

DOI: 10.19181/socjour.2022.28.1.8835

С.К. НАРЬЯН¹, А.В. БЫКОВ^{1,2}

¹ Национальный исследовательский университет
«Высшая школа экономики».
101000, Москва, ул. Мясницкая, д. 11.

² Институт социологии ФНИСЦ РАН.
109544, Москва, ул. Большая Андроньевская, д. 5, стр. 1.

ПРОБЛЕМА МОРАЛЬНОЙ АГЕНТНОСТИ АКТОРОВ: ПЕРСПЕКТИВЫ СОЦИОЛОГИЧЕСКОГО ПОДХОДА В КОНТЕКСТЕ ТЕОРИИ «МОРАЛЬНОЙ ДИАДЫ»¹

Аннотация. В статье в контексте развития новой социологии морали рассматривается область изучения моральной агентности и моральной пациентности, понимаемых как приписываемая способность актора выступать субъектом или объектом морально-релевантных действий. Анализируется концепция «моральной диады» как наиболее популярный общий подход, объясняющий различные аспекты приписывания моральной агентности и пациентности через атрибуцию сознания. Авторы эксплицируют четыре ключевых положения концепции «моральной диады»: причинение вреда как обязательное условие моральной релевантности действия, неразрывная связь приписывания сознания и моральной агентности/пациентности, феномен морального типажирования, а также принцип обязательного дополнения моральной схемы. На основе критики этих базовых положений обсуждаются перспективы дальнейших исследований в области изучения данных моральных качеств акторов различного типа. В этой связи авторы определяют и описывают три ключевые проблемы концепции «моральной диады», разработка которых требует привлечения социологической перспективы: объяснение морального статуса акторов в случае просоциальных действий, а также морального статуса коллективных (группы, корпорации, государства, институты) и нечеловеческих (роботы, компьютерные алгоритмы) акторов. Делается вывод о том, что фокус на изучении указанных проблем в контексте реальных социальных взаимодействий может стать одним из ключевых исследовательских направлений в рамках новой социологии морали.

Ключевые слова: мораль; теория «моральной диады»; моральная агентность; моральная пациентность; социология морали.

¹ Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 20-311-90038.

Для цитирования: Нарьян С.К., Быков А.В. Проблема моральной агентности акторов: перспективы социологического подхода в контексте теории «моральной диады» // Социологический журнал. 2022. Том 28. № 1. С. 8–23. DOI: 10.19181/socjour.2022.28.1.8835

Введение

Мораль традиционно служит предметом пристального интереса для самых разных наук о социальном поведении человека. В последние годы наряду с популярностью психологических исследований моральных оценок [23] наблюдается заметное возрождение социологического интереса к изучению моральных феноменов [3; 39]. При этом исследователи часто указывают на продуктивность обращения к концептуальным и методологическим ресурсам психологии морали с точки зрения развития нового социологического подхода (напр.: [16; 52]), однако все еще заметен недостаток обсуждения конкретных возможностей интеграции двух дисциплин применительно к специфическим предметным полям. На наш взгляд, одним из наиболее перспективных направлений взаимного обогащения психологии и социологии морали выступает область изучения моральной агентности и пациентности², понимаемых (в общем случае) как способность актора выступать в качестве соответственно субъекта или объекта морально-релевантных действий.

Несмотря на то что в отечественной социологической литературе можно обнаружить некоторые шаги в сторону обсуждения этой исследовательской области (напр.: выступление И.Ф. Девятко в [4]), сфокусированного обсуждения места подобной проблематики в исследовательской программе новой социологии морали [39] до сих пор не предпринималось. Между тем, как мы попытаемся показать, сфера изучения моральной агентности и пациентности обладает значительным потенциалом для развития в социологии — не в последнюю очередь по той причине, что социологическая точка зрения может серьезно расширить и уточнить те концептуальные и эмпирические результаты, которые были достигнуты в этой области в основном представителями психологии морали.

Следует оговориться, что рассматриваемая тема весьма сложна, поскольку напрямую связана с фундаментальными проблемами определения сущности и базовых принципов морали (таких, например, как проблема морального реализма, о важности которой для социологии см.: [6]), которых мы сможем коснуться лишь поверхностно. Здесь мы сфокусируемся на детальном обсуждении ключевых положений популярной теории «моральной диады» [34], которая является наиболее влиятельной

² Авторы понимают, что транскрибированные англоязычные термины *agency* и *patiency* (последний — в особенности) несколько неблагозвучны. Тем не менее нам не удалось найти русскоязычных эквивалентов, чтобы наиболее адекватно отразить стоящее за этими терминами важное концептуальное различие, о котором речь пойдет далее.

в области концептуализации восприятия моральной агентности и пациентности. На основе критического анализа этой теории мы определим три ключевые концептуальные проблемы, для решения которых может оказаться полезным привлечение социологической перспективы, а также обозначим наиболее перспективные направления развития исследований в рамках соответствующих областей.

Основные положения теории «моральной диады»

В междисциплинарной науке о морали мейнстрим в изучении моральной агентности и пациентности составляют психологические исследования, посвященные приписыванию сознания (*mind perception*) актерам различного вида [34]. С точки зрения авторов этого направления, именно приписывание сознания выступает ключевым свойством морали и восприятия действий, как подлежащих моральной оценке. Под приписыванием сознания Курт Грей и другие представители этого направления понимают восприятие акторов как имеющих два типа ментальных свойств — способность к переживанию внутреннего опыта и собственной агентности, то есть осуществлению (зло)намеренных действий [30]. Прототипическими (а)моральными действиями, таким образом, выступают случаи прямого и непосредственного причинения вреда активным «агентом» «пациенту» — убийство, изнасилование, избиение и т. п. [34, р. 109], поскольку в подобных ситуациях оба члена «моральной диады» могут быть прямо и однозначно идентифицированы наблюдателем соответственно как намеренно причинившая вред и испытывающая в результате этого страдание стороны.

