разработки постнеклассической методологии анализа единичных случаев (часто нетривиальных, к чему мы еще вернем-

ся), которые ученые предпочитают не замечать, «боясь потерять научное реноме» (Петренко, 2018, с. 90). В качестве новых норм научности предлагается использовать диалогичность, плюралистичность, множественность истин и т. п. В рамках этой стратегии не только признается, что психологические исследования плохо воспроизводимы, но утверждается, что это не лишает их научной ценности. Так, например, Д. А. Стебаков в качестве «постнеклассического образца познания» приводит так называемые партисипаторные исследования (PAR), о которых он сам пишет, что «знания, получаемые в таких исследованиях, еще более уникальны, результаты еще менее воспроизводимы, что, кажется, ставит под сомнение их научность» (Стебаков, 2017, с. 56). При этом еще в 1992 г. Дж. Фрайдерс указывал, что в случае PAR речь идет не о научных исследованиях, а скорее о развитии сообщества (Frideres, 1992). Термин «исследование» в данном случае используется некорректно, а сам подход смешивает идеологию и политику с научным исследованием.

Проблема «постнеклассической стратегии» разрешения кризиса связана с тем, что воспроизводимость является одной из ядерных характеристик научного знания. Если мы исключим воспроизводимость из перечня критериев научности, оставшееся *нечто* перестанет быть наукой. Это не означает, что возникшее *нечто* будет хуже науки, но оно не будет наукой и не может претендовать на право использовать авторитет, который наука заработала за прошедшие столетия. Таким образом, в случае постнеклассической стратегии разрешения кризиса воспроизводимости речь идет об интеллектуальной уловке: постнеклассическая психология не желает соответствовать критериям науки и утверждает, что они к ней не применимы, но при этом и не желает отказываться от того, чтобы именоваться наукой и использовать ее авторитет. Безусловно, наука – эволюционирующий организм, но это не означает, что вопреки всем изменениям она остается наукой. Продолжая аналогию с эволюцией, вполне можно ожидать, что наука как вид после ряда изменений станет *другим* видом, при этом важно понимать, что возникший вид далеко не обязательно будет «лучше» и «прогрессивней». Можно предложить и иную аналогию, которая также исходит из того, что воспроизводимость является сущностной характеристикой научного знания. Представим науки разного типа в качестве параллелограммов (прямоугольник, квадрат, ромб) – для них верно то, что их противоположные стороны попарно параллельны. Исключим этот критерий, и нечто возникшее останется геометрической фигурой, но перестанет быть параллелограммом. Равно и с наукой: исключим из науч-

ного знания критерий воспроизводимости, и нечто возникшее останется формой знания, но перестанет быть наукой. Как указывал К. Поппер, «невоспроизводимые отдельные события не имеют значения для науки» (Поппер, 2005, с. 78–79).

Важно отметить при этом, что сама по себе уникальность объекта не является особым препятствием для научного исследования, опирающегося на критерий воспроизводимости (Фёдоров, 2020). Если существует уникальный человек, обладающий даром телепатии, ничто не мешает тому, чтобы эта экстрасенсорная способность была продемонстрирована множество раз в разных условиях в экспериментах, проводимыми разными людьми. Что является проблемой для научного исследования, так это уникальные невоспроизводимые события, не представляющие собой манифестации более общих случаев и закономерностей и не имеющие воспроизводимых эмпирических следствий. Ссылка на уникальность объекта изучения несостоятельна и для гуманитарных дисциплин, в которых репликационные исследования не только возможны, но и желательны (Peels, Bouter, 2018).

Кроме того, критерий воспроизводимости выполняет важную функцию защиты науки от мошенничества и ошибок. Если мы его исключим, то каким образом будем отличать уникальное «научное исследование» (которое больше *никто* и *никогда* не может повторить) от разных форм мошенничества, которое, по некоторым оценкам, носит в науке, включая психологию, массовый характер (Kohn, 1986; Martin, 1992)?

Возможно ли воспроизведение?

Одно из направлений критики репликационных исследований, не обязательно связанное с призывами к отказу от принципа воспроизводимости, заключается в указании на невозможность точного воспроизведения психологических исследований. Так, в своей критике проекта Open Science Collaboration Д. Т. Гилберт с соавт. в качестве одного из главных аргументов приводят тот факт, что многие репликационные исследования отличались от оригинальных по ряду параметров: выборке, используемым методикам и пр. (Gilbert et al., 2016). Сторонники постнеклассической стратегии также это отмечают, и тут уместно привести довольно объемную цитату, раскрывающую логику такой аргументации: «Воссоздать идентичные условия психологического эксперимента, видимо, просто невозможно: слишком большое число факторов оказывает влияние на его про-

ведение. Даже в физических экспериментах это достаточно трудно. А на подвижную, особенно детскую, психику оказывают влияние погода, настроение, успех (или провал) на вчерашнем уроке, утреннее опоздание в школу, цвет стен в комнате, где проводится исследование, внешний вид, пол, возраст экспериментатора, воспоминание об уехавшем в командировку отце... Все этого учесть и воспроизвести не представляется возможным» (Вачков, Вачкова, 2016, с. 100). Итак, полное воспроизведение всех условий невозможно (что верно для любого исследования), следовательно, нужно либо исключать воспроизводимость из списка критериев научного знания (постнеклассическая стратегия отказа от воспроизводимости), либо на основании этого говорить, что никакого кризиса воспроизводимости нет (стратегия отказа от нарратива кризиса). В действительности же эта аргументация совершенно неубедительна, поскольку вряд ли можно оправдать принципиально *невоспроизводимый* физический, химический или биологический эксперимент на том основании, что попытки его репликации проведены в другом городе, в другое время суток, учеными другого пола и с другим цветом кожи и пр. Собственно, на это указывают и авторы OSC, отвечая своим критикам: «Нет такой вещи, как точная репликация. Все репликационные исследования во многих отношениях отличаются от оригинальных. Они проводятся на другом оборудовании, в других погодных условиях, другими экспериментаторами, с помощью других компьютеров и мониторов, на других языках, в другие момента времени и т. д.» (Anderson et al., 2016, p. 1037c).

Аргументация, согласно которой низкий процент воспроизводимости связан с отхождением от абсолютно точной репликации оригинальных работ, возлагает вину за сложившуюся ситуацию на сами репликационные исследования. Так, Д. Канеман видит причину в том, что авторы репликаций не консультируются с авторами оригинальных работ, и прямо пишет, что «такое поведение следует запретить, и не только потому, что это не по-товарищески, но потому, что это плохая наука» (Kahneman, 2014, p. 310). С его точки зрения, убеждение, что методическая секция статьи включает в себя все необходимое для точной репликации, – это миф, особенно в психологии. Как он отмечает, «экспериментальные инструкции в методической секции обычно просто пересказываются, хотя известно, что имеет значение не только точный порядок слов, но даже шрифт, которым они напечатаны» (ibid., p. 310). Предлагаемое им решение, так называемый «новый этикет репликаций», заключается в том, чтобы *обязать* автора репликации связаться с автором оригиналь-

ного исследования до сбора данных и предоставить ему детальное описание планируемой процедуры. Затем автор оригинальной работы в течение ограниченного времени отвечает, предлагая свои комментарии и возможные изменения. Далее, хотя автор репликации и не обязан принимать эти предложения, он обязан предоставить полное описание окончательного плана с подробным объяснением, почему он отказался от тех или иных предложений автора исходной работы. Наконец вся переписка предоставляется рецензентам (это, по задумке Канемана, обеспечит конструктивное и благоразумное общение, даже если авторы оригинального и репликационного исследований раздражают друг друга).

