

Научная статья

УДК 159.9

DOI 10.25205/2658-4506-2024-17-1-41-61

Личностные и безличные моральные дилеммы: объяснительные модели и критерии разграничения

Алексей Павлович Шабалин

Новосибирский государственный университет

Новосибирск, Россия

a.shabalin3@g.nsu.ru

Аннотация

В статье рассматривается проблема классификации жертвенных моральных дилемм, активно используемых в психологии и нейрофизиологии морального выбора. Анализируются истоки и развитие противопоставления личностных и безличных сценариев, предложенное в рамках теории двойного процесса Дж. Грина, а также альтернативные подходы, включая концепцию универсальной моральной грамматики Дж. Михаила и модели, основанные на вычислительной нейронауке. Особое внимание уделяется критериям разграничения, их эволюции и современным изменениям в подходах к классификации дилемм. Подчеркивается важность дальнейшего уточнения параметров классификации для повышения валидности исследований морального суждения.

Ключевые слова

моральные дилеммы, теория двойного процесса, универсальная моральная грамматика, дилемма вагонетки, дилемма моста, доктрина двойного эффекта

Для цитирования

Шабалин А. П. Личностные и безличные моральные дилеммы: объяснительные модели и критерии разграничения // Reflexio. 2024. Т. 17, № 1. С. 41–61. DOI 10.25205/2658-4506-2024-17-1-41-61

© Шабалин А. П., 2024

Personal and Impersonal Moral Dilemmas: Explanatory Models and Distinction Criteria

Alexey P. Shabalin

Novosibirsk State University,
Novosibirsk, Russian Federation
a.shabalin3@g.nsu.ru

Abstract

The article considers the problem of classification of sacrificial moral dilemmas actively used in psychology and neurophysiology of moral reasoning. The origins and development of the opposition of personal and impersonal scenarios proposed within the framework of J. Greene's dual process theory are analyzed, as well as alternative approaches, including the concept of universal moral grammar by J. Mikhail and models based on computational neuroscience. Particular attention is paid to the distinction criteria, their evolution and modern changes in approaches to the classification of dilemmas. The importance of further clarification of the classification parameters for increasing the validity of moral judgment studies is emphasized.

Keywords

Moral dilemmas, dual process theory, universal moral grammar, trolley dilemma, footbridge dilemma, doctrine of double effect

For citation

Shabalin A. P. Personal and impersonal moral dilemmas: explanatory models and distinction criteria. *Reflexio*, 2024, vol. 17, no. 1, pp. 41–61. (in Russ.) DOI 10.25205/2658-4506-2024-17-1-41-61

Введение

Использование дилемм является стандартной практикой в рамках психологических исследований морали уже более 60 лет, начиная с работ Лоуренса Колберга, признаваемого основоположником «рационалистского» подхода к моральным суждениям (см. [Greene et al., 2001; Christensen & Gomila, 2012]) – более того, предложенная им «дилемма Хайнца», вопрос о допустимости совершения кражи во имя спасения близкого человека, активно применяется в качестве стимульного материала по сей день, совершенно вне контекста оригинальной теории морального развития (см., к примеру, [Vezirian et al., 2025]). Так, согласно данным обзора, проведенного Н. Эллемерс и соавторами, почти треть (30 %) задействованных в рассмотренных ими работах по психологии морали количественных методик опираются на гипотетические моральные дилеммы, предполагающие выбор между двумя альтернативами в соответствии с теми или иными моральными принципами [Ellemers et al., 2019].

Традиционно моральные дилеммы понимаются как короткие истории о ситуациях, содержащих в себе моральный конфликт, т. е. противоречие между двумя несовместимыми способами поведения, обусловленными конкури-

рующими моральными соображениями [Christensen & Gomila, 2012]. Отдельным классом моральных дилемм, которого и будет касаться данный обзор, являются *жертвенные дилеммы* (*sacrificial dilemmas*) – описания гипотетических ситуаций, где единственным способом спасти группу людей от опасности является причинение вреда (очень часто – смерти) другому человеку или группе людей. Примечательно, что жертвенные дилеммы зачастую носят вычурный, гротескный характер или же просто не отличаются сколько-нибудь высоким реализмом, что порождает сомнения в валидности получаемых с их помощью результатов (см. [Bauman et al., 2014]); тем не менее вопрос о том, является ли низкая экологическая валидность и реалистичность жертвенных дилемм принципиальным препятствием на пути к их использованию, можно признать дискуссионным (см., к примеру, [Фёдоров, Злобина, 2023]).

Как бы то ни было, активно продолжающееся использование жертвенных дилемм в рамках психологии и нейроэтики предполагает учет специфики данной разновидности стимульного материала: если перед исследователями стоит задача изучения связей между отдельными аспектами предъявляемых испытуемым дилемм и демонстрируемыми поведенческими реакциями, актуальным становится вопрос выделения релевантных в контексте выносимых моральных суждений параметров стимульных ситуаций и, как следствие, проблема классификации последних. Таким образом, целью настоящей статьи является анализ существующих подходов к классификации жертвенных моральных дилемм с акцентом на дихотомию так называемых *личностных* (*personal*) и *безличных* (*impersonal*)¹ сценариев.

Вагонетка, мост и теория двойного процесса

Хрестоматийным примером жертвенной моральной дилеммы является впервые сформулированная в 1967 г. философом Филиппой Фут *проблема*, или *дилемма вагонетки* (*trolley problem*) – сценарий, согласно наиболее известной версии которого тяжелая неуправляемая вагонетка движется по рельсам навстречу пяти рабочим, и единственный способ их спасти заключается в том, чтобы перевести стрелку на запасной путь, где находится только один рабочий, тем самым убив его [Foot, 1967]. Аналогичным образом, в схожей *дилемме моста*, предложенной Джудит Джарвис Томсон в 1976 г., единственный способ остановить вагонетку и спасти пятерых рабочих от гибели заключается в том, чтобы столкнуть с пешеходного моста крупного прохожего на проходящие внизу рельсы [Thomson, 1976]. Суть данных мысленных экспериментов сводится к противопоставлению *консеквенциалистской* (в частности, *утилитаристской*) и *деонтологической* этики; если первый класс нормативных этических теорий постулирует допустимость или обязательность действия

¹ В русскоязычной литературе, касающейся моральных дилемм, встречаются переводы эпитета “personal” и как «личностный», и как «личный»; в рамках данной статьи используется первый вариант. Однако, пожалуй, наиболее адекватным по смыслу переводом должен быть «влияющая на конкретного человека» или «затрагивающая конкретного человека».