Ключевым фактором «распознавания» этой диады выступает именно приписывание сознания ее членам, вариации в котором, помимо прочего, могут объяснять существующие в обществе споры, например, относительно моральной допустимости абортов, эвтаназии или употребления в пищу мяса, поскольку предполагается, что позиция в них будет в значительной мере обусловлена убеждением относительно степени развития сознания у предполагаемого морального пациента (см., напр.: [40; 49]). При этом Грей и коллеги утверждают, что само приписывание сознания может отличаться в зависимости от ситуации и в некоторых случаях быть даже более приоритетным в вынесении моральных суждений, чем объективные сведения о биологическом состоянии субъектов. В целом теория «моральной диады» может быть сведена к следующим ключевым положениям.

1. *Причинение вреда — обязательное условие моральной релевантности.* Все морально-релевантные действия фундаментально сводимы к воспринимаемому причинению вреда [34]. Даже в случае, когда от действия никто очевидным образом не страдает, если это действие считают аморальным, то предполагают, что в результате него кому-то все-таки был нанесен вред. И напротив, действие, причиняющее вред, автоматически попадает в моральный домен [47].

2. *Приписывание сознания — основа морали.* Восприятие актора как морального агента или морального пациента связано с приписыванием ему сознания, причем «моральный агент» воспринимается как интенциональный и ответственный, а «моральный пациент» — как способный испытывать эмоции, чувствовать боль или страдание [34].

3. *Моральное типажирование.* Грей и коллеги утверждают, что одним из ключевых свойств моральной категоризации выступает так называемое «моральное типажирование» (*moral typecasting*) — явление, связанное с тем, что приписывание сознания моральным акторам дихотомично и взаимоисключающе, то есть кто-то или что-то может восприниматься либо как моральный агент, либо как моральный пациент. Лишь частичное заполнение «моральной схемы» соответствует трудностям в вынесении моральных суждений [32].

4. *Дополнение моральной схемы.* Моральное типажирование универсально и характеризует даже те ситуации внутри морального домена, в которых нет явного агента или пациента. Наличие актора, которому можно приписать вину, — морального агента — способствует в восприятии людей дополнению «моральной диады» при помощи конституирования морального пациента, и, напротив, наличие морального пациента характеризуется стремлением предположить существование ответственного лица — морального агента [34].

В ходе развития концепции «моральной диады» было проведено множество эмпирических исследований, демонстрирующих неоднозначный характер приписывания агентности и пациентности акторам различного типа — от животных и младенцев до Бога и роботов. К примеру, было показано, что более строгое осуждение морального выбора, который делают беспилотные автомобили, по сравнению с тем, который совершают люди, по крайней мере, отчасти связано с приписыванием людям сознания [55]. Кроме того, было выявлено, что моральные трансгрессии человека осуждаются меньше в случае, если подчеркивается его роль жертвы, пострадавшей в прошлом, а не роль героя, совершавшего в прошлом выдающиеся поступки [33]. Эксперименты также показывают приписывание сознания мертвым, если смерть наступила, когда человек был в сознании, а не во время сна или комы [22]. Кроме того, различные смещения в приписывании сознания и агентности наблюдаются у людей с такими личностными особенностями (измеренными на субклиническом уровне), как аутизм, шизотипия и психопатия [31].

Подобные данные представляют несомненный интерес, поскольку в целом свидетельствуют о контекстуальности, а не об универсальности приписывания морально-релевантных качеств. Тем не менее, несмотря на существование корпуса ценных эмпирических исследований, представляется важным проанализировать критические оценки теории Грея и коллег, обозначив ряд сложностей, с которыми сталкивается эта концепция. Это позволит, в свою очередь, определить ключевые проблемы в области изучения моральной агентности и пациентности,

для преодоления которых, на наш взгляд, необходимо привлечение в том числе и социологической перспективы.

Критика основных положений теории «моральной диады»

Дж. Грэм и Р. Айер верно замечают [29], что используемые Греем и коллегами в их программной статье «Приписывание сознания — основа морали» (“Mind Perception Is the Essence of Morality”) и в других работах утверждения о том, что приписывание сознания — это «сущность», или «основа», морали [34], и т. д. практически не фальсифицируемы. Тем не менее многие составные части теории описаны в терминах связи между разными переменными, относящимися к моральной способности людей, в результате чего, в частности, неясно, какого рода каузальные отношения предполагаются между приписыванием сознания и моральными качествами. Кроме того, обозначенные выше четыре постулата концепции «моральной диады» также подвергаются критике, которую мы попытаемся здесь кратко суммировать.

Так, постулат о причинении вреда (в первую очередь межличностного) как основе морали, несмотря на важность этого фактора с точки зрения ряда философских и психологических концепций [18; 27], выглядит проблематичным, поскольку существует вполне обоснованная точка зрения, согласно которой этим моральный домен не ограничивается. В частности, в известной мере более социологически ориентированная теория моральных оснований предполагает, что вред — это только одно из моральных оснований — внеличных структур, влияющих на вариативные моральные интуиции [28]. Пять основных из них — это *вред/забота*, *справедливость/реципрокность*, *ингруппа/преданность*, *авторитет/уважение* и *чистота/непорочность*. Теория моральных оснований предполагает, что индивидуальные (individualizing) основания включают моральные ценности, развивавшиеся, чтобы совершенствовать межличностные отношения через продвижение равноправия и автономии, которые предотвращают нанесение вреда индивидам [10]. Связующие (binding) основания, напротив, включают моральные ценности, развивавшиеся, чтобы повышать сплоченность в больших социальных группах, поощряя послушание, верность и телесную чистоту. Данная теория и ряд связанных с ней подходов, таким образом, свидетельствуют о необходимости серьезного дополнения концепции «моральной диады» с учетом прогруппового характера значительной части моральных интуиций людей.