Вероятно, нет ничего плохого в том, чтобы связаться с автором оригинального исследования (хотя в случае изучения психологических феноменов это потенциально может угрожать конструктной валидности), но предложение *обязать* автора репликации сделать это представляется мне радикальным, если не странным. Безусловно, существует проблема того, что методический раздел секции не всегда позволяет провести точную репликацию. Но почему тогда не обязать автора оригинальной работы указывать всю необходимую для потенциальной репликации информацию? Для этого уместно использовать не только методическую секцию, но и дополнительные к статье материалы, в которых можно разъяснить тонкости процедуры, если по каким-то причинам это нельзя сделать в самой статье. В конечном итоге, если шрифт инструкции и цвет бумаги, на которой она напечатана, оказывают значимое влияние на поведение испытуемого и об этом не сообщено, то это проблема автора, а не того, кто будет пытаться повторить его исследование (оставим в стороне вопрос о том, насколько интересны поведенческие эффекты, возникающие только при прочтении инструкции, напечатанной на белой бумаге формата A5 синим шрифтом Open Sans, 12 кеглем).

Такое предложение может не устроить критиков репликационных исследований. Как отмечает Р. Зваан, в то время как авторы репликаций рассматривают опубликованные исследования как общественную собственность, авторы оригинальных работ часто видят в них «произведения искусства», свою интеллектуальную собственность, в которой скрыто то, что сложно облечь в слова (Zwaan, 2014). Иначе говоря, репликация изначально дефектна, поскольку всегда упускает нечто невыразимое. Трудно, однако, не согласиться с Р. Звааном, которые пишет, что «метафора „эксперимент – это произведение искусства“ несостоятельна и противоречит самой сути научного знания» (ibid.).

Вообще, далеко не очевидно, что репликация обязана *предельно точно* воспроизводить оригинальное исследование. Напротив, очевидно, что повторение исследования невозможно без изменений. Так, и набор новой выборки, и повторное использование тех же испытуемых является существенным изменением исследования. И даже если воспроизведение осуществляет сам ученый, нельзя говорить о полной тождественности последнего самому себе в разные моменты времени. В целях дальнейшего анализа необходимо уточнить терминологию, что позволит разграничить разные варианты воспроизведения.

Нередко в дискуссии термины «воспроизводимость» и «реплицируемость» используются в качестве взаимозаменяемых. Так, представители Open Science Collaboration пишут: «Некоторые проводят различие между „воспроизводимостью“ и „реплицируемостью“, рассматривая первую как частный случай второй (например, в вычислительных науках) или наоборот (например, в биологических науках). Мы игнорируем это различие» (Open Science Collaboration, 2012, p. 659). Эти же авторы позднее предложили рассматривать воспроизводимость трояко. В самом узком смысле воспроизводимость – это повторение анализа существующих данных. В более широком смысле это прямая репликация, предусматривающая сбор новых данных при помощи методов оригинального исследования. Наконец, в самом широком смысле воспроизводимость – это концептуальная репликация, которая представляет собой «попытку валидизировать *интерпретацию* оригинального наблюдения путем манипуляции или измерения тех же концептуальных переменных, используя другие методы» (Open Science Collaboration, 2014, p. 300).

Исследователи, которые видят необходимость в разграничении воспроизводимости и реплицируемости, могут использовать эти понятия в противоположных смыслах, что отмечено в цитате, приведенной выше. Показателен случай с Ассоциацией вычислительной техники (ACM), крупнейшей и старейшей организацией в области компьютерных технологий. В 2016 г. она предприняла попытку унификации терминологии, опираясь на Международный словарь по метрологии (Association for Computing Machinery, 2016). С ее точки зрения, о *повторяемости* следовало говорить, когда исследование проводит одна и та же команда, используя одни и те же экспериментальные условия; о *реплицируемости* – когда исследование проводит другая команда, используя одинаковые с оригинальным исследованием экспериментальные условия; о *воспроизводимости* – когда исследование проводит другая команда, используя другие экспериментальные условия. Но в 2020 г. в ходе дискуссии с Национальной

организацией по стандартизации информации (NISO) ACM поменяло местами *воспроизводимость* и *реплицируемость* (Association for Computing Machinery, 2020). Более того, сейчас в области компьютерных технологий воспроизводимость предполагает не просто использование исходной процедуры и методов, но и использование исходных данных (National Academies of Sciences, Engineering and Medicine, 2019).

Еще одна попытка разграничения предпринята К. Драммондом (Drummond, 2018). Он предлагает различать воспроизводимость (reproducibility), реплицируемость (replicability) и перепроверяемость (retestability). *Воспроизводимость* требует того, чтобы оригинальное исследование было воспроизведено настолько точно, насколько это представляется уместным и разумным. По сути дела, целью воспроизведения исследования является независимая верификация результата. *Реплицируемость,* по мнению К. Драммонда, является статистическим понятием и связана с проверкой первичных результатов на независимой выборке. *Перепроверяемость* относится к устойчивости и генерализуемости полученных в исследовании результатов. В данном случае, как указывает К. Драммонд, реплицируется не столько исследование, сколько его результат: «Цель заключается в том, чтобы преднамеренно увеличивать различия между двумя экспериментами, чтобы определить, повлияют ли изменения на результат» (Drummond, 2018, p. 4). К. Драммонд утверждает, что именно перепроверяемость представляет собой золотой стандарт научного исследования, поскольку и воспроизводимость, и реплицируемость являются слишком слабыми критериями.

В психологии популярно уже упоминавшееся разделение репликаций на *прямые* и *концептуальные* (Nosek, Errington, 2017; Schmidt, 2009; Zwaan et al., 2018a). Об их определениях идут дискуссии (в частности, обсуждается вопрос о том, можно ли *прямую* репликацию считать *точной*), но обычно о прямой репликации говорят, когда предпринимается попытка воспроизвести ключевые элементы (не обязательно все!) оригинального исследования с намерением повторить его результат. Концептуальная репликация представляет собой попытку проверить теоретическую гипотезу оригинального исследования, целенаправленно используя другие методы.

Иногда отмечается важность учета того, кто проводит воспроизводящее исследование: сам автор (авторы) или другие ученые. Так, Ф. Штейнле *повторением* называет случай, когда воспроизведение осуществляет автор оригинального исследования, а *репликацией* – когда его осуществляют другие; термин *«воспроизведение»* используется

при этом как обобщающее понятие (Steinle, 2016, p. 41). Схожего разделения придерживается Ю. Шикоре (Schickore, 2011). Учитывается авторство и в типологии, которую предложили Й. Хёфмайер с соавт. (Hüffmeier et al., 2016). Они выделяют пять типов репликаций: *точная репликация* проводится самим автором (авторами), *близкая репликация* – другими исследователями, а оставшиеся три могут быть проведены и теми, и другими (*конструктивная репликация, концептуальная репликация в лаборатории* и *концептуальная репликация в поле*).

Некоторые авторы строят сложные классификационные схемы, учитывая, что репликационное и оригинальное исследование могут отличаться друг от друга разными аспектами. Так, Г. Бар с соавт. выделяют 16 типов репликаций, опираясь на 4 характеристики: время, место, испытуемые и методы (Bahr et al., 1983). Каждая характеристика рассматривается дихотомически (похожа репликация на оригинальное исследование или нет), а типы репликаций кодируются буквами (от *a* до *p*). Тип *p* представляет собой крайний вариант триангуляции, который позволяет ответить на вопрос, выдерживает ли гипотеза проверку, когда используются другие методы на другой выборке в другом месте и в другое время. Авторы отмечают, что нужно также учитывать и пятую характеристику – цели исследователя, что увеличивает число типов репликаций до 32.

Существуют и другие подходы к пониманию и классификации воспроизводящих исследований. Для целей настоящей статьи, однако, достаточно будет классификации, которая, по сути, представляет собой комбинацию подхода Ф. Штейнле с идеей, что воспроизведение может копировать оригинал с разной степенью точности. В качестве обобщающего понятия используется термин «воспроизводимость», а типы воспроизводящих исследований выделяются по двум основаниям: авторство и наличие вариаций (таблица 2).

