DOI: 10.31857/S023620070026107-4

©2023 И.Г. СНЕТКОВ

СТРАТЕГИИ НАТУРАЛИСТИЧЕСКОГО ОБЪЯСНЕНИЯ МОРАЛЬНОЙ ОТВЕТСТВЕННОСТИ: КОМПАТИБИЛИЗМ И ЭЛИМИНАТИВИЗМ

Снетков Иван Геннадьевич — приглашенный преподаватель, аспирант Школы философии и культурологии.
Национальный исследовательский университет «Высшая школа экономики».
Российская Федерация, 101000 Москва, ул. Мясницкая, д. 11.
ORCID: 0000-0003-4314-3346
i.snetkov@yandex.ru

Аннотация. Статья посвящена исследованию интуиции моральной ответственности с позиций натурализма. Целью статьи было проанализировать, какие возможны натуралистические подходы к концепции моральной ответственности с нейрофилософской точки зрения. Ставится задача пересмотреть концепцию моральной ответственности с целью преодоления либертарианской и классической компатибилистской интуиции и перехода к натуралистическому объяснению. Это исследование показывает, что интуиция моральной ответственности основана на том, что мозг человека наделен двумя различными параллельными системами восприятия каузальности. С учетом этих эмпирических данных предлагаются две возможные стратегии натуралистического объяснения моральной ответственности, концептуально обозначенные как нейрофилософский компатибилизм (НК) и нейрофилософский элиминативизм (НЭ). В защиту первой стратегии НК выдвигается эпистемический аргумент, который утверждает, что знание того, как именно кто-то детерминирован, принципиально ограничено, и поэтому мы не должны основывать

УДК 167.7

НАУЧНЫЕ ИССЛЕДОВАНИЯ

наши представления о моральной ответственности только на фактическом компатибилистском контроле. Вторая стратегия НЭ разделена по силе элиминации моральной ответственности: элиминируется либо вся концепция моральной ответственности, либо только ее онтологическая реальность. В первом варианте перед нами встает необходимость разработки альтернативной системы наказания, права и преодоления устойчивой интуиции моральной ответственности. Во втором варианте допускается сохранение моральной ответственности как эмерджентного функционального свойства и осуществляется полная редукция к онтологическому натурализму или при нередуктивном натуралистическом подходе — к онтологическому плюрализму.

Ключевые слова: моральная ответственность, нейрофилософия, свобода воли, нейронаука, натурализм, компатибилизм, элиминативизм, каузальность, передняя парамедиальная кора, эпистемический аргумент.

Ссылка для цитирования: Снетков И.Г. Стратегии натуралистического объяснения моральной ответственности: компатибилизм и элиминативизм // Человек. 2023. Т. 34, № 3. С. 93–110. DOI: 10.31857/S023620070026107-4

У всех нас есть устойчивая интуиция — мы склонны морально оценивать поступки людей. На этой интуиции основывается концепция моральной ответственности, которая необходима для корректного функционирования общества и его судебно-правовой системы, системы наказаний и т.д. Эта разделяемая всеми интуиция основана на привычке (в терминологии Д. Юма[1]) приписывать ответственность другому, то есть на повседневной практике одобрения или осуждения поступков других людей. Такое поведение настолько фундаментально для социальной жизни людей, что некоторые философы называют эту интуицию фактом [Логинов, 2023: 161].

В настоящей статье не будем затрагивать концептуальный анализ понятия моральной ответственности. Допустим, что возможность приписывания моральной ответственности концептуально обоснована [см.: Логинов, 2023]. Но как обоснована моральная ответственность эмпирически? И какие последствия для философских концептуальных схем имеют данные эмпирической науки?

Нейронаука расширила философские дискуссии о сознании, и ее данные стали неотъемлемой частью философии сознания

[1] «…Руководителем в жизни является не разум, а привычка. Лишь она понуждает ум во всех случаях предполагать, что будущее соответствует прошлому» [Юм, 1996: 665].

в целом и дискуссий по проблеме свободы воли в частности. В данной статье я рассмотрю роль нейронауки для концепции моральной ответственности. Я буду защищать тезис о том, что нейронаука необходима для формулирования эмпирически обоснованной версии моральной ответственности. Моральная ответственность должна быть пересмотрена в некоторых важнейших аспектах: необходимо отказаться от либертарианской и классической компатибилистской интуиции конечной ответственности морального агента как от когнитивного искажения, вводящего в заблуждение, и перейти к концепциям моральной ответственности, которые будут ближе к концепциям, основанным на жестком детерминизме. Таким образом, задача настоящей статьи состоит в пересмотре моральной ответственности с целью преодоления либертарианской и классической компатибилистской интуиции и переходу к натуралистическому объяснению. И в этом пересмотре когнитивная нейронаука играет решающую роль.

Как именно следует пересмотреть концепцию моральной ответственности? Я утверждаю, что с позиций натурализма[2] существуют два основных пути в объяснении моральной ответственности, которые можно условно обозначить как:

1) нейрофилософский компатибилизм;
2) нейрофилософский элиминативизм.