Постулат о непременной связи приписывания сознания и моральных суждений также нуждается в существенном прояснении. Если предположить, что с точки зрения теории «моральной диады» приписывание сознания — ключевой фактор морали, то становится очевидно, что она серьезно сужает многообразный моральный домен, содержащий не только моральные суждения, но и моральное поведение и моральные эмоции [23]. Кроме того, используемое авторами определение приписывания сознания как приписывания агентности

(способности быть субъектом морального действия) и пациентности (способности быть его объектом) [34] не учитывает, что приписывание сознания может быть связано с другими аспектами воспринимаемой ментальной способности, такими как способность нести ответственность, иметь свободу воли или осуществлять осознанный выбор [13]. Остается также неясным, какую роль приписывание сознания должно играть в случаях, когда моральными агентами и пациентами выступают группы и сообщества — к примеру, нарушение обычаев и традиций, развязывание войны, предательство национальных интересов и т. п.

Феномен морального типажирования как обязательное наличие отдельных друг от друга «агента» и «пациента», в свою очередь, исключает множество ситуаций, в которых люди выносят моральные суждения относительно самих себя [5; 8]. В частности, существуют данные о том, что такие действия, как самоповреждение [44] или мастурбация [57], могут восприниматься как аморальные; при этом сложно понять, как это соотносится с феноменом морального типажирования. Другая важная критика этой идеи состоит в том, что результаты исследований ее сторонников можно объяснить не обязательным существованием стремления к моральному типажированию, а специфическими характеристиками использованного исследователями стимульного материала [8].

Наконец, принцип дополнения моральной схемы, то есть необходимость обязательного конституирования второго члена диады при наличии первого, также вызывает определенную критику, основная суть которой заключается в том, что это теоретическое допущение требует всесторонней эмпирической проверки. Еще одна важная проблема этого принципа состоит в том, что он тесно связан с другими идеями теории «моральной диады», каждая из которых применима далеко не ко всем морально релевантным ситуациям. Так, если расширить взгляд на моральные трансгрессии и учитывать не только те, которые влекут непосредственное причинение вреда, неочевидным становится и дополнение моральной схемы. Например, неясно, как будет дополнена схема, если человек был обманут, но «белая ложь» не повлекла страдания, — ситуация, относящаяся к моральному домену и воспринимаемая как морально неоднозначная [7]. Кроме того, Дж. Хайдт описывает феномен «морального ошарашивания» (*moral dumbfounding*), суть которого заключается в том, что люди продолжают считать аморальными некоторые действия, в которых очевидный «пациент», то есть сторона, которой был причинен вред, отсутствует [35], что тоже позволяет усомниться в необходимости обязательного наличия обоих членов диады как ключевой характеристики морального домена.

Тем не менее, несмотря на эту и другую критику, концепция «моральной диады» и стоящая за ней проблематика приписывания моральных «метакачеств», на наш взгляд, имеют серьезный потенциал для дальнейшего развития и расширения с учетом социологической перспективы. В следующем разделе мы обозначим три ключевые про-

блемные области для потенциальной социологической рецепции и модификации теории «моральной диады», разработка которых позволит социологии морали содержательно обогатить базовые психологические модели приписывания моральной агентности и пациентности.

Ключевые проблемы концепции «моральной диады» и социология морали

Перед тем как представить наше видение ключевых концептуальных вопросов, рассмотрение которых открывает возможности для интеграции теории «моральной диады» и социологии морали, следует сделать важное пояснение относительно методологии. Проблема соотношения психологических и социологических методов и методик в применении к исследованию моральных явлений активно осмысливается в новой социологии морали [16; 39]. В частности, С. Вейзи и Э. Майлз убедительно показывают [22], что методические инструменты, разработанные в рамках психологии морали, вполне могут использоваться для изучения интересующих социологов явлений (например, «личностной моральной культуры»), в то время как психологические концептуальные модели морали могут тестироваться традиционными социологическими методами, включая контент-анализ «больших» текстовых данных и полуструктурированные интервью. Отечественные социологи, в свою очередь, вполне успешно используют методологию когнитивной психологии (включая применение гипотетических сценариев в качестве стимульного материала), адаптируя их под собственные содержательные задачи (см., напр.: [2; 20; 21]). В целом же современная «наука о морали» характеризуется большой степенью междисциплинарности, что, помимо прочего, предполагает скорее взаимное методическое обогащение, нежели выстраивание дисциплинарных границ через педалирование специфических методологических стратегий.

Таким образом, здесь мы не склонны придавать чрезмерное значение методологическим вопросам, поэтому сосредоточимся в первую очередь на концептуальных проблемах, связанных с потенциальной социологической рецепцией, уточнением и расширением теории «моральной диады». Мы предполагаем, что для решения этих проблем социологи могут использовать весь доступный арсенал методов социальных и поведенческих наук.