Таблица 2

Классификация типов воспроизведений (reproduction)

Кто воспроизводит	**Существенные вариации в исследовании**	
	Нет	**Есть**
Повторение (сам исследователь/команда)	Конвергентное повторение (convergent repetition)	Дивергентное повторение (divergent repetition)
Репликация (другой исследователь/ другая команда)	Конвергентная репликация (convergent replication)	Дивергентная репликация (divergent replication)

Важно отметить следующее. Во-первых, вместо широко представленной в литературе дихотомии «прямая/концептуальная репликация» используется разделение воспроизводящих исследований на *конвергентные* и *дивергентные*. Идея концептуальной репликации, с моей точки зрения, весьма противоречива. Невозможно воспроизвести теорию. Использование других методов для проверки теоретической гипотезы оригинального исследования направлено не на воспроизведение, а на получение новых данных, позволяющих верифицировать (или фальсифицировать в зависимости от философских предпочтений ученого) исходную *теорию*. Это, безусловно, важный тип исследований, но по своим целями и результатам он существенно расходится с целями воспроизводящих исследований. Так, Р. Зваан с соавт. рассматривают «концептуальную репликацию» как ошибку именования, считая, что в этом случае корректнее говорить о проверке расширения теории на новый контекст (Zwaan et al., 2018). Они, однако, не отказываются от термина, чтобы сохранить связь с предшествующей литературой. М. Дерксен и Дж. Моравски рассматривают разделение между прямыми и концептуальными репликациями как бесполезное, а Носек и Эррингтон – как потенциально контрпродуктивное (Derksen, Morawski, 2022; Nosek, Errington, 2020). С точки зрения последних, «чтобы была репликация, две вещи должны быть истинными: результаты, совпадающие с предшествующим утверждением, увеличивают уверенность в этом утверждении, а несовпадающие результаты – снижают уверенность в утверждении. Многие „концептуальные репликации“ соответствуют первому критерию, но не второму» (Nosek, Errington, 2020, p. 7). Для концептуальной репликации, результаты которой не совпадают с оригинальной работой, могут быть и будут найдены многочисленные причины, объясняющие возникшее расхождение различиями методологии (вероятно, особо усердно такие причины будет искать автор исходного исследования). С другой стороны, если результаты будут совпадать, это с готовностью интерпретируют как поддержку результатов оригинального исследования. Такая «эпистемологическая асимметрия» не позволяет рассматривать концептуальные репликации как форму воспроизводящих исследований.

Вообще точка зрения, согласно которой воспроизведение должно стремиться как можно точнее повторить оригинальное исследование, не так уж наивна. Особенно это верно для репликаций, непосредственно следующих за исходным экспериментом во времени. С другой стороны, как уже указывалось, абсолютно точное воспроизведение невозможно, поэтому степень совпадения оригинально-

го и воспроизводящего исследований может быть различной. Кроме того, на каком-то этапе в воспроизводящее исследование могут преднамеренно вводиться различия с оригинальной работой, чтобы узнать, насколько стабилен результат, или чтобы сделать измерения более точными. Такие исследования я и называю *дивергентными*. Их ключевое отличие от концептуальных репликаций заключается в том, что они направлены именно на воспроизведение исходного исследования, а не на проверку теоретической гипотезы, лежащей в его основе. Степень дивергенции не должна превращать исследование, целями которого является вынесение суждения о результатах предшествующей работы, в исследование, верифицирующее или фальсифицирующее теорию. В *конвергентных* воспроизведениях целенаправленно минимизируются различия с исходным исследованием, таким образом, их недостижимым идеалом является *точное* повторение (на практике же они остаются воспроизведениями, которые называют *прямыми*).

Во-вторых, границы между повторением и репликацией также являются нечеткими. Так, если воспроизводящее исследование проводит другая команда с участием автора (или авторов) оригинального исследования, то оно не будет ни чистым повторением, ни чистой репликацией. Будет ли прямым повторением исследование, которое проводится частью первоначальной команды, но не включает других людей? Прямое повторение исследования самим автором спустя год или годы будет отличаться от прямого повторения, которое проводится сразу после окончания оригинального исследования.

Предложенную схему можно усложнить, если учесть, что возможно «воспроизведение» исследования на тех же самых эмпирических данных. В сфере вычислительных наук точное повторение оригинального исследования на том же программном обеспечении с использованием исходных данных называется «вычислительной воспроизводимостью» (Stodden, 2015), при этом иногда термины «вычислительная воспроизводимость» и «воспроизводимость» употребляются как взаимозаменяемые (National Academies of Sciences, Engineering and Medicine, 2019, p. 46). В области эмпирического знания такое отождествление не кажется оправданным. Таким образом, далее воспроизведение с получением новых данных будем назвать *воспроизведением* (более точно – *эмпирическим воспроизведением*), а воспроизведение, осуществляющееся на ранее полученных данных, – *вычислительным воспроизведением*. Последнее, по сути, представляет собой повторный анализ. Это вовсе не означает, что такие исследования лишены смысла. Напротив, по мере увеличения объема

собираемых данных, усложнения процедур их кодирования и методов математической обработки *вычислительное воспроизведение* будет востребоваться все больше. При этом сохраняется смысл в учете конвергенции/дивергенции воспроизведения, поскольку для старых «сырых» данных можно применить иные методы кодирования и математической обработки с тем, чтобы оценить устойчивость полученного ранее результата. Допустим, исследователь проводит изучение юмористического поведения у детей, используя в качестве «сырых» данных видеозаписи. Для проверки своих гипотез он использует ту или иную систему кодирования поведения, самостоятельно просматривая материал, а затем обрабатывает полученные данные, используя определенные методы статистической обработки и математического моделирования. Чтобы снизить риск ошибки, он может повторить все шаги (и тогда речь будет идти о конвергентном повторении) или, например, использовать при воспроизведении продвинутое программное обеспечение (если оно существует), позволяющее определять на видеозаписи разные типы юмористического поведения, и иные методы математической обработки (дивергентное повторение). Этим могут заняться другие исследователи, тогда воспроизведение примет форму репликации. Показательный пример серии параллельных вычислительных исследований представляет собой работа, в которой 29 групп ученых анализировали *один и тот же* набор данных, отвечая на *один и тот же вопрос:* дают ли футбольные арбитры красные карточки темнокожим игрокам с большей вероятностью, чем светлокожим (Silberzahn et al., 2018). Команды использовали разные аналитические подходы, при этом 20 команд (69%) обнаружили статистически значимый положительный эффект, а 9 команд (31%) – не обнаружили. Размер найденного эффекта при этом варьировался в достаточно широких пределах. Таким образом, результаты анализа сложных данных могут существенно зависеть от выбранных методов, что делает важным проведением вычислительных воспроизводящих исследований разного рода.

Учитывая вышесказанное и отказываясь от жесткой дихотомии между конвергентными и дивергентными воспроизводящими исследованиями, таблицу 1 можно преобразовать следующим образом (см. таблицу 3).

Подобно границам между конвергентным и дивергентным воспроизведением или между повторением и репликацией, границы между эмпирическим и вычислительным воспроизведением также размыты. В конечном итоге «сырых данных» не существует, все они в том или ином смысле приготовлены, отличается только степень

Таблица 3

Классификация типов воспроизведений (reproduction)

		Степень изменений по отношению к оригиналу					
		Низкая		← →		Высокая	
Новые данные	Сам исследователь/ команда	–	Полюс конвергенции	**Повторение** (repetition)	Полюс дивергенции	+	Эмпирическое воспроизведение
	Другой исследователь/ другая команда	–		**Репликация** (replication)		+	
Старые данные	Сам исследователь/ команда	–		**Вычислительное повторение** (computational repetition)		+	Вычислительно воспроизведение
	Другой исследователь/ другая команда	–		**Вычислительная репликация** (computational replication)		+	

«прожарки». Как пишет Дж. Боукер, «сырые данные – это и оксюморон, и плохая идея» (Bowker, 2006, p. 184). Будет ли повторный анализ видеозаписей, полученных другим ученым при изучении юмористического поведения детей, использованием «прежних данных» или их обработка будет извлечением новых данных?