или бездействия в том случае, если оно ведет к наилучшим возможным последствиям в свете соотношения выгод и издержек, то второй предполагает верховенство определенных этических правил независимо от возможных последствий (см., к примеру, [Schaich Borg et al., 2006; Ryazanov et al., 2023]). При этом в свете дальнейшего обсуждения следует упомянуть следующие деонтологические принципы:

1. *Доктрина деяния и допущения (doctrine of doing and allowing)*. Согласно данному постулату, вред, причиненный действием, сложнее оправдать, чем вред, причиненный бездействием.

2. *Доктрина двойного эффекта (doctrine of double effect)*. В соответствии с этим постулатом сложнее оправдать вред, причиненный как цель или как средство для достижения цели, чем как предвидимый, но неумышленный побочный результат достижения цели [Schaich Borg et al., 2006; Cushman & Young, 2011; Ryazanov et al., 2023].

Возвращаясь к дилеммам вагонетки и моста, следует заметить, что данные гипотетические сценарии, будучи перенесенными из сферы нормативной этики в область психологии и нейрофизиологии, во многом определили направление эмпирических исследований в области морали в начале XXI в. К этому моменту доминирующие в психологии взгляды относительно природы моральных суждений неоднократно кардинально изменялись – от эмоциональных и нерациональных процессов в оптике психоанализа к рационализму в духе идей Ж. Пиаже и Л. Колберга, одержавшему триумф после «когнитивной революции», но который снова уступил место аффекту и автоматическим процессам после 1980-х гг. [Greene & Haidt, 2002].

Своеобразной поворотной точкой стало исследование Дж. Грина и соавторов [Greene et al., 2001], обративших свое внимание на «загадку для моральных философов»: в случае проблемы вагонетки спасение пятерых человек ценой жизни одного («утилитаристский» сценарий) оказывается для людей более морально приемлемым, чем в случае дилеммы моста, чаще вызывающей «деонтологические» ответы у испытуемых. Выдвинутая исследователями гипотеза гласила, что «...некоторые моральные дилеммы (относительно схожие с дилеммой моста) задействуют эмоциональную обработку в большей степени, чем другие (относительно схожие с дилеммой вагонетки), и эти различия в эмоциональной вовлеченности влияют на суждения людей» [Ibid. P. 2106].

В рамках проведенного исследования с использованием методов нейровизуализации в качестве стимульного материала Грин и соавторы использовали набор из 60 дилемм, условно разделенных на «моральную» и «неморальную» категории. В свою очередь, дилеммы из категории «моральных» также разбивались на две подгруппы, исходя из их сходства с дилеммой моста или дилеммой вагонетки – предполагалось, что в первом случае совершаемое действие, приводящее к гибели одного человека, интуитивно переживается как более «близкое и личное» (*close and personal*), а во втором – как «без-

личное» (*impersonal*); как можно понять, это и положило начало выделению групп личностных и безличных моральных дилемм. В качестве формальных критериев отнесения к одной из данных категорий Грин и соавторы выделили следующие пункты:

- а) можно ли обоснованно ожидать, что совершаемое действие приведет к серьезным телесным повреждениям;
- б) является ли вред результатом перенаправления существующей угрозы на другого человека/группу лиц;
- с) будет ли причинен результирующий вред отдельному человеку или участникам группы людей (находится ли жертва «на сцене» дилеммы) [Ibid.].

Те дилеммы, которые (а) ожидаемо должны были привести к серьезным телесным повреждениям у (б) отдельного человека или участников некоей группы людей, и в которых (с) вред не является результатом перенаправления существующей угрозы на других людей, были отнесены «моральным личностным», а остальные – к «моральным безличным». В дальнейшем Дж. Грин предложил более лаконичную формулу ME HURT YOU, по-прежнему предполагающую причинение вреда (HURT), наличие «жертвы» (YOU) и агентность, «авторство» самого субъекта (ME) [Greene & Haidt, 2002; Greene et. al., 2004]. Кроме того, личностные моральные дилеммы были разбиты на простые и сложные: если в первом случае негативная социально-эмоциональная реакция доминирует над слабыми или несуществующими «когнитивными» доводами в пользу действия, что приводит к быстрым и единообразным ответам у всех испытуемых, то во втором негативная социально-эмоциональная реакция вступает в борьбу с более абстрактным «когнитивным» пониманием, что в случае совершения этого действия человеку нечего терять, но получено может быть многое [Greene et. al., 2004].

Как отмечали Грин и соавторы, в терминах психологических процессов, связанных с их порождением, суждения по поводу безличных моральных дилемм больше напоминали суждения по поводу неморальных дилемм, чем личностных моральных дилемм, что проявилось на уровне выявленных паттернов активации коры головного мозга. Также, согласно полученным данным, участки головного мозга, продемонстрировавшие повышенную активацию в случае морального личностного условия, были вовлечены в эмоциональную переработку, что соответствовало выдвинутой гипотезе [Greene et al., 2001].

На основе полученных эмпирических данных Дж. Грин сформулировал *теорию двойного процесса (dual-process theory)*², призванную занять промежуточное место между рационалистскими и эмоционалистскими подходами; был выдвинут тезис о том, что в вынесении морального суждения участвуют как быстрые социально-эмоциональные реакции (направляющие личностные моральные суждения), так и «когнитивные» процессы (как антипод аффективных; задействуются в безличных суждениях). Само разде-

² Другой возможный вариант перевода – «теория двух процессов».

ление на «личностное» и «безличное» получило эволюционное толкование: по мнению Дж. Грина и Дж. Хайдта, предки человека, ведя социальный образ жизни, руководствовались преимущественно эмоциями, но современный человек также способен и к сложному абстрактному рассуждению. Как результат, суждения в отношении нарушений морали, знакомые нашим эволюционным предкам (т. е. личностные) управляются социально-эмоциональными реакциями, а моральные суждения в ответ на исключительно человеческие (безличные) моральные нарушения являются более «когнитивными» по своему характеру. Социально-эмоциональные реакции, как предполагалось, лежат в основе абсолютных запретов, центральных для деонтологии, а моральные вычисления, определяющие утилитаризм, возможны только благодаря более поздним структурам лобных долей, участвующим в абстрактном мышлении и высокоуровневом когнитивном контроле [Greene & Haidt, 2002; Greene et. al., 2004].

В целом, согласно Дж. Грину и Дж. Хайдту, моральная система не является «естественным видом» (*natural kind*): «Точно так же, как привычное понятие памяти относится к множеству разрозненных познавательных процессов (рабочая память, эпизодическая память, моторная память и т. д.), привычное понятие морального суждения, как мы считаем, означает множество более тонких и раздельных процессов, “аффективных” и “когнитивных”» [Greene & Haidt, 2002. P. 523].