Проблема 1. Объяснение моральной оценки просоциальных действий

Как уже было сказано, ключевой характеристикой концепции «моральной диады» является представление о намеренном нанесении вреда другому как прототипическом действии, вызывающем моральную оценку. На наш взгляд, этот факт вполне объясним исходя из логики развития психологии морали в последние десятилетия [23], которая в целом была чрезмерно сфокусирована на изучении моральных трансгрессий

(ср., например, знаменитую «проблему вагонетки»³) в ущерб просоциальным действиям. Вместе с тем развитие упомянутой теории моральных оснований [27] позволило показать, в частности, что действия внутри морального домена часто сопряжены с (воспринимаемым) нанесением вреда не индивиду, а группе, что отстоит довольно далеко от того представления, из которого исходит вся концепция «моральной диады».

В связи с этим необходимо отметить, что альтруистические и групповые действия в целом остаются практически неконцептуализированными в рамках теории «моральной диады». Между тем предположить возможное различие в факторах восприятия моральной агентности и пациентности⁴ для случаев вреда и помощи можно, в частности, исходя из известной асимметрии приписывания интенциональности (ключевой атрибут «агентности» по Грею и соавт.) актерам в условиях позитивных и негативных побочных эффектов действий (так называемый «эффект Ноуба» [24], впрочем, также подверженный вариации в зависимости от институционального контекста [21]). Кроме того, существуют экспериментальные данные о том, что люди высоко оценивают моральную ценность альтруистических действий как таковых [20], это также может быть сопряжено со спецификой распознавания позитивной роли моральных агентов в подобных ситуациях.

Для социологии морали в целом исторически характерен фокус на феноменах альтруизма и солидарности как ключевых составляющих моральных качеств человека [15]. Проявление моральной агентности через просоциальные действия по отношению к другим индивидам или группам как целостным образованиям (и, напротив, действия групп по отношению к индивидам), таким образом, нуждается в специальном изучении, которое с большой вероятностью повлечет за собой модификацию имеющихся представлений о восприятии моральных качеств. В этом смысле одним из наиболее социологически релевантных примеров потенциального смещения моральной агентности просоциальных действий является активность благотворительных организаций, которая в ряде случаев оказывается ответственной за, казалось бы, индивидуальное решение о донировании крови или органов [36]. Отдельный важный с точки зрения теории альтруизма вопрос: как восприятие/конституирование пациентности связано с мотивацией к оказанию жертве альтруистической помощи и от каких

³ Классический мысленный эксперимент в этике [14]. В эксперименте предлагается представить, что тяжелая неуправляемая вагонетка быстро едет по рельсам в направлении стрелки, которую можно переключить. На одном из возможных путей вагонетки лежат пять человек, а на другом пути привязан и не может убежать один человек.

⁴ Если данный термин останется применим для характеристики реципиента альтруистического действия, который, очевидно, едва ли выступает пострадавшей стороной.

социально-контекстуальных факторов это может зависеть? Внимание к подобного рода социальным процессам позволит социологически насытить общие психологические модели и более адекватно отразить сложность восприятия моральных свойств участников различных социальных взаимодействий.

Проблема 2. Моральные качества коллективных акторов

Второй важной проблемой концепции «моральной диады», которая связана с предыдущей, является спорный моральный статус коллективных акторов и, как следствие, отсутствие ясности относительно связи приписывания агентности, пациентности и сознания группам и сообществам. Грей и соавторы явным образом исходят из модели межиндивидуального морального действия, которая в этом смысле оказывается изначально ограниченной. В то же время дюркгеймианская традиция концептуализации морали в социологии, напротив, предполагает понимание морального действия как акта, направленного на благо группы, а не конкретного индивида [15], что позволяет концептуально расширить «моральную диаду» за счет включения в нее коллективных и институциональных акторов различного уровня.

Нужно отметить, что до сих пор предпринимались некоторые теоретические усилия к изучению коллективных субъектов со спорно воспринимаемым моральным статусом, таких как корпорации [54] или государства [26]. Эти усилия связаны с рассмотрением двух отдельных предполагаемых в теории «моральной диады» аспектов: моральной агентности и моральной пациентности. Так, большое количество литературы посвящено теоретическому обсуждению того, могут ли корпорации нести моральную ответственность, отдельную от моральной ответственности их членов [19; 37]. В то же время ряд работ связаны с обсуждением вопроса возможности существования прав у корпораций и отличия этих прав от прав людей, в том числе и в юридическом контексте [12]; также существуют дискуссии о моральной ответственности наций [43]. Немногочисленные эксперименты, направленные на изучение приписываемой таким акторам моральной агентности [45], фокусируются на приписывании сознания, не разделяя эти два процесса, что не позволяет напрямую оценить применимость концепции «моральной диады» к коллективным субъектам.

Вполне возможно предположить, что восприятие агентности и пациентности коллективных акторов может оказаться не связанным с приписыванием им ментальных качеств, по крайней мере в ряде случаев. В частности, известное различие моральных представлений либералов и консерваторов [28] вполне может объясняться тем, что первые не распознают моральных пациентов в коллективных акторах (таких, например, как нация с ее традициями или церковь с ее догматами) и, как следствие, не видят морального содержания в действиях, наносящих, с точки зрения вторых, вред традиционным устоям.

На наш взгляд, эта гипотеза заслуживает комплексной эмпирической проверки в будущих исследованиях.

Кроме того, еще одна важная идея, которую социология может привнести в данную проблематику, — рассмотрение приписывания агентности и пациентности на разных уровнях социального взаимодействия. Приписывание моральной ответственности и сопряженных с ней моральных «метакачеств» может значительно отличаться в случае, когда в действии проявляются те или иные коллективные или институциональные акторы. Например, полицейский, применивший силу к участникам митинга, может рассматриваться не как самостоятельный (а)моральный агент, а как проявление (а)моральных качеств института полиции или целого государства. Другой пример: когда на войне солдат армии страны N стреляет в солдата армии страны M, его действие ассоциировано с самим солдатом, его непосредственным командиром, армией страны N, руководством страны N, самой страной N. Вопросы о том, в какой степени все эти акторы проявляют в глазах людей качества моральной агентности и пациентности, вполне могут составить одну из основ социологической исследовательской программы в данной предметной области.