В целом приведенный выше анализ показывает, что воспроизведение исследования отнюдь не означает исключительно его точное повторение. Ссылки на то, что воспроизводящее исследование как-то отличается от воспроизводимого, не являются аргументом в пользу того, что психология должна отказаться от критерия воспроизводимости (в широком смысле). Более того, можно утверждать, что воспроизводящее исследование *должно* отличаться от оригинального и при этом воспроизводить его результат (а если результат не воспроизведен, то определять переменные, влияющие на это). В чем научное значение исследования, результаты которого нельзя повторить на другой выборке, используя другие методы и пр.? Если мы получили, скажем, отрицательную корреляционную связь между интеллектом и религиозностью, используя некоторые оценочные шкалы, мы вправе ожидать, что использование других шкал, измеряющих интеллект и религиозность, даст схожий результат. Если же мы его не получаем, разумно предположить, что методы используемые либо в оригинальном, либо в повторяющем исследовании не-

удовлетворительны. Если на результат исследования влияет пол экспериментатора, национальность испытуемых или их культурный бэкграунд, время проведения исследования, цвет стен в комнате и пр., то все это необходимо эксплицировать и обозначить как переменные. Вариация условий исследования, по сути дела, выступает в качестве двигателя науки, но только в том случае, если эта вариация фиксируется и включается в корпус научных теорий. Вода закипает при достижении температуры 100 °C; однако это верно лишь при определенном атмосферном давлении, что должно быть эксплицировано. Таким образом, результаты психологического исследования должны быть воспроизводимы не только конвергентно, но и дивергентно, т. е. при целенаправленном изменении условий в некотором диапазоне.

Открытая наука

Как уже указывалось выше, одной из стратегий повышения воспроизводимости научного знания является построение открытой науки. Это связанный комплекс мероприятий, направленный на повышение прозрачности и открытости научной деятельности. Как сказано в «Манифесте открытой науки», «прозрачность превыше доверия» (Munafò et al., 2017, p. 5). Б. Носек с соавт. разработали правила по продвижению прозрачности и открытости (TOP), которые на разных уровнях приняты многими научными издательствами, включая Springer Nature и Американскую ассоциацию содействия развитию науки (Nosek et al., 2015). Среди психологических журналов, использующих TOP, можно назвать European Journal of Personality (Wiley), Experimental Psychology (Hogrefe), Journal of Personality Assessment (Taylor & Francis), Journal of Personality and Social Psychology (APA), Psychological Science (Sage) и др.

Принципы открытой науки предполагают публикацию данных, аналитических методов, кода, исследовательского дизайна и т. д., что обеспечивает доступность результатов для независимой проверки. Самый высший уровень TOP-правил требует, чтобы все это размещалось в доверенных репозиториях, а до публикации осуществлялось независимое воспроизведение проведенного анализа. Также в арсенал открытой науки входит пререгистрация исследований, обеспечивающая разработку теоретических оснований и выбор методов анализа до начала сбора данных. Как отмечают Носек с соавт., «пререгистрация планов анализа гарантирует различение конфирматорных и поисковых исследований или, как они еще на-

зываются, исследований, которые проверяют и которые генерируют гипотезы» (Nosek et al., 2015, p. 1423). Еще одной мерой является повышение прозрачности процедуры рецензирования статей. Ряд журналов уже внедрили практику публикации вместе со статьей полного протокола коммуникации между авторами, рецензентами и редакцией (Wolfram et al., 2020). Реализация стандартов открытой науки направлена на повышение доверия к научным результатам, развитие статистической грамотности и борьбу с сомнительными исследовательскими практиками, включая p-хакинг, харкинг (HARKing – выдвижение гипотез после того, как результаты станут известны) и пр.

Безусловно, открытая наука не является панацеей от мошенничества. Так, можно осуществлять пререгистрацию исследований *после* сбора данных, делать открытой только часть данных или размещать в репозиториях данные, которые частично или полностью придуманы. Еще в 1830 г. Ч. Бэббидж выделил четыре варианта мошенничества: мистификацию (*hoaxing*), подлог (*forging*), приукрашивание (*trimming*) и стряпню (*cooking*) (Babbage, 1830, p. 175). Политика открытых данных позволяет обнаружить приукрашивание или стряпню, только если автор открыл полные исходные данные. Обнаружить же *грамотную* мистификацию или подлог без проведения новых эмпирических исследований практически невозможно.

Открытая наука – критически важный элемент в повышении вычислительной реплицируемости научного знания. Публикация в открытых репозиториях данных, программного кода, аналитических процедур и пр. позволяет другим исследователям воспроизводить полученные результаты, используя те же или другие методы. Более того, как уже указывалось, репликация до публикации явным образом указана в третьем, наиболее строгом уровне TOP-правил. Очевидно, что речь здесь идет о вычислительном воспроизведении. Но если говорить об эмпирической воспроизводимости, то открытая наука влияет на нее только косвенно, делая прозрачнее нюансы процедуры исследования.

Статистическая реформа

Вторая стратегия повышения воспроизводимости, обозначенная в библиометрическом анализе, связана с призывами к реформам статистического вывода и статистической практики. Предлагается меньше внимания обращать на p-level, а больше на размер эффекта (Cohen, 1994; Flora, 2020); прекратить бездумно следовать ритуалам

«противоречивого гибрида» фишеровской проверки нулевой гипотезы и теории Неймана – Пирсона (Gigerenzer, 2004, 2018); больше внимания уделять байесовской статистике (Gelman, 2015) и пр. Некоторые исследователи полагают, что статистическая реформа сама по себе не является решением кризиса воспроизводимости, скорее требуется, чтобы ученые научились лучше делать выводы, опираясь и на частотную, и на байесовскую статистику (Colling, Szücs, 2021, p. 143). Соглашаясь с тем, что повышение статистической грамотности, бесспорно, полезно, я все же полагаю, что ни оно, ни предлагаемые статистические реформы не решают проблемы, которые выявил кризис воспроизводимости. Более того, смещение дискуссии, связанной с кризисом воспроизводимости, в статистическую плоскость пагубно, поскольку статистический анализ, сколь бы сложным он ни был, не способен заменить правильное измерение переменных, эксперимент и его непосредственное повторение. Усложнение статистического анализа в психологической науке, как правило, связано с отказом от экспериментального дизайна и увеличением списка переменных (корректность измерения которых в психологии всегда под вопросом), и нельзя допускать, чтобы чары статистики мешали видеть нам действительный источник проблемы. Как пишет Э. Дерман, «давайте не будем слишком сильно очаровываться наукой о данных, главные победы которой пока что заключаются в громогласной рекламе и попытках убедить мир в собственной значимости» (Дерман, 2017, с. 673).

Безусловно, ужесточение требований к используемым статистическим методам и обоснованию их использования может улучшить ситуацию, но вряд ли существенно. За прошедшие десятилетия в психологии глубоко укоренилась идея, что статистический анализ является неотъемлемой характеристикой психологического исследования (стал он неотъемлемой частью и репликационных исследований). Но это не так. Наука – это поиск количественных *законов,* и никакой статистический анализ не спасет исследование, проваленное на этапе измерения переменных или планирования дизайна. Как сказал Э. Дерман, статистика похожа на «неприкаянного дикаря, обитающего в области где-то посреди между математикой и естественными науками» (Дерман, 2017, с. 673). Следует сосредоточиться не столько на улучшении статистических практик, сколько на решении проблемы психологического измерения, ужесточении требований к экспериментальному дизайну, математизации языка психологии и, что самое, на мой взгляд, очевидное, повторении проведенных исследований.