Тезис Дж. Грина о самом наличии дискретных систем, совместно принимающих участие в вынесении морального суждения («мультисистемность»), оказался крайне влиятельным – однако вопрос о том, какие именно процессы в нем участвуют (познавательные vs. эмоциональные, контролируемые vs. автоматические, интуитивные vs. рациональные), до сегодняшнего дня остается в той или иной степени открытым; более того, варьируют даже предполагаемые характеристики протекания самих процессов (см. [Cushman, 2013; Greene, 2023]). Что касается разделения моральных дилемм на личностные и безличные, то, несмотря на яркость и лаконичность, формула ME HURT YOU оказалась произвольной – попытка формализовать различие между ситуациями, схожими с дилеммами вагонетки и дилеммами моста, не опиралась ни на принципы философии морали (так, Дж. Грин не учитывал доктрину деяния и допущения и доктрину двойного эффекта), ни на нейрофизиологические данные (о чем, в частности, говорит последующее возникновение подклассов простых и сложных личностных дилемм); недостаточно обоснованным оказалось и утверждение о необходимости сочетания всех трех компонентов формулы для того, чтобы дилемма вызвала сильный эмоциональный отклик. Кроме того, критике была подвергнута методология оригинального исследования: в ходе репликаций было показано, что ряд дилемм, использованных Дж. Грином и соавторами, не может считаться настоящими дилеммами, а также отмечалось то, что предлагаемые ситуации слишком отличны друг от друга, в частности, формулировки личностных дилемм характеризовались

большей эмоциональностью, чем безличных и неморальных [Schaich Borg et al., 2006; McGuire et al., 2009]. Идея сужения определения личностных дилемм для изолирования частных факторов, ответственных за автоматические реакции, высказывалась и сторонниками теории двойного процесса [Moore et al., 2008; Moore et al., 2011]. В своем ответе на высказанные критические замечания Дж. Грин признал произвольность и предварительность разделения «личностного» и «безличного», объяснив это изначальным отсутствием знаний о релевантных свойствах дилемм, а также выдвинул тезис о независимости теории двойного процесса от идеи разграничения личностных и безличных дилемм [Greene, 2009].

Тем не менее эвристичность как теории, так и методологии Дж. Грина была достаточно высоко оценена, в силу чего идея «мультисистемности» нашла свое продолжение в ряде других моделей, некоторые из которых будут рассмотрены далее.

Вычислительная нейронаука и развитие теории двойного процесса

Как упоминалось выше, идея проявления в процессах морального суждения двух дискретных систем получала различное содержательное наполнение; очевидно, что в целом само противопоставление быстрого, автоматического и интуитивного процесса медленному, произвольному и рациональному было продиктовано «духом времени»: так, Дж. Грин признавал, что источником вдохновения для его теории являлись описанные Даниелем Канеманом Система 1 и Система 2 («быстрое» и «медленное» мышление) [Greene, 2023]; следует вспомнить и социально-интуиционистскую модель Дж. ХайдТА [Haidt, 2001], также предполагающую существование быстрых аффективно-интуитивных процессов и рациональных обоснований *post hoc*, которая стала вехой в развитии эмпирической психологии морали [Ellemers et al., 2019].

По мнению Ф. Кашмана, существует достаточно широкий консенсус относительно необходимости обращения к работающей оптике двух процессов, но их контуры, функции и взаимовлияния остаются предметом споров. Не соглашаясь с дихотомией «когнитивное vs. аффективное» (поскольку оба процесса / обе группы процессов должны включать в себя и некоторую аффективную составляющую, и познание как переработку информации) и опираясь на принципы обучения с подкреплением из вычислительной нейронауки, Ф. Кашман предлагает различие между системами на основе оценки внутреннего статуса действий (процессы, приписывающие ценность непосредственно действиям; к примеру, «Я должен сказать Джорджу правду, потому что лгать неправильно») и оценки ожидаемых последствий действий (процессы, выбирающие действия на основе исходов; к примеру, «Если я обману Джорджа, это в конечном счете причинит ему вред»). Первая система, *модельно-свободная (Model-Free)*, присваивает ценность действиям напрямую, на ос-

нове прошлого опыта; она оказывается связанной с деонтологическими суждениями, в частности, осуждением «личностного» вреда в случае дилеммы моста; на нейрофизиологическом уровне она оказывается связана с работой дофаминовой системы и базальными ганглиями. Вторая система, *модельно-зависимая (Model-Based)*, оценивает исходы действий посредством внутренней модели причинно-следственных связей; соответственно, она вовлечена в утилитаристские суждения в случае «безличных» дилемм, сходных с проблемой вагонетки; ее нейрофизиологической основой является префронтальная кора. В ряде случаев модельно-свободная и модельно-зависимая системы взаимодействуют; примером может служить влияние привычек на сложные решения. Ф. Кашман признает, что некоторые моральные нормы могут формироваться независимо от опыта подкрепления или наказания, а быть врожденными или обусловленными биологическими механизмами; также данная теоретическая модель лучше всего применима для описания автоматических моральных интуиций. Как бы то ни было, несмотря на все ограничения, исследователь предполагает, что подобная конкретизация двух систем является шагом на пути к большей точности по сравнению с традиционной дихотомией «эмоции vs. рациональность» [Cushman, 2013].

Сходная мультисистемная модель была предложена М. Крокетт. С ее точки зрения, в моральном суждении оказываются задействованными три системы:

1. *Модельно-зависимая система* порождает прогнозирующее дерево решений, репрезентирующее связи между действиями и исходами, а также их ценностью. Действия оцениваются путем поиска по дереву и определения того, какая последовательность действий с наибольшей вероятностью приведет к наилучшему варианту; сам поиск при этом оказывается ресурсозатратным.

2. *Модельно-свободная система* оценивает действия на основе их усвоенной ценности в специфических контекстах и является ретроспективной (в отличие от проспективной модельно-зависимой), а также может предлагать неоптимальные рекомендации в условиях, где традиционно полезные действия приводят к неблагоприятным результатам.

3. *Павловская система* включает в себя автоматические рефлекторные реакции приближения или удаления к приятным и аверзивным стимулам соответственно; данная система может оказывать влияние на поведение, управляемое другими системами, а также на оценку модельно-зависимой системы [Crockett, 2013].

М. Крокетт полагает, что утилитаристские и деонтологические решения проистекают из сложных взаимодействий между системами, а не связаны напрямую с одной из них. В целом, процесс принятия морального решения выглядит в рамках данной модели следующим образом: модельно-зависимая система преобразует гипотетические сценарии в структурные описания действий и исходов (дерево решений); путем поиска по этому дереву си-

стема оценивает все возможные исходы и рекомендует действия, ведущие к наилучшему. Одновременно с этим модельно-свободная система оценивает контекстуализированные действия, приписывая то или иное значение парам «состояние – действие», исходя из истории подкрепления; в случае никогда не осуществлявшихся действий оценка основывается на наблюдении или на обучении модельно-зависимой системой путем симуляции. В свою очередь, павловская система может отвечать на прогнозы аверзивных исходов модельно-зависимой системы или аверзивные значения модельно-свободной системы. По мнению М. Крокетт, каждая система «голосует» за предпочитаемое действие, и выбор является результатом их совместного влияния. Контроль, осуществляемый системой, основанной на модели, оптимален для простых решений; в ситуациях, если результирующее дерево решений слишком велико и поиск слишком ресурсозатратен, модельно-свободный и павловский механизмы обеспечивают полезные эвристики. Если консеквенциализм дает гибкую рамку для максимизации выгоды и минимизации ущерба, то деонтология позволяет избежать проблем в ситуациях, где исходы являются неопределенными [Ibid.].