Проблема 3. Моральные качества нечеловеческих акторов

Наконец, третьей важной для социологии проблемой концепции «моральной диады» является неясный моральный статус нечеловеческих акторов, в первую очередь различных форм искусственного интеллекта. Несмотря на то что Грей и соавторы предпринимали попытки выяснить, насколько люди склонны к приписыванию моральной агентности и пациентности роботам [30], они оставались довольно абстрактными и не были помещены в контекст конкретных ситуаций взаимодействия человека с искусственным интеллектом.

Между тем сегодня многие решения, сопряженные с моральной оценкой действий и их последствий, принимаются с использованием нечеловеческих акторов, таких как роботы [48] и компьютерные алгоритмы [42]. Уже существуют дроны, которые могут автономно проводить разведку и бомбардировку вражеских комбатантов с вероятным причинением человечески жертв [17], что порождает соответствующие вопросы об их моральном статусе. Ведется дискуссия об использовании алгоритмов в принятии решений о приговорах в уголовном процессе [56]. В контексте пандемии COVID-19 искусственный интеллект используется в разработке лекарственных средств и вакцин, а также для снижения нагрузки на медицинский персонал [53]. В то же время беспилотные автомобили перестают быть объектами абстрактного теоретизирования — они широко тестируются на дорогах общего пользования, в том числе в России; все более актуальной темой в социальных науках становится этика их создания, программирования, использования и регулирования [11; 41].

В связи с этим представляется, что проблема восприятия моральных качеств искусственного интеллекта в ближайшее время будет

приобретать все большую актуальность. Можно предположить, что приписывание моральной агентности роботам будет определяться в том числе таким уникальным для этих акторов фактором, как антропоморфность [25], или автономность, понимаемая как уровень независимости от программы [9]. Кроме того, существуют данные о том, что восприятие моральной пациентности робота связано с оказанием ему помощи [50]; это говорит о том, что важно учитывать контекст взаимодействия человека и искусственного интеллекта, а не только приписывать последнему сознание в той или иной форме [38]. Вероятно также, что немалую роль в оценке моральных качеств алгоритмов будут играть представления о причинности того или иного действия, а также о том, насколько эти причины могут быть эксплицированы и объяснены, что составляет серьезную проблему для технологий машинного обучения [51]. Так, данные недавнего эксперимента [46] свидетельствуют о том, что люди склонны винить во вредоносных действиях искусственного интеллекта не столько самого робота, сколько команду программистов и компанию, в которой они работают, что указывает на перспективность рассмотрения проблемы моральных качеств машин в широком социальном контексте. В целом фокусирование на взаимодействиях людей и компьютерных алгоритмов в морально-релевантных ситуациях будет расширять традиционную для социологии область изучения морального сознания [1] за счет анализа качественно новых социальных процессов.

Заключение

В данной работе мы попытались показать, что проблематика изучения моральной агентности и пациентности, в наибольшей степени разработанная в когнитивной психологии, представляет собой перспективную область развития в рамках новой социологии морали. Нами обозначены ключевые проблемы концепции «моральной диады», которые, на наш взгляд, требуют социологического анализа, а также намечены возможные пути дальнейших исследований в этой области с применением социологического подхода. Мы полагаем, что феномен приписывания моральной агентности и пациентности акторам различного типа, до сих пор не получивший должного внимания в социологии, вполне может стать одним из интереснейших и плодотворных направлений исследований в рамках активно возрождающейся сегодня социологии морали.

ЛИТЕРАТУРА НА РУССКОМ ЯЗЫКЕ⁵

1. *Быков А.В.* Понятие морального сознания в социологической традиции // Социологический журнал. 2017. Том. 23. № 3. С. 26–43. DOI: 10.19181/socjour.2017.23.3.5362

⁵ Полный список литературы — в списке References.

2. *Гаврилов К.А.* Целостность и ценность сообщества в исследованиях риска // Социологический журнал. 2013. № 3. С. 92–113. DOI: 10.19181/socjour.2013.3.422
3. *Девятко И.Ф., Абрамов Р.Н., Быков А.В., Гаврилов К.А., Катерный И.В., Подвойский Д.Г., Романовский Н.В., Сапов В.В.* Нормы и мораль в социологической теории: от классических концепций к новым идеям. М.: Весь Мир, 2017. — 288 с.
4. *Девятко И.Ф., Катерный И.В., Кирилина Т.Ю., Перов В.Ю., Семенов Е.В., Соколов В.М., Черныш М.Ф.* Нравственность, мораль, этика: что происходит в теории и социальной практике? (Круглый стол) // Социологические исследования. 2021. № 3. С. 28–43. DOI: 10.31857/S013216250014275-5
5. *Настина Е.А., Девятко И.Ф.* Моральная оценка себя: разработка и апробация русскоязычного инструментария // Мониторинг общественного мнения: Экономические и социальные перемены. 2021. № 2. С. 4–27. DOI: 10.14515/monitoring.2021.2.1720

СВЕДЕНИЯ ОБ АВТОРАХ

Нарьян Светлана Кареновна — аспирант, Аспирантская школа по социологическим наукам; стажер-исследователь, департамент социологии, Национальный исследовательский университет «Высшая школа экономики». **Телефон:** +7 (495) 771-32-32, доб. 12454. **Электронная почта:** snaryan@hse.ru

Быков Андрей Вячеславович — кандидат социологических наук, доцент, кафедра анализа социальных институтов, департамент социологии, Национальный исследовательский университет «Высшая школа экономики»; научный сотрудник, Институт социологии ФНИСЦ РАН. **Телефон:** +7 (495) 771-32-32, доб. 12454. **Электронная почта:** a.bykov@hse.ru

Дата поступления: 02.12.2021.