«Заткнись и повторяй!»

Когда заходит речь о том, кто должен воспроизводить исследования, ответ обычно такой: другие ученые. Иными словами, воспроизведение должно проходить в форме репликаций. В недавней статье С. Левандовски и К. Оберауэр провели симуляцию эффективности двух «репликационных режимов»: *режима частных репликаций* и *режима публичных репликаций* (Lewandowsky, Oberauer, 2020). Режим частных репликаций предполагает, что исследователи повторяют любые полученные ими значимые результаты, и если повторение успешно, то публикуются *оба* исследования. Далее научное сообщество решает, какие результаты представляют интерес. Режим публичных репликаций предполагает, что исследователи публикуют любые значимые результаты без их повторения. Далее научное сообщество, не зная степени воспроизводимости этих результатов, определяет, какие из них интересны и реплицирует их. В качестве меры «интересности» использовался уровень цитирования. Итоговой мерой эффективности режима является соотношение цены («общее количество проведенных экспериментов») к произведенному знанию («открытые истинные эффекты»). По результатам симуляций с учетом предположений, которые сделали авторы, публичный режим продемонстрировал большую эффективность независимо от того, использовалась ли для анализа частотная или байесовская статистика. Для исследований, ориентированных на открытие, режим частных репликаций требует проведения примерно на 10% больше исследований, чем режим публичных публикаций. Если говорить об исследованиях, направленных на проверку теории, то эффективность второго режима выше на 40–50%. Полученные результаты устойчивы при введении в симуляцию различных отклонений от идеализированного научного поведения, таких как *p*-хакинг или мошенничество.

Таким образом, проведенная в рамках парадигмы эффективности затрат симуляция демонстрирует нецелесообразность частных репликаций или, если следовать предложенной выше терминологии, повторений. Однако, на мой взгляд, она обладает рядом ограничений, которые ставят под сомнение применимость полученных результатов к реальному исследовательскому процессу. Во-первых, в данной симуляции мерой произведенного знания являются не просто открытые истинные эффекты (для которых $p<0{,}05$), а реплицированные истинные эффекты, признанные интересными (т. е. часто цитируемые). Это видно, например, на графиках «произведенного знания», где по оси ординат отложено число открытых интересных эффектов,

а не число открытых истинных эффектов (Lewandowsky, Oberauer, 2020). «Открытый» здесь следует понимать как «реплицированный». Таким образом, для режима частных репликаций применялась вероятностная модель оценки того, что реплицированный результат будет признан научным сообществом интересным, а для режима публичных репликаций сначала применялось вероятностное решение, воспроизводить исследование или нет, а затем считалось количество успешных репликаций. При этом число открытых интересных истинных эффектов может не совпадать с числом открытых интересных значимых эффектов (так как реплицированный значимый эффект может быть ложным). Смещение по основанию «интересности» искажает симуляцию в пользу режима публичных репликаций, поскольку любой реплицированный в публичном режиме эффект является интересным, но не любой эффект, реплицированный в частном режиме, будет позднее признан интересным. При этом общее число реплицированных эффектов безотносительно к их «интересности» будет выше для режима частных репликаций, так как в рамках данной модели режим публичных репликаций не предусматривает «неинтересных» реплицированных результатов. Как отмечают сами авторы симуляции, есть некая ирония в том, что «эффективность» режима частных публикаций особо сильно страдает в исследованиях, направленных на проверку хороших теорий: большое число значимых эффектов, многие из которых истинные, автоматически приводит к росту количества дополнительных экспериментов (Lewandowsky, Oberauer, 2020, p. 4). Увы, далеко не все эти реплицированные эффекты потом вызовут интерес у других ученых.

Во-вторых, симуляция предполагает, что в случае успеха частной репликации публикуются оба исследования, но не эксплицировано, как осуществляется публикация: в рамках одной или в рамках двух статей. Вообще режим частных репликаций, хотя он и сопряжен с увеличением числа исследований до публикации, должен приводить к публикации меньшего количества статей (особенно, если и оригинальное исследование, и его повторение публикуются в одной работе). Таким образом, модель распределения цитирований, полученная на массиве текущих публикаций, может быть неприменима к режиму частных репликаций.

В-третьих, режим публичных репликаций оставляет открытыми вопросы, связанные с тем, кто и как именно будет заниматься воспроизведением чужих работ. Как писал А. Элтер, «в сложившейся научной практике слава приходит к тем, кто создает что-то новое, а не перепроверяет старое» (Элтер, 2017, с. 503). Авторы симуляции

осознают проблему и указывают, что режим публичных репликаций может выполнить свои обещания, если будут соблюдены 5 условий. «1. Нереплицированные результаты будут публиковаться предварительно и с эмбарго (например, в 1 год) на освещение в СМИ или цитирование. 2. Предварительным публикациям сопутствует приглашение к репликации другими исследователями. 3. Если репликация успешна, проводившие ее ученые становятся соавторами, а архивная публикация заменяет предварительную версию. 4. Провал репликации ведет к публичному отзыву предварительной публикации, чему сопутствует публичное подтверждение репликации, на которое можно ссылаться. Это гарантирует, что провалы репликаций будут известны и, следовательно, устранит публикационную предвзятость. 5. Если никто не пытается воспроизвести предварительные результаты, первоначальная публикация становится архивной после истечения срока действия эмбарго с примечанием, что она не вызвала интереса к репликации. Этот статус может измениться на 3) или 4), если репликация будет осуществлена позднее» (Lewandowsky, Oberauer, 2020, p. 10).

Эти условия, на мой взгляд, противоречат самой идее режима публичных репликаций. В предложенной симуляции в рамках этого режима реплицируются только исследования, достигшие определенного порога цитирования. Если имплементировать первое условие, запрещающее цитировать исследования, то как будут отбираться исследования для репликации? Далее, эти условия не отвечают на вопрос, кто будет проводить репликацию. Очевидно, те, кого заинтересует нецитируемое исследование или возможность стать автором публикации без особых творческих усилий. Поскольку условия не ограничивают, кто имеет право реплицировать чужие работы, что мешает малоизвестному ученому выбрать для репликации исследование, которое предварительно принято в высокорейтинговый журнал, и стать его соавтором (не очень понятно, конечно, на каких основаниях)? Более того, эти условия создают несправедливую ценностную асимметрию между оригинальным исследованием и его репликацией. Если репликация неуспешна, то на этом основании осуществляется отзыв исходной публикации. Но почему именно первоначальные результаты в этом случае рассматриваются как ошибочные? Неужели ошибочной не может быть репликация, в рамках которой могут быть критические ошибки или расхождения с оригинальной процедурой? В симуляции проверена устойчивость модели относительно случаев мошенничества, когда все первоначальные результаты признаются значимыми вне зависимости от ре-

ального исхода, а последующие репликации – честные. Но что мешает фальсифицировать исход самой репликации, при этом выгоду несет любая фальсификация: если незначимые результаты объявить значимыми, то станешь соавтором публикации, а если значимые – незначимыми, то получишь цитируемое признание, а статью авторов оригинального исследования отзовут (открывается невероятный простор для нечестной конкуренции и даже «мести»).

Таким образом, хотя проведенная симуляция представляет собой интересный подход к анализу кризиса воспроизводимости, она малоприменима к реальной научной деятельности. Безусловно, публичные репликации как независимая проверка научных результатов необходимы. Но обязанность воспроизведения должна быть возложена на самих ученых, публикующих отчет об исследовании. Сами ученые должны непосредственно повторить свое исследование (а, возможно, и повторить, внеся определенные преднамеренные изменения в оригинальный дизайн) и только потом предлагать его к публикации.