Таким образом, модели Ф. Кашмана и М. Крокетт, придерживаясь общей позиции «мультисистемности», наполняют исходную идею существования двух процессов несколько иным содержанием: особо подчеркивается роль научения (в противовес «врожденным» интуициям), а линия демаркации личностных и безличных случаев во многом оказывается зависимой от прошлого опыта и «вычислимости» ситуации, с которой сталкивается человек.

Моральный «монизм»: универсальная моральная грамматика

Своеобразным антиподом рассмотренной выше идеи мультисистемности является теория универсальной моральной грамматики, предложенная Дж. Михайлом, согласно которой мораль обусловлена функционированием специфического биологически обусловленного модуля³, а в основе морального суждения лежат рационально-интуитивные вычислительные процессы [Mikhail, 2011].

Прежде всего, Дж. Михайл оспаривает представления Дж. Грина о том, что эмоциональные процессы могут лежать в основе каких бы то ни было моральных суждений; с его точки зрения, эмоции являются эпифеноменом, ответом на независимые от них моральные оценки. Разграничение личностного и безличного в таком случае является псевдопроблемой, поскольку в действительности предъявляемые моральные дилеммы неосознаваемо анализируются людьми с позиции содержащихся в них элементов преступлений и правонарушений, что обусловлено наличием имплицитного знания моральных правил и законов⁴ [Ibid.]: «...различение «личностного» и «безличного» вреда пред-

³ В понимании Дж. Фодора.

⁴ В частности, различие в ответах на дилемму вагонетки и моста объясняется доктриной двойного эффекта.

стает слишком грубым для достижения описательной адекватности. Обычные юридические справочники – хранилища накопленных в течение веков моральных проблем и вызываемых ими интуиций – полны правдоподобных контрпримеров. Так, понятия *причинения физического вреда, цели, средства и побочного результата* являются вычислительными формулами, которые прошли проверку временем» [Mikhail, 2007. Р. 143].

В целом, согласно представлениям Дж. Михаила, люди, сталкиваясь с проблемами вагонетки, бессознательно строят репрезентации (дендрограммы), обладающие эмпирически верифицируемыми структурными свойствами. Построение структурного описания ситуации моральной дилеммы, в свою очередь, включает:

- a) распознавание различных описаний действий в стимулах;
- b) размещение их в надлежащем временном порядке;
- c) разложение их на лежащие в основе причинные и семантические структуры;
- d) применение определенных моральных и логических принципов по отношению к этим основополагающим структурам для создания репрезентаций положительных и негативных последствий;
- e) вычисление интенциональной структуры данных действий и бездействий путем принятия допущения о том, что акторы стремятся к положительным последствиям и избегают негативных (при отсутствии доказательств противоположного);
- f) получение репрезентаций морально значимых действий, подобных причинению физического вреда, а также размещение их в правильном порядке в дендрограмме действий.

Хотя каждая из этих операций сама по себе является достаточно простой, общая продолжительность, сложность и абстрактный характер подобных вычислений, равно как и их быстрый, интуитивный и как минимум частично недоступный характер, поддерживают, согласно мнению Дж. Михаила, гипотезу об их зависимости от врожденных и специализированных алгоритмов [Там же].

По аналогии с универсальной грамматикой Ноама Хомского, проявляющейся в сфере освоения языка, Дж. Михаил выдвигает идею существования универсальной моральной грамматики (UMG) как некоей врожденной моральной способности, в некоторых отношениях аналогичной языковой способности. Как и в случае языка, UMG содержит три элемента:

- 1) *деонтические правила*, действующие в процессе осуществления морального суждения (по аналогии с грамматическими правилами);
- 2) *структурные описания*, по которым осуществляются вычислительные действия (по аналогии с глубинными синтаксическими структурами);
- 3) *правила преобразования*, по которым стимул конвертируется в подходящее структурное описание (по аналогии с правилами преобразования глубинных языковых структур в поверхностные).

К аргументам в поддержку данной позиции Дж. Михаил относит достаточно сложную интуитивную юриспруденцию у детей раннего возраста (различение «истинных» нарушений морали и нарушений социальных конвенций, обращение к понятиям намерения и цели, и т. п.), наличие в каждом естественном языке слов или фраз, выражающих основные деонтические понятия («обязательное», «допустимое» и «запрещенное»), универсальность запретов на убийство, изнасилование и других типов агрессии, а также наличие достаточно устойчивой сети участков мозга, вовлеченных в моральное познание [Mikhail, 2007].

Примечательно, что М. Крокетт полагает, что по принципам, описанным Дж. Михаилом, функционирует модельно-зависимая система [Crockett, 2013], однако предложенная ей теоретическая модель все же выходит за пределы построения структурных описаний, не ограничиваясь бессознательными репрезентациями ситуации моральной дилеммы.

Одной из уязвимых точек теории универсальной моральной грамматики может быть ее доменоспецифичность, т. е. предположение о том, что механизмы морального суждения работают с моральными же репрезентациями (действие, бездействие, средство, побочный результат), осуществляя вычисления – то, что было обозначено Ф. Кашманом и Л. Янг как «базовая модель». С их точки зрения, в действительности («производная модель») основу моральных суждений формируют каузальная и интенциональная атрибуция, не являющиеся специфичными для морали: «Комплексные паттерны морального суждения могут быть укоренены в сложных процессах каузальной и интенциональной атрибуции, проходящих через фильтр относительно простых моральных вычислений. Эти разграничения кодируются как эксплицитные моральные правила путем рассуждения *post-hoc*, в равной мере обычного для философов, законодателей и обычных людей» [Cushman & Young, 2011. P. 1071].

Другим моментом в теории Дж. Михаила, подвергающимся критике, является представление об «изолированности» моральной системы, проистекающее из ее модулярного характера. Если моральное познание недоступно для влияний со стороны других систем, то, к примеру, становятся необъяснимыми результаты экспериментов, где было продемонстрировано влияние отвлечения на моральные суждения (см. [Schnall et al., 2008]). Также отмечается, что на моральные суждения оказывают влияние явные убеждения [Dipoux & Jacob, 2007].