SOTSIOLOGICHESKIY ZHURNAL = SOCIOLOGICAL JOURNAL. 2022.
VOL. 28. No. 1. P. 8–23. DOI: 10.19181/socjour.2022.28.1.8835

SVETLANA K. NARYAN¹, ANDREY V. BYKOV^{1,2}

¹ HSE University.

20, Myasnitckaya str., 101000, Moscow, Russian Federation.

² Institute of Sociology of FCTAS RAS.

5, bl. 1, Bolshaya Andronievskaya str., 109544, Moscow, Russian Federation.

THE PROBLEM OF MORAL AGENCY: PROSPECTS OF THE SOCIOLOGICAL APPROACH IN THE CONTEXT OF THE “MORAL DYAD” THEORY

Abstract. In the context of developing a new sociology of morality, the article examines such phenomena as moral agency and moral patiency, understood as an actor’s attributed ability to be seen as a subject or object of morally relevant actions. The authors analyze the concept of the “moral dyad” as the most popular general approach that addresses

various aspects of attributing moral agency and patiency through mind ascription. We also examine the four key principles of the moral dyad concept: causing harm as a prerequisite for the moral relevance of the action, the inseparable link between mind ascription and moral agency/patiency, the phenomenon of moral typecasting, and the principle of the moral scheme's completion being a necessity. Based on the criticisms of these basic principles, we consider the prospects for further research in the field of studying these moral qualities of actors of various types. For this purpose, we identify and discuss three key problems of the concept of the “moral dyad”, which, as we argue, should be addressed from a sociological perspective: explaining the moral status of actors in the case of pro-social actions, as well as the moral status of collective (groups, corporations, states, institutions) and non-human (robots, computer algorithms) actors. In conclusion we suggest that focusing on these problems in the context of actual social interactions can become one of the key research areas within the framework of the new sociology of morality.

Keywords: morality; “moral dyad” theory; moral agency; moral patiency; sociology of morality.

For citation: Naryan, S.K., Bykov, A.V. The Problem of Moral Agency: Prospects of the Sociological Approach in the Context of the “Moral Dyad” Theory. *Sotsiologicheskii Zhurnal = Sociological Journal*. 2022. Vol. 28. No. 1. P. 8–23. DOI: 10.19181/socjour.2022.28.1.8835

Acknowledgements. The reported study was funded by RFBR, project number 20-311-90038.

REFERENCES

1. Bykov A.V. The concept of moral conscience in the sociological tradition. *Sotsiologicheskii Zhurnal = Sociological Journal*. 2017. Vol. 23. No. 3. P. 26–43. DOI: 10.19181/socjour.2017.23.3.5362 (In Russ.)
2. Gavrilov K.A. Entitativity and value of the community in risk research. *Sotsiologicheskii Zhurnal = Sociological Journal*. 2013. No. 3. P. 92–113. DOI: 10.19181/socjour.2013.3.422 (In Russ.)
3. Devyatko I.F., Abramov R.N., Bykov A.V., Gavrilov K.A., Katernyi I.V., Podvoiskii D.G., Romanovskii N.V., Sapov V.V. *Normy i moral' v sotsiologicheskoi teorii: ot klassicheskikh kontseptsii k novym ideyam*. [Norms and morality in sociological theory: from classical conceptions to new ideas.] Moscow: Ves' Mir publ., 2017. 288 p. (In Russ.)
4. Devyatko I.F., Katernyi I.V., Kirilina T.Yu., Perov V.Yu., Semenov E.V., Sokolov V.M., Chernysh M.F. *Nravstvennost', moral', etika: chto proiskhodit v teorii i sotsial'noi praktike? (Kruglyi stol)*. [Morality, Ethics: What Happens in Theory and Social Practice? (Round Table Discussion).] *Sotsiologicheskie issledovaniya*. 2021. No. 3. P. 28–43. DOI: 10.31857/S013216250014275-5 (In Russ.)
5. Nastina E.A., Devyatko I.F. Moral Self-Evaluation: Developing and Validating Methodology for Russian-Speaking Populations. *Monitoring obshchestvennogo mneniya: Ekonomicheskie i sotsial'nye peremeny*. 2021. No. 2. P. 4–27. DOI: 10.14515/monitoring.2021.2.1720 (In Russ.)
6. Abend G. Two main problems in the sociology of morality. *Theory and Society*. 2008. No. 37 (2). P. 87–125. DOI: 10.1007/s11186-007-9044-y
7. Argo J., Shiv B. Are white lies as innocuous as we think? *Journal of Consumer Research*. 2010. Vol. 38. No. 6. P. 1093–1102. DOI: 10.1086/661640
8. Arico A.J. Breaking out of moral typecasting. *Review of Philosophy and Psychology*. 2012. Vol. 3. No. 3. P. 425–438. DOI: 10.1007/s13164-012-0108-1