Важно отметить следующее. Во-первых, повторение исследования не ведет к появлению второй публикации: отчет о нем включается в одну статью. Безусловно, это может показаться неприятным для ученых, которые оценивают свою продуктивность количеством публикаций. Но публикация – это не просто отчет об исследовании, публикация – это отчет об исследовании, которое удовлетворяет определенным критериям. Почему нам кажется разумным требовать от психологов использования валидных методик и не публиковать статьи, в которых психологические конструкты оцениваются при помощи опросников из журнала Cosmopolitan? Потому что валидность – важный критерий оценки исследований. Почему мы требуем определенного числа испытуемых от исследований, в которых делается статистический вывод о популяции на основе выборки? Потому что это повышает надежность результатов. Почему не публикуются исследования, в которых данные собраны, но не обработаны? Потому что один из необходимых шагов исследования – обработка данных и их интерпретация. Почему же тогда мы не рассматриваем воспроизводимость как характеристику результатов, которая должна быть получена прямым образом? Безусловно, воспроизводимость можно пытаться оценивать по косвенным параметрам. Есть интересные данные, свидетельствующие о том, что люди без психологического образования успешно предсказывают результаты репликаций социально-психологических исследований в 59% случаев по их простому описанию, а если в описание включена информация о силе ис-

ходного результата, то успешность предсказаний возрастает до 67% (Hoogeveen et al., 2020). Также изучается возможность использования для прогноза успешности репликаций так называемых рынков предсказаний, в которых процент успеха достигает 70–75% (Dreber et al., 2015; Forsell et al., 2019). Очевидная проблема подобных исследований заключается в том, что для оценки точности прогноза необходима информация о том, успешна репликация или нет, т. е. они не могут заменить прямую оценку воспроизводимости, что отмечают и сами авторы. Вероятно, прогнозы простых людей или участников рынка предсказаний могут использоваться для отбора исследований, которые следует реплицировать. Это, однако, не отменяет необходимости обеспечить воспроизводимость результатов на уровне повторяемости до их публикации.

Приведу слова К. Поппера, который утверждает, что ученые не только должны сами воспроизводить свои эксперименты, но и давать подробные инструкции, следуя которым это может сделать любой исследователь: «Каждый физик-экспериментатор знает те поразительные и необъяснимые мнимые „эффекты“, которые могут даже в течение некоторого времени воспроизводиться в его лаборатории, но которые затем исчезают бесследно. Конечно, ни один физик в таком случае не скажет, что он совершил научное открытие (хотя он и может попытаться так перестроить свой эксперимент, чтобы сделать этот результат воспроизводимым). В действительности имеющий научную значимость физический эффект следует определить как такой, который может быть неоднократно воспроизведен любым человеком, выполняющим соответствующий эксперимент предписанным образом. Ни один серьезный физик не предложил бы для публикации в качестве научного открытия сообщение о любом таком „оккультном эффекте“ (как я предлагаю называть явления такого рода), для воспроизведения которого он не мог бы дать никаких инструкций. Такого рода „открытие“ было бы немедленно отвергнуто как химерическое просто потому, что попытки проверить его привели бы к отрицательным результатам. (Отсюда следует, что любые споры по вопросу о том, действительно ли встречаются события, которые в принципе неповторимы и уникальны, не могут быть разрешены наукой, это споры в области метафизики)» (Поппер, 2005, с. 43). Это справедливо не только в отношении физиков, но и других ученых, включая психологов. О том же говорит Р. Фейнман в своей знаменитой речи о науке-самолетопоклонников (наука карго-культа) (Feynman, 1974). Научная честность требует, чтобы сам ученый повторял собственные эксперименты, выискивая в сво-

их исследованиях любые нюансы, которые могут подорвать доверие к полученным результатам, и сообщая о них. На практике это означает, что будет нельзя принимать к публикации исследования, которые проведены единожды, пусть и на достаточно большой выборке. Очевидно, что такое требование входит в противоречие с доминирующим принципом «публикуйся или умри». Также это требование отлично от идеи так называемых пререгистрационных прямых репликационных исследований (Lindsay, 2017).

Таким образом, идея заключается в том, чтобы рассматривать повторение как стадию научного исследования, а не отдельное исследование. Если на этапе проектирования дизайна исследования мы создаем процедуру и отбираем методы, минимизируя угрозы конструктной и внутренней валидности исследования, а на этапе обработки результатов пытаемся обеспечить валидность статистических выводов, то на этапе повторения мы убеждаемся в воспроизводимости полученных результатов. Более того, нет особого смысла интерпретировать результаты до того, как мы убедимся в их воспроизводимости. Иными словами, в идеализированной модели научного исследования между этапами сбора и анализа данных и этапом интерпретации результатов следует поставить этап повторения.

Принятие подобной практики, на мой взгляд, может принести немало положительных результатов. Во-первых, снизится общее количество публикаций, что уменьшит информационную нагрузку на исследователей. Возможно, это и не приведет к утопии «одна-статья-в-год» (Nelson et al., 2012), но заставит ученого больше исследовать и меньше писать. Кроме того, уменьшение числа публикаций способно изменить существующий паттерн цитирования, когда значительная часть публикаций не цитируется или цитируется очень мало (хотя, безусловно, цитируемость не является самоцелью научной работы). Так, в области медицинских наук в рамках 27-летнего «окна» (с 1990 по 2017 г.) ни разу не цитировались около 12% публикаций, в области социальных наук – около 30%, а в области искусства и гуманитарных наук – порядка 70% (Sugimoto, Larivière, 2018).

Во-вторых, императив повторений заставит ученого подходить более ответственно к выбору темы. Никто не хочет повторять «пустые» исследования, и сам их автор в первую очередь. Если исследователь верит, что он действительно открыл истину, не он ли должен быть первым человеком, который захочет увидеть ее повторно? Прекрасно об этом сказал И. П. Павлов: «Мало того, чтобы истину захватить и этим удовлетвориться. Истиной надо любоваться, ее надо любить. Когда я был в молодые годы за границей и слушал великих профес-

соров-стариков, я был изумлен, каким образом они, читавшие по десяткам лет лекции, тем не менее читают их с таким подъемом, с такою тщательностью ставят опыты. Тогда я это плохо понимал. А затем, когда мне самому пришлось сделаться стариком, – это для меня стало понятно. Это совершенно естественная привычка человека, который открывает истины. У такого человека есть потребность постоянно на эту истину смотреть. Он знает, чего это стоило, каких напряжений ума, и он пользуется каждым случаем, чтобы еще раз убедиться, что это действительно твердая истина, несокрушимая, что она всегда такая же, как и в то время, когда была открыта. И вот теперь, когда я ставлю опыты, я думаю, едва ли есть хоть один слушатель, который бы с таким интересом, с такой страстью смотрел на них, как я, видящий это уже в сотый раз» (Павлов, 2001б, с. 105–106). И добавляет: «Люди часто проводят всю жизнь в кабинете, отыскивая истину. Но это стремление распадается на два акта. Во-первых, стремление к приобретению новых истин, любопытство, любознательность. А другое – это стремление постоянно возвращаться к добытой истине, постоянно убеждаться и наслаждаться тем, что то, что ты приобрел, есть действительно истина, а не мираж. Одно без другого теряет смысл. Если вы обратитесь к молодому ученому, научному эмбриону, то вы отчетливо видите, что стремление к истине в нем есть, но у него нет стремления к абсолютной гарантии, что это – истина. Он с удовольствием набирает результаты и не задает вопроса, а не есть ли это ошибка? В то время как ученого пленяет не столько то, что это новизна, а что это действительно прочная истина. А что же у нас? А у нас прежде всего первое – это стремление к новизне, любопытство. Достаточно нам что-либо узнать, и интерес наш этим кончается. („А, это все уже известно“.)» (Павлов, 2001а, с. 122).