Наконец, аргументами против идеи универсальной моральной грамматики являются тезисы о том, что человеческий мозг эволюционировал, скорее, для решения социальных, нежели моральных проблем, а явные параллели между языком и моралью все же отсутствуют – к примеру, мораль не обладает такими свойствами грамматики, как обратимость [Dipoux & Jacob, 2007; Kirchmair, 2017].

Несмотря на оригинальность нативистского, модулярного и «монистического» подхода к морали, теория Дж. Михаила, по всей видимости, не выдер-

живает эмпирической проверки; тем не менее, как было отмечено ранее, отдельные ее положения (к примеру, тезис о построении структурных описаний) представляют интерес и находят свое воплощение в других моделях.

Эволюция критериев?

Как отмечалось выше, изначальные критерии разграничения личностных и безличных дилемм были сочтены неудовлетворительными как самим Дж. Грином, так и рядом других исследователей, что привело к их трансформации, а также появлению альтернативных вариантов классификаций.

Так, Ф. Кашман, Л. Янг и М. Хаузер выделили три моральных принципа, на основе которых происходило решение моральных дилемм [Cushman, Young & Hauser, 2006]:

1. *Принцип действия*: вред, причиненный действием, хуже с моральной точки зрения, чем эквивалентный вред, причиненный бездействием; данный принцип соотносится с явлением «недооценки бездействия» (*omission bias*).
2. *Принцип намерения*: вред, предполагаемый как средство достижения цели, хуже с моральной точки зрения, чем эквивалентный вред, предполагаемый как побочный эффект достижения цели; этот принцип соотносится с доктриной двойного эффекта.
3. *Принцип контакта*: использование физического контакта для причинения вреда жертве хуже с моральной точки зрения, чем причинение эквивалентного вреда жертве без использования физического контакта.

Ф. Кашман, Л. Янг и М. Хаузер отмечают, что принцип контакта недостаточно изучался психологами и философами; хотя в нем есть внешнее сходство с разделением личностных и безличных моральных дилемм, физический контакт, согласно Дж. Грину, никогда не являлся необходимым или достаточным условием для личностной моральной дилеммы. По мнению исследователей, понятия, которые задействовали в своих моделях Л. Колберг (авторитет, сотрудничество и автономия), Дж. Хайдт (страдание, взаимность, иерархия, чистота и групповая идентичность), а также Дж. Грин, «...следует объединить с более подробным анализом таких базовых понятий, как действие, намерение и контакт» [Там же. Р. 1088].

Соответственно, в упоминавшейся выше «производной модели» Ф. Кашмана и Л. Янг основополагающими критериями являются действие/бездействие, средство/побочный результат и физический контакт/отсутствие контакта, опосредуемые каузальной и интенциональной атрибуцией [Cushman & Young, 2011].

Я. Шайх Борг и соавторы также указывали на недостаточность создания или перенаправления вреда для разграничения личностных и безличных ситуаций; так, в случае дилеммы моста актер приближается к жертве. В итоге исследователями были выделены такие факторы, как *последствия* (причинение

вреда), *действие/бездействие* и *намерение*; было выдвинуто предположение, что в основе консеквенциализма, доктрины двойного эффекта и доктрины делания и допущения лежат разные эвристики и разные мозговые механизмы [Schaich Borg et al., 2006].

Сам Дж. Грин ушел от формулы ME HURT YOU к понятию *личной силы* (*personal force*). С его точки зрения, к наиболее общим факторам, обуславливающим разницу в суждениях в ситуации жертвенной моральной дилеммы, можно отнести *намерение*, или *интенциональность* (разграничение между вредом как средством для достижения большего блага и вредом как предвидимым, но «ненамеренным» побочным эффектом достижения большего блага) и *непосредственность*, *личную силу* или *личностный характер* (наличие физического контакта, локус вмешательства, наличие или отсутствие механической опосредованности и т. п.). Как полагает Грин, актер применяет личную силу, когда сила, воздействующая на другого, генерируется его собственными мышцами; применения личной силы не могут быть опосредованы механизмами, высвобождающими или производящими другой вид силы и применяющими его к другому человеку – поэтому выстрел или сбрасывание груза, как в некоторых жертвенных дилеммах, не могут считаться примером задействования личной силы. При этом, согласно полученным эмпирическим данным, вредоносные действия оценивались как менее морально приемлемые, когда к жертве применялась личная сила; кроме того, факторы личной силы и намерения оказываются взаимозависимыми, из чего Дж. Грин делает вывод о существовании системы морального суждения, оперирующей интегрированными репрезентациями целей и личной силы (*цель-в-пределах-достигаемости-личной-силы*) – этот механизм является видом воплощенного познания. По мнению Грина, связь чувства моральной неправильности действия с базовыми двигательными функциями, равно как и фактора намерения с чувствительностью к личной силе, опровергает идеи универсальной моральной грамматики, поскольку в последней остается слишком «мало места» воплощенным репрезентациям [Greene et al., 2009].

Сходным образом, с точки зрения А. Мура и соавторов, к факторам, обуславливающим принятие морального решения, можно отнести *прямой физический контакт* (аналогичный упомянутой выше личной силе), *личный интерес* (вовлеченность или невовлеченность респондента в качестве участника гипотетической ситуации) и *интенциональность* [Moore et. al., 2011].

Й. Людвиг и соавторы также рассматривают в качестве двух наиболее важных переменных концептуализации дилемм *инструментальность вреда* (аналог намерения; вопрос о том, является ли причиняемый вред средством или побочным следствием попытки предотвратить больший вред) и *личную силу* (как роль активности актора в причинении вреда другому человеку в попытках предотвратить больший вред). При этом, как отмечают исследователи, ментальные процессы, ответственные за эффекты инструментальности и личностной силы, оказываются не до конца понятыми – в частности, неясно,

как они соотносятся с избеганием вреда или максимизацией общего блага (деонтологические и утилитаристские варианты поведения). В целом, Й. Людвиг и соавторы противопоставляют «традиционной» модели двойного процесса более «мягкие» варианты, в которых отвергается представление об исключительных связях между аффектом и деонтологией, а также когнитивными процессами и утилитаризмом, равно как и представление о том, что аффективные процессы идут первыми, и где сохраняется предположение о том, что деонтологические суждения преимущественно основаны на аверзивных чувствах по поводу причиняемого вреда, в то время как утилитаристские – на сознательных рассуждениях о максимизации выгоды [Ludwig et al., 2020].

Достаточно интересной представляется попытка классификации дилемм, предложенная Л. Лотто и соавторами: так, используя дилеммы, аналогичные дилеммам моста и вагонетки, исследователи заменяют дихотомию личностных и безличных на *инструментальные* и *случайные*, в соответствии с доктриной двойного эффекта. Они отмечают, что разграничение личностных и безличных дилемм основано на слишком обширных критериях, применимых к большому количеству сценариев, и более того, критерии ME HURT YOU касаются лишь личностных дилемм, оставляя класс безличных крайне расплывчатым в своих границах. Вторым основанием классификации является *вовлеченность* (по аналогии с личным интересом) [Lotto et. al, 2014].