9. Banks J.A. Perceived moral agency scale: Development and validation of a metric for humans and social machines. *Computers in Human Behavior*. 2019. Vol. 90. P. 363–371. DOI: 10.1016/j.chb.2018.08.028
10. Bettache K., Hamamura T., Amrani Idrissi J., Amenyogbo R., Chiu C. Monitoring moral virtue: When the moral transgressions of in-group members are judged more severely. *Journal of Cross-Cultural Psychology*. 2019. Vol. 50. No. 2. P. 268–284. DOI: 10.1177/0022022118814687
11. Bigman Y.E., Gray K. Life and death decisions of autonomous vehicles. *Nature*. 2020. Vol. 579. No. 7797. P. E1–E2. DOI: 10.1038/s41586-020-1987-4
12. Bishop J.D. The limits of corporate human rights obligations and the rights of for-profit corporations. *Business Ethics Quarterly*. 2012. Vol. 22. No. 1. P. 119–144. DOI: 10.5840/beq20122217
13. Björnsson G., Shepherd J. Determinism and attributions of consciousness. *Philosophical Psychology*. 2020. Vol. 33. No. 4. P. 549–568. DOI: 10.1080/09515089.2020.1743256
14. Bruers S., Braeckman J. A Review and Systematization of the Trolley Problem. *Philosophia*. 2014. Vol. 42. P. 251–269. DOI: 10.1007/s11406-013-9507-5
15. Bykov A. Altruism: new perspectives of research on a classical theme in sociology of morality. *Current Sociology*. 2017. Vol. 65. No. 6. P. 797–813. DOI: 10.1177/0011392116657861
16. Bykov A. Rediscovering the moral: The “old” and “new” sociology of morality in the context of the behavioural sciences. *Sociology*. 2019. Vol. 53. No. 1. P. 192–207. DOI: 10.1177/0038038518783967
17. Crootof R. The killer robots are here: Legal and policy implications. *Cardozo Law Review*. 2015. Vol. 36. No. 5. P. 1837–1916.
18. Decety J., Cowell J.M. Interpersonal harm aversion as a necessary foundation for morality: A developmental neuroscience perspective. *Development and Psychopathology*. 2018. Vol. 30. No. 1. P. 153–164. DOI: 10.1017/S0954579417000530
19. Dempsey J. Corporations and non-agential moral responsibility. *Journal of Applied Philosophy*. 2013. Vol. 30. No. 4. P. 334–350. DOI: 10.1111/japp.12029
20. Deviatko I.F., Bykov A. Weighing the moral worth of altruistic actions: a discrepancy between moral evaluations and prescriptive judgments. *Philosophical Psychology*. 2021. Vol. 22. No. 1. P. 95–121. DOI: 10.1080/09515089.2021.1950666
21. Deviatko I.F., Gavrilov K.A. Causality and blame judgments of negative side effects of actions may differ for different institutional domains. *SAGE Open*. October 2020. DOI:10.1177/2158244020970942
22. Doyle C., Gray K. How people perceive the minds of the dead: the importance of consciousness at the moment of death. *Cognition*. 2020. Vol. 202. Article 104308. DOI: 10.1016/j.cognition.2020.104308
23. 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*. 2019. Vol. 23. No. 4. P. 332–366. DOI: 10.1177/1088868318811759
24. Feltz A. The Knobe effect: A brief overview. *Journal of Mind and Behavior*. 2007. Vol. 28 (3–4). P. 265–278.
25. Ferrari F., Paladino M.P., Jetten J. Blurring human–machine distinctions: Anthropomorphic appearance in social robots as a threat to human distinctiveness. *International Journal of Social Robotics*. 2016. Vol. 8. No. 2. P. 287–302. DOI: 10.1007/s12369-016-0338-y
26. Fleming S. Moral agents and legal persons: The ethics and the law of state responsibility. *International Theory*. 2017. Vol. 9. No. 3. P. 466–489. DOI: 10.1017/S1752971917000100
27. Gert B. *Morality: Its Nature and Justification*. N.Y.: Oxford University Press, 1998. 438 p.