В-третьих, императив повторений будет способствовать уходу от ориентации на ритуалы статистических практик и большие выборки. Часто именно необходимость повторно собирать данные на многих людях повышает «стоимость» исследования и отвращает ученых от повторения собственных работ. Вообще широкое распространение статистики в психологии – это некий субститут количественных законов, который позволяет сделать вид, что идет прирост научного знания. Переориентация на дизайны с малым N упростит проведение повторений и репликаций, и, можно надеяться, сместит акцент с использования статистики на установление количественных законов, связывающих измеренные переменные. Как еще в 1960-х годы отмечал Р. Бауэрфайнд, призывая к проведению воспроизводящих исследований, «несколько репликаций обеспечат гораздо более су-

щественные результаты, чем статистика стандартных ошибок, примененная к одному набору данных. Более того, множественные репликации устранят необходимость применения индуктивной статистики» (Bauernfeind, 1968, p. 126). Возможно, итоговая «стоимость» исследований в рамках режима повторений окажется ниже, чем в рамках режима, ориентированного на изучение больших выборок.

В-четвертых, возникнет богатое многообразие форм воспроизводящих исследований. Успешное повторение или репликация – это не обязательно исследование, снова получившее результат на уровне $p<0,05$. Как остроумно замечает автор статьи, описывающей летучих мышей, поедающих лягушек: «Ели ли летучие мыши лягушек? Да, мы это видели. Повторили ли мы наблюдение? Да, мы поймали несколько летучих мышей, лакомящихся этими лягушками» (Ryan, 2011, p. 1229).

Вместо того чтобы искать причины не повторять свои исследования, следует закатать рукава и взяться за работу. Можно перефразировать знаменитый, пусть и несколько грубоватый, принцип квантовой механики: «Заткнись и вычисляй!». Любовь к истине вместе с научной честностью, налагающей на ученого обязанность повторить собственное исследование *до* публикации его результатов, и есть путь выхода из кризиса воспроизводимости, который является лишь одним из проявлений общего кризиса психологической науки.

Технологическая кода

В заключение следует отметить, что в статье намеренно исключен из рассмотрения вопрос, какую сущность имеет научное знание, продуцируемое психологией. В действительности, это не влияет на то, важна ли для психологии воспроизводимость результатов. Если производится научное знание, то оно должно быть в той или иной форме воспроизводимо; если знание невоспроизводимо, то оно не имеет отношения к науке. Мой краткий ответ заключается в том, что психология должна осознать себя как технологическую область знания, в рамках которой изучаются естественные и конструируются искусственные психологические системы. Законы, определяющие функционирование таких систем, имеют биологическую (шире – физическую) природу и *per se* выходят за пределы психологии. Воспроизводимость психологического знания заключается, прежде всего, в том, чтобы демонстрировать стабильность результата технологической реализации этих законов в рамках психологических систем. Иными словами, воспроизводить не наблюдение, но изменение.

Литература

Вачков И. В., Вачкова С. Н. Воспроизводимость психологических экспериментов как проблема постнеклассической науки // Культурно-историческая психология. 2016. Т. 12. № 1. С. 97–101.

Дерман Э. Сила статистики // Эта идея должна умереть. Научные теории, которые блокируют прогресс. М.: АСТ, 2017. С. 671–673.

Павлов И. П. Об уме вообще // Рефлекс свободы. СПб.: Питер, 2001а. С. 96–106.

Павлов И. П. О русском уме // Рефлекс свободы. СПб.: Питер, 2001б. С. 107–125.

Петренко В. Ф. К проблеме коллективного бессознательного в рамках философии постнеклассической рациональности и психологии конструктивизма // Вопросы философии. 2018. № 2. С. 89–101.

Поппер К. Логика научного исследования. М.: Республика, 2005.

Стебаков Д. А. Психологический эксперимент в контексте проблемы сложности // Философия науки и техники. 2017. Т. 22. № 1. С. 46–60.

Фёдоров А. А. Постнеклассическая психология и псевдонаука: реалистическая перспектива // Вопросы философии. 2020. № 7. С. 93–103.

Элтер А. Повторяемость – это гарантия // Эта идея должна умереть. Научные теории, которые блокируют прогресс. М.: АСТ, 2017. С. 502–504.

Anderson C. J., Bahník Š., Barnett-Cowan M., Bosco F. A., Chandler J., Chartier C. R., Cheung F. et al. Response to Comment on "Estimating the reproducibility of psychological science" // Science. 2016. V. 351. № 6277. P. 1037–1037.

Association for Computing Machinery. Artifact Review and Badging – Version 1.0. 2016. URL: https://www.acm.org/publications/policies/artifact-review-badging (дата обращения: 08.05.2022).

Association for Computing Machinery. Artifact Review and Badging – Current. 2020. URL: https://www.acm.org/publications/policies/artifact-review-and-badging-current (дата обращения: 08.05.2022).

Babbage C. Reflections on the Decline of Science in England: And on Some of Its Causes. L.: B. Fellowes, 1830.

Bahr H. M., Caplow T., Chadwick B. A. Middletown III: Problems of Replication, Longitudinal Measurement and Triangulation // Annual Review of Sociology. 1983. V. 9. № 1. P. 243–264.

Bauernfeind R. H. The Need for Replication in Educational Research // The Phi Delta Kappan. 1968. V. 50. № 2. P. 126–128.

Bench S. W., Rivera G. N., Schlegel R. J., Hicks J. A., Lench H. C. Does expertise matter in replication? An examination of the reproducibility project: Psychology // Journal of Experimental Social Psychology. 2017. V. 68. P. 181–184.

Bowker G. C. Memory Practices in the Sciences. Cambridge, MA, USA: MIT Press, 2006.

Camerer C. F., Dreber A., Forsell E., Ho T.-H., Huber J., Johannesson M., Kirchler M. et al. Evaluating replicability of laboratory experiments in economics // Science. 2016. V. 351. № 6280. P. 1433–1436.

Camerer C. F., Dreber A., Holzmeister F., Ho T.-H., Huber J., Johannesson M., Kirchler M. et al. Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015 // Nature Human Behaviour. 2018. V. 2. № 9. P. 637–644.

Cohen J. The earth is round (p <.05) // American Psychologist. 1994. V. 49. № 12. P. 997–1003.

Colling L. J., Szücs D. Statistical Inference and the Replication Crisis // Review of Philosophy and Psychology. 2021. V. 12. № 1. P. 121–147.

Cova F., Strickland B., Abatista A., Allard A., Andow J., Attie M., Beebe J. et al. Estimating the Reproducibility of Experimental Philosophy // Review of Philosophy and Psychology. 2021. V. 12. № 1. P. 9–44.

Derksen M., Morawski J. Kinds of Replication: Examining the Meanings of "Conceptual Replication" and "Direct Replication" // Perspectives on Psychological Science. 2022. doi: 1177/17456916211041116

Dreber A., Pfeiffer T., Almenberg J., Isaksson S., Wilson B., Chen Y., Nosek B. A. et al. Using prediction markets to estimate the reproducibility of scientific research // Proceedings of the National Academy of Sciences. 2015. V. 112. № 50. P. 15343–15347.

Drummond C. Reproducible research: a minority opinion // Journal of Experimental & Theoretical Artificial Intelligence. 2018. V. 30. № 1. P. 1–11.

Ebersole C. R., Atherton O. E., Belanger A. L., Skulborstad H. M., Allen J. M., Banks J. B., Baranski E. et al. Many Labs 3: Evaluating participant pool quality across the academic semester via replication // Journal of Experimental Social Psychology. 2016. V. 67. P. 68–82.