Дж. Кристенсен и Э. Гомила, исчерпывающим образом анализируя особенности экспериментального дизайна в исследованиях с применением моральных дилемм, выделяют три категории параметров: формулировка дилеммы (каким образом представлен сценарий, какой требуется ответ, влияние стиля и формулировок), характеристики участника и концептуализация дилеммы (релевантные с моральной точки зрения элементы, которые характеризуют ситуацию как понятную участником); именно к разделу концептуализации относятся факторы, которые обуславливают разницу в моральных суждениях. Авторы относят к данной категории *интенциональность* (намеренный вред vs. случайный вред, действие vs. бездействие), *тип морального нарушения*, *прямолинейность причиняемого вреда* (личностное vs. безличное, личное расстояние, личная сила), *компромисс*, *нормальность вреда*, *определенность событий* [Christensen & Gomila, 2012]. Соответственно, предложенная ими классификация дилемм основывается на таких концептуальных факторах, как *личная сила* (роль актора в причинении вреда), *выгодоприобретатель*, *избегаемость* причиняемого вреда и *интенциональность* [Christensen et al., 2014].

Примечательно, что в рассмотренных критериях концептуализации не фигурируют такие переменные, как соотношение числа спасенных и погибших, а также вероятность того или иного исхода, которые способны оказывать значительное влияние на выносимое суждение, что соответствует принципам *пороговой деонтологии* – позиции, согласно которой категорические запреты могут быть нарушены, если последствия превышают определенный

порог, долгое время игнорируемой в психологических исследованиях морального познания с применением моральных дилемм [Ryazanov et al., 2023].

Таким образом, временные критерии, предложенные Дж. Грином как отражающие некое «естественное» и «интуитивно понятное» разграничение, подверглись пересмотру, в силу чего прежняя дихотомия уступила место более разнородной картине. Произошло обращение к нормативной этике, в силу чего в качестве критериев классификации дилемм стали задействоваться параметры, апеллирующие к доктрине двойного эффекта (интенциональность, инструментальность), а также доктрине делания и допущения (действие или бездействие). Предложенное понятие личной силы все еще представляется достаточно неопределенным, так как в различных интерпретациях подразумевает наличие контакта, расстояние, активность или даже наличие воплощенного компонента. Вероятно, что и такой вариант выделения критериев окажется не окончательным и в дальнейшем уступит место более частным параметрам; возможен и переход от дискретной к более континуальной картине, а также включение принципов пороговой деонтологии.

Заключение

История разделения моральных дилемм на личностные и безличные во многом примечательна: философские мысленные эксперименты, не предназначенные для проведения эмпирических изысканий, стали тем не менее частью ведущей экспериментальной парадигмы на стыке нейрофизиологических и психологических исследований, постепенно продвигаясь от набора невалидизированных и нестандартизированных ситуаций к состоянию адекватного инструмента.

Идея «мультисистемности», множественности морального познания оказалась продуктивной, и несмотря на неоднократный пересмотр и уточнение характера его процессов (аффективное и когнитивное, автоматическое и сознательное, интуитивное и рациональное или модельно-зависимое и модельно-свободное), остается актуальной и эвристичной. В свою очередь, альтернативные «монистические» подходы, как показывает опыт универсальной моральной грамматики, все же обнаруживают себя не столь состоятельными в контексте объяснения природы морального суждения и обладают недостаточной поддержкой со стороны эмпирических данных, хотя отдельные их положения могут быть непротиворечиво инкорпорированы в «мультисистемные» подходы.

Изначальное спекулятивное и «интуитивное» разделение моральных дилемм на основе субъективных критериев, осуществленное в рамках теории двойного процесса, выполнило свою функцию, поспособствовав привлечению внимания к проблематике механизмов, лежащих в основе морального познания, и постепенно сменяется иным подходом, задействующим более глубокое теоретическое осмысление, а также предполагающим междисциплинарный

синтез на стыке экспериментальной психологии, нейрофизиологии и философии морали – остается лишь выразить осторожную надежду на сохранение тренда на поступательное развитие.

Список литературы

Фёдоров А. А., Злобина М. В. Игра в мораль: связана ли оценка инструментальной пригодности моральных дилемм с их решением // *Reflexio*. 2023. Т. 16. № 1. С. 5–28. <https://doi.org/10.25205/2658-4506-2023-16-1-5-28>

Bauman C. W., McGraw A. P., Bartels D. M., Warren C. Revisiting external validity: Concerns about trolley problems and other sacrificial dilemmas in moral psychology // *Social and Personality Psychology Compass*. 2014. Vol. 8 (9). P. 536–554. <https://doi.org/10.1111/spc3.12131>

Christensen J. F., & Gomila A. Moral dilemmas in cognitive neuroscience of moral decision-making: A principled review // *Neuroscience and Biobehavioral Reviews*. 2012. Vol. 36 (4). P. 1249–1264. <https://doi.org/10.1016/j.neubio-rev.2012.02.008>

Christensen J. F., Flexas A., Calabrese M., Gut N. K., & Gomila A. Moral judgment reloaded: A moral dilemma validation study // *Frontiers in Psychology*. 2014. Vol. 5. Article 607. <https://doi.org/10.3389/fpsyg.2014.00607>

Crockett M. J. Models of morality // *Trends in cognitive sciences*. 2013. Vol. 17 (8). P. 363–366. <https://doi.org/10.1016/j.tics.2013.06.005>

Cushman F. Action, outcome, and value: a dual-system framework for morality // *Personality and social psychology review : an official journal of the Society for Personality and Social Psychology, Inc.* 2013. Vol. 17 (3). P. 273–292. <https://doi.org/10.1177/1088868313495594>

Cushman F., & Young L. Patterns of moral judgment derive from nonmoral psychological representations // *Cognitive science*. 2011. Vol. 35 (6). P. 1052–1075. <https://doi.org/10.1111/j.1551-6709.2010.01167.x>

Cushman F., Young L., & Hauser M. The role of conscious reasoning and intuition in moral judgment: testing three principles of harm // *Psychological science*. 2006. Vol. 17 (12). P. 1082–1089. <https://doi.org/10.1111/j.1467-9280.2006.01834.x>

Dupoux E., & Jacob P. Universal moral grammar: a critical appraisal // *Trends in cognitive sciences*. 2007. Vol. 11 (9). P. 373–378. <https://doi.org/10.1016/j.tics.2007.07.001>

Ellemers N., van der Toorn J., Paunov Y., & van Leeuwen T. The Psychology of Morality: A Review and Analysis of Empirical Studies Published From 1940 Through 2017 // *Personality and social psychology review : an official journal of the Society for Personality and Social Psychology, Inc.* 2019. Vol. 23 (4). P. 332–366. <https://doi.org/10.1177/1088868318811759>

Foot P. The Problem of Abortion and the Doctrine of Double Effect // *Oxford Review*. 1967. Vol. 5. P. 5–15.