28. Graham J., Haidt J., Koleva S., Motyl M., Iyer R., Wojcik S.P., Ditto P.H. Moral foundations theory: The pragmatic validity of moral pluralism. *Advances in Experimental Social Psychology*. Ed. by P. Devine, A. Plant. Amsterdam: Academic Press, 2013. Vol. 47. P. 55–130. DOI: 10.1016/B978-0-12-407236-7.00002-4
29. Graham J., Iyer R. The unbearable vagueness of “essence”: Forty-four clarification questions for Gray, Young, and Waytz. *Psychological Inquiry*. 2012. Vol. 23. No. 2. P. 162–165. DOI: 10.1080/1047840X.2012.667767
30. Gray H., Gray K., Wegner D. Dimensions of mind perception. *Science*. 2007. Vol. 315. No. 5812. Accessed 25.02.2022. URL: <http://www.bio-nica.info/biblioteca/Gray2007MindPerception.pdf> DOI: 10.1126/science.1134475
31. Gray K., Jenkins A., Heberlein A., Wegner D. Distortions of mind perception in psychopathology. *Proceedings of the National Academy of Sciences*. 2011. Vol. 108. No. 2. P. 477–479. DOI: 10.1073/pnas.1015493108
32. Gray K., Wegner D. Moral typecasting: Divergent perceptions of moral agents and moral patients. *Journal of Personality and Social Psychology*. 2009. Vol. 96. No. 3. P. 505–520. DOI: 10.1037/a0013748
33. Gray K., Wegner D.M. To escape blame, don’t be a hero — be a victim. *Journal of Experimental Social Psychology*. 2011. Vol. 47. No. 2. P. 516–519. DOI: 10.1016/j.jesp.2010.12.012
34. Gray K., Young L., Waytz A. Mind perception is the essence of morality. *Psychological Inquiry*. 2012. Vol. 23. No. 2. P. 101–124. DOI: 10.1080/1047840X.2012.651387
35. Haidt J. The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*. 2001. Vol. 108. P. 814–834. DOI: 10.1037/0033-295X.108.4.814
36. Healy K. Altruism as an organizational problem: The case of organ procurement. *American Sociological Review*. 2004. Vol. 69. No. 3. P. 387–404. DOI: 10.1177/000312240406900304
37. Hess K.M. “If you tickle us...”: How corporations can be moral agents without being persons. *The Journal of Value Inquiry*. 2013. Vol. 47. No. 3. P. 319–335. DOI: 10.1007/s10790-013-9391-z
38. Himma K.E. Artificial agency, consciousness, and the criteria for moral agency: What properties must an artificial agent have to be a moral agent? *Ethics and Information Technology*. 2009. Vol. 11. No. 1. P. 19–29. DOI: 10.1007/s10676-008-9167-5
39. Hitlin S., Vaisey S. The new sociology of morality. *Annual Review of Sociology*. 2013. Vol. 39. No. 1. P. 51–68. DOI: 10.1146/annurev-soc-071312-145628
40. Loughnan S., Haslam N., Bastian B. The role of meat consumption in the denial of moral status and mind to meat animals. *Appetite*. 2010. Vol. 55. No. 1. P. 156–159. DOI: 10.1016/j.appet.2010.05.043
41. Marres N. Co-existence or displacement: do street trials of intelligent vehicles test society? *The British Journal of Sociology*. 2020. Vol. 71. No. 3. P. 537–555. DOI: 10.1111/1468-4446.12730
42. McBride N. The ethics of driverless cars. *ACM SIGCAS Computers and Society*. 2016. Vol. 45. No. 3. P. 179–184. DOI: 10.1145/2874239.2874265
43. Miller D. Holding nations responsible. *Ethics*. 2004. Vol. 114. No. 2. P. 240–268. DOI: 10.1086/379353
44. Patterson P., Whittington R., Bogg J. Measuring nurse attitudes towards deliberate self-harm: The Self-Harm Antipathy Scale (SHAS). *Journal of Psychiatric and Mental Health Nursing*. 2007. Vol. 14. No. 5. P. 438–445. DOI: 10.1111/j.1365-2850.2007.01102.x
45. Rai T.S., Diermeier D. Corporations are cyborgs: Organizations elicit anger but not sympathy when they can think but cannot feel. *Organizational Behavior and Human Decision Processes*. 2015. Vol. 126. P. 18–26. DOI: 10.1016/j.obhdp.2014.10.001

46. Sullivan Y., Fosso Wamba S. Moral Judgments in the Age of Artificial Intelligence. *Journal of Business Ethics*. 2022. DOI: 10.1007/s10551-022-05053-w
47. Schein C., Gray K. The theory of dyadic morality: Reinventing moral judgment by re-defining harm. *Personality and Social Psychology Review*. 2018. Vol. 22. No. 1. P. 32–70. DOI: 10.1177/1088868317698288
48. Sharkey A. Can we program or train robots to be good? *Ethics and Information Technology*. 2020. Vol. 22. No. 4. P. 283–295. DOI: 10.1007/s10676-017-9425-5
49. Singer P. *Animal liberation*. N.Y.: Ecco Press, 2002. 324 p.
50. Tanibe T., Hashimoto T., Karasawa K. We perceive a mind in a robot when we help it. *PLOS ONE*. 2017. Vol. 12. No. 7. Article e0180952. DOI: 10.1371/journal.pone.0180952
51. Tsamados A., Aggarwal N., Cowl J., et al. The ethics of algorithms: key problems and solutions. *Artificial Intelligence and Society*. 2021. DOI: 10.1007/s00146-021-01154-8
52. Vaisey S., Miles A. Tools from moral psychology for measuring personal moral culture. *Theory and Society*. 2014. Vol. 43. P. 311–332. DOI: 10.1007/s11186-014-9221-8
53. Vaishya R., Javaid M., Khan I.H., Haleem A. Artificial intelligence (AI) applications for COVID-19 Pandemic. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*. 2020. Vol. 14. No. 4. P. 337–339. DOI: 10.1016/j.dsx.2020.04.012
54. Werhane P.H. Corporate moral agency and the responsibility to respect human rights in the UN guiding principles: Do corporations have moral rights? *Business and Human Rights Journal*. 2016. Vol. 1. No. 1. P. 5–20. DOI: 10.1017/bhj.2015.1
55. Young A., Monroe A. Autonomous morals: Inferences of mind predict acceptance of AI behavior in sacrificial moral dilemmas. *Journal of Experimental Social Psychology*. 2019. Vol. 85. Article 103870. DOI: 10.1016/j.jesp.2019.103870
56. Završnik A. Algorithmic justice: Algorithms and big data in criminal justice settings. *European Journal of Criminology*. 2021. Vol. 18. No. 5. P. 623–642. DOI: 10.1177/1477370819876762
57. Zhong C.-B., Strejcek B., Sivanathan N. A clean self can retrnder harsh moral judgment. *Journal of Experimental Social Psychology*. 2010. Vol. 46. No. 5. P. 859–862. DOI: 10.1016/j.jesp.2010.04.003

INFORMATION ABOUT THE AUTHORS

Svetlana K. Naryan — Research Assistant, Graduate Student,
Department of Sociology, HSE University. **Phone:** +7 (495) 771-32-32, ext. 12454.
Email: snaryan@hse.ru

Andrey V. Bykov — Candidate of Sociological Sciences, Associate Professor,
Department of Sociology, HSE University; Research Fellow,
Institute of Sociology of FCTAS. **Phone:** +7 (495) 771-32-32, ext. 12454.
Email: a.bykov@hse.ru

Received: 02.12.2021.