Fanelli D. Is science really facing a reproducibility crisis, and do we need it to? // Proceedings of the National Academy of Sciences. 2018. V. 115. № 11. P. 2628–2631.

Feynman R. P. Cargo Cult Science // Engineering and Science. 1974. V. 37. № 7. P. 10–13.

Flora D. B. Thinking about effect sizes: From the replication crisis to a cumulative psychological science // Canadian Psychology/Psychologie canadienne. 2020. V. 61. № 4. P. 318–330.

Forsell E., Viganola D., Pfeiffer T., Almenberg J., Wilson B., Chen Y., Nosek B. A. et al. Predicting replication outcomes in the Many Labs 2 study // Journal of Economic Psychology. 2019. V. 75. P. 102–117.

Frideres J. Participatory research: An illusionary perspective // A World of Communities: Participatory Research Perspectives. North York: Captus University Publications, 1992.

Gelman A. The connection between varying treatment effects and the crisis of unreplicable research: a Bayesian perspective // Journal of Management. 2015. V. 41. № 2. P. 632–643.

Gigerenzer G. Mindless statistics // The Journal of Socio-Economics. 2004. V. 33. № 5. P. 587–606.

Gigerenzer G. Statistical rituals: the replication delusion and how we got there // Advances in Methods and Practices in Psychological Science. 2018. V. 1. № 2. P. 198–218.

Gilbert D. T., King G., Pettigrew S., Wilson T. D. Comment on "Estimating the reproducibility of psychological science" // Science. 2016. V. 351. № 6277. P. 1037–1037.

Hoogeveen S., Sarafoglou A., Wagenmakers E.-J. Laypeople can predict which social-science studies will be replicated successfully // Advances in Methods and Practices in Psychological Science. 2020. V. 3. № 3. P. 267–285.

Hüffmeier J., Mazei J., Schultze T. Reconceptualizing replication as a sequence of different studies: A replication typology // Journal of Experimental Social Psychology. 2016. V. 66. P. 81–92.

Kahneman D. A New Etiquette for Replication // Social Psychology. 2014. V. 45. № 4. P. 310–311.

Klein R. A., Ratliff K. A., Vianello M., Adams R. B. Jr, Bahník Š., Bernstein M. J., Bocian K. et al. Investigating variation in replicability: A "many labs" replication project // Social Psychology. 2014. V. 45. № 3. P. 142–152.

Klein R. A., Vianello M., Hasselman F., Adams B. G., Adams R. B., Alper S., Aveyard M. et al. Many Labs 2: Investigating variation in replicability across samples and settings // Advances in Methods and Practices in Psychological Science. 2018. V. 1. № 4. P. 443–490.

Kohn A. False prophets. Fraud and error in science and medicine. Oxford: Basil Blackwell, 1986.

Lewandowsky S., Oberauer K. Low replicability can support robust and efficient science // Nature Communications. 2020. V. 11. № 1. P. 358.

Lindsay D. S. Preregistered direct replications in psychological science // Psychological Science. 2017. V. 28. № 9. P. 1191–1192.

Martin B. Scientific fraud and the power structure of science // Prometheus. 1992. V. 10. № 1. P. 83–98.

Munafò M. R., Nosek B. A., Bishop D. V. M., Button K. S., Chambers C. D., Percie du Sert N., Simonsohn U. et al. A manifesto for reproducible science // Nature Human Behaviour. 2017. V. 1. № 1. P. 1–9.

National Academies of Sciences, Engineering and Medicine. Reproducibility and Replicability in Science. Washington, DC: The National Academies Press, 2019.

Nelson L. D., Simmons J. P., Simonsohn U. Let's publish "fewer" papers // Psychological Inquiry. 2012. V. 23. № 3. P. 291–293.

Nosek B. A., Alter G., Banks G. C., Borsboom D., Bowman S. D., Breckler S. J., Buck S. et al. Promoting an open research culture // Science. 2015. V. 348. № 6242. P. 1422–1425.

Nosek B. A., Hardwicke T. E., Moshontz H., Allard A., Corker K. S., Dreber A., Fidler F. et al. Replicability, Robustness and Reproducibility in Psychological Science // Annual Review of Psychology. 2022. V. 73. № 1. P. 719–748.

Nosek B. A., Errington T. M. Making sense of replications // eLife. 2017. V. 6. P. e23383.

Nosek B. A., Errington T. M. What is replication? // PLoS Biology. 2020. V. 18. № 3. e3000691.

Open Science Collaboration. An open, large-scale, collaborative effort to estimate the reproducibility of psychological science // Perspectives on Psychological Science. 2012. V. 7. № 6. P. 657–660.

Open Science Collaboration. The Reproducibility Project: A model of large-scale collaboration for empirical research on reproducibility // Implementing Reproducible Research / V. Stodden, F. Leisch, R. D. Peng (Eds). Boca Raton: Chapman and Hall/CRC, 2014. P. 299–323.

Open Science Collaboration. Estimating the reproducibility of psychological science // Science. 2015. V. 349. № 6251. P. aac4716.

Pashler H., Harris C. R. Is the replicability crisis overblown? Three arguments examined // Perspectives on Psychological Science. 2012. V. 7. № 6. P. 531–536.

Peels R., Bouter L. The possibility and desirability of replication in the humanities // Palgrave Communications. 2018. V. 4. № 1. P. 95.

Ryan M. J. Replication in field biology: the case of the frog-eating bat // Science. 2011. V. 334. № 6060. P. 1229–1230.

Schickore J. The significance of re-doing experiments: a contribution to historically informed methodology // Erkenntnis. 2011. V. 75. № 3. P. 325–347.

Schmidt S. Shall we really do it again? The powerful concept of replication is neglected in the social sciences // Review of General Psychology. 2009. V. 13. № 2. P. 90–100.

Silberzahn R., Uhlmann E. L., Martin D. P., Anselmi P., Aust F., Awtrey E., Bahník Š. et al. Many analysts, one data set: Making transparent how variations in analytic choices affect results // Advances in Methods and Practices in Psychological Science. 2018. V. 1. № 3. P. 337–356.

Soto C. J. How Replicable are links between personality traits and consequential life outcomes? The life outcomes of personality replication project // Psychological Science. 2019. V. 30. № 5. P. 711–727.

Steinle F. Stability and replication of experimental results: a historical perspective // Reproducibility: Principles, Problems, Practices and Prospects / H. Atmanspacher, S. Maasen (Eds). Hoboken: John Wiley & Sons Ltd, 2016. P. 39–63.

Stodden V. Reproducing statistical results // Annual Review of Statistics and Its Application. 2015. V. 2. № 1. P. 1–19.

Sugimoto C. R., Larivière V. Measuring Research: What Everyone Needs to Know®. Oxford–N. Y.: Oxford University Press, 2018.

Wolfram D., Wang P., Hembree A., Park H. Open peer review: promoting transparency in open science // Scientometrics. 2020. V. 125. № 2. P. 1033–1051.

Zwaan R. Trying to Understand both Sides of the Replication Discussion. 2014. URL: https://rolfzwaan.blogspot.com/2014/05/trying-to-understand-both-sides-of.html (дата обращения: 19.04.2022).

Zwaan R. A., Etz A., Lucas R. E., Donnellan M. B. Making replication mainstream // Behavioral and Brain Sciences. 2018a. V. 41. e120.

Zwaan R. A., Pecher D., Paolacci G., Bouwmeester S., Verkoeijen P., Dijkstra K., Zeelenberg R. Participant Nonnaiveté and the reproducibility of cognitive psychology // Psychonomic Bulletin & Review. 2018b. V. 25. № 5. P. 1968–1972.