Greene J. D. Dual-process moral judgment beyond fast and slow // *The Behavioral and brain sciences*. 2023. Vol. 46. e123. <https://doi.org/10.1017/S0140525X22003193>

Greene J. D. Dual-process morality and the personal/impersonal distinction: A reply to McGuire, Langdon, Coltheart, and Mackenzie // *Journal of Experimental Social Psychology*. 2009. Vol. 45 (3). P. 581–584. <https://doi.org/10.1016/j.jesp.2009.01.003>

Greene J. D., Cushman F. A., Stewart L. E., Lowenberg K., Nystrom L. E., & Cohen J. D. Pushing moral buttons: the interaction between personal force and intention in moral judgment // *Cognition*. 2009. Vol. 111 (3). P. 364–371. <https://doi.org/10.1016/j.cognition.2009.02.001>

Greene J. D., Nystrom L. E., Engell A. D., Darley J. M., & Cohen J. D. The neural bases of cognitive conflict and control in moral judgment // *Neuron*. 2004. Vol. 44 (2). P. 389–400. <https://doi.org/10.1016/j.neuron.2004.09.027>

Greene J. D., Sommerville R. B., Nystrom L. E., Darley J. M., & Cohen J. D. An fMRI investigation of emotional engagement in moral judgment // *Science* (New York, N.Y.). 2001. Vol. 293 (5537). P. 2105–2108. <https://doi.org/10.1126/science.1062872>

Haidt J. The emotional dog and its rational tail: A social intuitionist approach to moral judgment // *Psychological Review*. 2001. Vol. 108 (4). P. 814–834. <https://doi.org/10.1037/0033-295X.108.4.814>

Kirchmair L. Morality between nativism and behaviorism: (Innate) intersubjectivity as a response to John Mikhail's "universal moral grammar" // *Journal of Theoretical and Philosophical Psychology*. 2017. Vol. 37 (4). P. 230–260. <https://doi.org/10.1037/teo0000067>

Lotto L., Manfrinati A., & Sarlo M. A new set of moral dilemmas: Norms for moral acceptability, decision times, and emotional salience // *Journal of Behavioral Decision Making*. 2013. Vol. 27 (1). P. 57–65. <https://doi.org/10.1002/bdm.1782>

Ludwig J., Reizenzein R., & Hiemisch A. Effects of instrumentality and personal force on deontological and utilitarian inclinations in harm-related moral dilemmas // *Frontiers in Psychology*. 2020. Vol. 11. <https://doi.org/10.3389/fpsyg.2020.01222>

McGuire J., Langdon R., Coltheart M., & Mackenzie C. A reanalysis of the personal/impersonal distinction in moral psychology research // *Journal of Experimental Social Psychology*. 2009. Vol. 45 (3). P. 577–580. <https://doi.org/10.1016/j.jesp.2009.01.002>

Mikhail J. Emotion, Neuroscience, and Law: A Comment on Darwin and Greene // *Emotion Review*. 2011. Vol. 3 (3). P. 293–295. <https://doi.org/10.1177/1754073911406150>

Mikhail J. Universal moral grammar: Theory, evidence and the future // *Trends in Cognitive Sciences*. 2007. Vol. 11 (4). P. 143–152. <https://doi.org/10.1016/j.tics.2006.12.007>

Moore A. B., Clark B. A., & Kane M. J. Who shalt not kill? Individual differences in working memory capacity, executive control, and moral judgment // *Psy-*

chological science. 2008. Vol. 19 (6). P. 549–557. <https://doi.org/10.1111/j.1467-9280.2008.02122.x>

Moore A. B., Lee N. Y. L., Clark B. A. M., & Conway A. R. A. In defense of the personal/impersonal distinction in moral psychology research: cross-cultural validation of the dual process model of moral judgment // *Judgment and Decision Making*. 2011. Vol. 6. P. 186–195.

Ryazanov A. A., Wang S. T., Nelkin D. K., McKenzie C. R. M., & Rickless S. C. Beyond killing one to save five: Sensitivity to ratio and probability in moral judgment // *Journal of Experimental Social Psychology*. 2023. Vol. 108. P. 1–13. <https://doi.org/10.1016/j.jesp.2023.104499>

Schaich Borg J., Hynes C., Van Horn J., Grafton S., & Sinnott-Armstrong W. Consequences, action, and intention as factors in moral judgments: an fMRI investigation // *Journal of cognitive neuroscience*. 2006. Vol. 18 (5). P. 803–817. <https://doi.org/10.1162/jocn.2006.18.5.803>

Schnall S., Haidt J., Clore G. L., & Jordan A. H. Disgust as embodied moral judgment // *Personality and Social Psychology Bulletin*. 2008. Vol. 34 (8). P. 1096–1109. <https://doi.org/10.1177/0146167208317771>

Thomson J. J. Killing, Letting Die, and The Trolley Problem // *The Monist*. 1976. Vol. 59 (2). P. 204–217.

Vezirian K., Sarda E., Bègue L., Laine P.-J., & IJzerman H. “Black-and-white” thinking: Does visual contrast polarize moral judgment? Independent replications and extension of study 1 // *Journal of Experimental Social Psychology*. 2025. Vol. 117. 104712. <https://doi.org/10.1016/j.jesp.2024.104712>

References

Fedorov A. A., Zlobina M. V. (2023). Igra v moral': svyazana li otsenka instrumental'noi prigodnosti moral'nykh dilemm s ikh resheniem [The Morality Game: Is the Evaluation of the Instrumental Utility of Moral Dilemmas Related to Decision Outcomes]. *Reflexio*, 16(1), 5–28. (In Russ.) <https://doi.org/10.25205/2658-4506-2023-16-1-5-28>

Bauman, C. W., McGraw, A. P., Bartels, D. M., & Warren, C. (2014). Revisiting external validity: Concerns about trolley problems and other sacrificial dilemmas in moral psychology. *Social and Personality Psychology Compass*, 8(9), 536–554. <https://doi.org/10.1111/spc3.12131>

Christensen, J. F., & Gomila, A. (2012). Moral dilemmas in cognitive neuroscience of moral decision-making: A principled review. *Neuroscience and Biobehavioral Reviews*, 36(4), 1249–1264. <https://doi.org/10.1016/j.neubiorev.2012.02.008>

Christensen, J. F., Flexas, A., Calabrese, M., Gut, N. K., & Gomila, A. (2014). Moral judgment reloaded: A moral dilemma validation study. *Frontiers in Psychology*, 5, Article 607. <https://doi.org/10.3389/fpsyg.2014.00607>

Crockett, M. J. (2013). Models of morality. *Trends in cognitive sciences*, 17(8), 363–366. <https://doi.org/10.1016/j.tics.2013.06.005>

Cushman, F. (2013). Action, outcome, and value: a dual-system framework for morality. *Personality and social psychology review : an official journal of the Society for Personality and Social Psychology, Inc.*, 17(3), 273–292. <https://doi.org/10.1177/1088868313495594>

Cushman, F., & Young, L. (2011). Patterns of moral judgment derive from nonmoral psychological representations. *Cognitive science*, 35(6), 1052–1075. <https://doi.org/10.1111/j.1551-6709.2010.01167.x>

Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment: testing three principles of harm. *Psychological science*, 17(12), 1082–1089. <https://doi.org/10.1111/j.1467-9280.2006.01834.x>

Dupoux, E., & Jacob, P. (2007). Universal moral grammar: a critical appraisal. *Trends in cognitive sciences*, 11(9), 373–378. <https://doi.org/10.1016/j.tics.2007.07.001>

Ellemers, N., van der Toorn, J., Paunov, Y., & van Leeuwen, T. (2019). The Psychology of Morality: A Review and Analysis of Empirical Studies Published From 1940 Through 2017. *Personality and social psychology review : an official journal of the Society for Personality and Social Psychology, Inc.*, 23(4), 332–366. <https://doi.org/10.1177/1088868318811759>

Foot, P. (1967) The Problem of Abortion and the Doctrine of Double Effect. *Oxford Review*, 5, 5-15.

Greene, J. D. (2023). Dual-process moral judgment beyond fast and slow. *The Behavioral and brain sciences*, 46, e123. <https://doi.org/10.1017/S0140525X22003193>

Greene, J. D. (2009). Dual-process morality and the personal/impersonal distinction: A reply to McGuire, Langdon, Coltheart, and Mackenzie. *Journal of Experimental Social Psychology*, 45(3), 581–584. <https://doi.org/10.1016/j.jesp.2009.01.003>

Greene, J. D., Cushman, F. A., Stewart, L. E., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2009). Pushing moral buttons: the interaction between personal force and intention in moral judgment. *Cognition*, 111(3), 364–371. <https://doi.org/10.1016/j.cognition.2009.02.001>

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, 44(2), 389–400. <https://doi.org/10.1016/j.neuron.2004.09.027>

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science (New York, N.Y.)*, 293(5537), 2105–2108. <https://doi.org/10.1126/science.1062872>

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108(4), 814–834. <https://doi.org/10.1037/0033-295X.108.4.814>

Kirchmair, L. (2017). Morality between nativism and behaviorism: (Innate) intersubjectivity as a response to John Mikhail's "universal moral grammar". *Journal of Theoretical and Philosophical Psychology*, 37(4), 230–260. <https://doi.org/10.1037/teo0000067>

Lotto, L., Manfrinati, A., & Sarlo, M. (2013). A new set of moral dilemmas: Norms for moral acceptability, decision times, and emotional salience. *Journal of Behavioral Decision Making*, 27(1), 57–65. <https://doi.org/10.1002/bdm.1782>

Ludwig, J., Reisenzein, R., & Hiemisch, A. (2020). Effects of instrumentality and personal force on deontological and utilitarian inclinations in harm-related moral dilemmas. *Frontiers in Psychology*, 11. <https://doi.org/10.3389/fpsyg.2020.01222>

McGuire, J., Langdon, R., Coltheart, M., & Mackenzie, C. (2009). A reanalysis of the personal/impersonal distinction in moral psychology research. *Journal of Experimental Social Psychology*, 45(3), 577–580. <https://doi.org/10.1016/j.jesp.2009.01.002>

Mikhail, J. (2007). Universal moral grammar: Theory, evidence and the future. *Trends in Cognitive Sciences*, 11(4), 143–152. <https://doi.org/10.1016/j.tics.2006.12.007>

Mikhail, J. (2011). Emotion, Neuroscience, and Law: A Comment on Darwin and Greene. *Emotion Review*, 3(3), 293–295. <https://doi.org/10.1177/1754073911406150>

Moore, A. B., Clark, B. A., & Kane, M. J. (2008). Who shalt not kill? Individual differences in working memory capacity, executive control, and moral judgment. *Psychological science*, 19(6), 549–557. <https://doi.org/10.1111/j.1467-9280.2008.02122.x>

Moore, A. B., Lee, N. Y. L., Clark, B. A. M., and Conway, A. R. A. (2011). In defense of the personal/impersonal distinction in moral psychology research: cross-cultural validation of the dual process model of moral judgment. *Judgment and Decision Making*, 6, 186–195.

Ryazanov, A. A., Wang, S. T., Nelkin, D. K., McKenzie, C. R. M., & Rickless, S. C. (2023). Beyond killing one to save five: Sensitivity to ratio and probability in moral judgment. *Journal of Experimental Social Psychology*, 108, 1–13. <https://doi.org/10.1016/j.jesp.2023.104499>

Schaich Borg, J., Hynes, C., Van Horn, J., Grafton, S., & Sinnott-Armstrong, W. (2006). Consequences, action, and intention as factors in moral judgments: an fMRI investigation. *Journal of cognitive neuroscience*, 18(5), 803–817. <https://doi.org/10.1162/jocn.2006.18.5.803>

Schnall, S., Haidt, J., Clore, G. L., & Jordan, A. H. (2008). Disgust as embodied moral judgment. *Personality and Social Psychology Bulletin*, 34(8), 1096–1109. <https://doi.org/10.1177/0146167208317771>

Thomson, J. J. (1976). Killing, Letting Die, and The Trolley Problem. *The Monist*, 59(2), 204–217.

Vezirian, K., Sarda, E., Bègue, L., Laine, P.-J., & IJzerman, H. (2025). "Black-and-white" thinking: Does visual contrast polarize moral judgment? Independent

replications and extension of study 1. *Journal of Experimental Social Psychology*, 117, 104712. <https://doi.org/10.1016/j.jesp.2024.104712>

Сведения об авторах

Шабалин Алексей Павлович, старший преподаватель кафедры психологии личности Института медицины и психологии им. Зельмана Новосибирского государственного университета, младший научный сотрудник Междисциплинарного центра прикладного анализа поведения ИМПЗ НГУ
Scopus Author ID 57206893194
AuthorID RSCI 1002509

Information about the Author

Alexey P. Shabalin, Senior Lecturer of the Section of Personality Psychology of the V. Zelman Institute for the Medicine and Psychology of Novosibirsk State University, Junior Researcher of the Center for Behavior Analysis at Novosibirsk State University
Scopus Author ID 57206893194
AuthorID RSCI 1002509

Материал поступил в редколлегию 19.05.2025
The article was submitted 19.05.2025