Обе концепции будут рассмотрены в рамках статьи. Еще одна задача этой статьи — объяснить, почему интуитивно кажется, что у нас есть свобода воли? Почему иногда возникает сильная уверенность в том, что наши действия происходят из беспричинного внутреннего источника в нашем «Я»? Когнитивная нейронаука может помочь объяснить происхождение таких интуиций и провести различие между иллюзиями о себе и реалистичным взглядом на наши способности и возможности [Wegner, 2002].

Прежде чем перейти к вопросу о моральной ответственности, необходимо будет показать, каким образом нейронаука объясняет некоторые из наших иллюзий относительно интуиций физической и ментальной каузальности[3], что напрямую связано и с проистекающей из них интуицией моральной ответственности. Затем

---

[2] Следуя здесь Данто, натурализм понимается полемически и определяется как отрицание мнения, что существуют или могут существовать какие-либо сущности, которые, в принципе, лежат за пределами научного объяснения [Danto, 2008: 2].

[3] «Ментальная каузальность — это способность ментальных состояний, убеждений и желаний агентов быть причиной других ментальных или физических состояний» [Волков, 2017]. Также о проблеме физической и ментальной казуальности в контексте нейронаук см.: [Кузнецов, 2014].

я проиллюстрирую, как эмпирически нейронаука подрывает идею классического компатибилизма и либертарианства через эпистемический аргумент в пользу нейрофилософского компатибилизма. Наконец я перейду к некоторым открытиям в области механизмов морального суждения, которые также требуют натуралистического подхода к моральной ответственности.

Объяснение интуиции физической и ментальной каузальности

Итак, возьмем, к примеру, интуицию о том, что мы сами являемся инициаторами наших действий и решений благодаря особой способности, которая не подчиняется обычным законам причинно-следственной связи. Эта интуиция основана на том, что мозг человека наделен двумя различными системами восприятия каузальности: одна из них отведена под ощущение нормальной, физической каузальности, то есть объектов, управляемых физическим законом действия и противодействия [Blakemore, 2001], а другая система предназначена для восприятия убеждений, желаний и намерений других биологических организмов [Gallagher, 2003]. Сравнивая свои восприятия, человек приходит к интуитивному выводу, что одни объекты не подчиняются обычной физической каузальности.

Способность понимать, что другие организмы имеют ментальные состояния, такие как убеждения, желания и намерения, назвали «Теорией Разума» (ТР) [ibid.]. По-видимому, это уникальная человеческая способность. Она развивается после трехлетнего возраста. Ее можно проверить, например, с помощью задач, которые можно решить, только если испытуемые принимают во внимание, что персонажи, изображенные в сказках или мультфильмах, действительно имеют убеждения, желания и намерения.

Эксперименты с Теорией Разума, активирующие восприятие физической и ментальной каузальности, а также противопоставляющие их друг другу, были проведены над испытуемыми с поражением мозга, с психическими расстройствами, такими как аутизм или шизофрения, а также с нормальными испытуемыми с использованием методов функциональной нейровизуализации. Было продемонстрировано, что у пациентов с поражением мозга повреждение лобной коры ухудшает способность решать задачи Теории Разума. Еще более интересными являются результаты экспериментов по нейровизуализации. Было обнаружено, что при выполнении задач Теории Разума задействован целый ряд областей мозга: две латеральные области (верхняя височная борозда

(STS) и передние височные столбы) и две медиальные области (задняя поясная кора/прекунеус и передняя парамедиальная кора (APCC)). Все эти области представляют интерес, но здесь я сосредоточусь на двух — STS и APCC. Было установлено, что STS предпочтительно реагирует на движение биологического объекта по сравнению с небиологическим. Но что такого особенного в таком «биологическом движении»? Движение живой материи определяется не только внешними силами, то есть объектами, сталкивающимися друг с другом и подчиняющимися закону действия и противодействия, но и внутренними силами с особыми паттернами движения. Когда человек видит паттерны, не обусловленные очевидными для него внешними силами, он автоматически склонен интерпретировать их как волевые или намеренные [Castelli et al., 2002].

Вторая область, передняя парамедиальная кора (APCC), активна, когда нормальные испытуемые должны оценивать истории, в которых необходимо «чтение мыслей», то есть понимание ментальных состояний персонажей историй. Это было показано в нескольких нейровизуализационных исследованиях. В одном из них было обнаружено, что APCC специфически активизируется, когда испытуемые занимают интенциональную позицию [Gallagher, 2002]. Людей сажали за компьютеры, и они играли в специальную версию игры «камень, ножницы, бумага» номинально либо против человека, либо против компьютера. И игра против человека была связана со специфической активацией APCC. Функция этого крошечного участка медиальной части префронтальной коры, APCC, каким-то образом связана с представлением ментальных состояний (по сравнению с обычными физическими движениями) других людей. Важно, что испытуемому достаточно было лишь предположить, что его товарищ по игре является человеком, чтобы активировать APCC, хотя на самом деле в этом эксперименте противником всегда был компьютер, а испытуемым лишь в некоторых случаях говорили, что это человек. Для представления ментальных состояний другого достаточно лишь того, чтобы испытуемый взаимодействовал так, как будто другой игрок — это человек. Представление о таких состояниях как обладающих особым статусом «ментальных» состояний появляется из-за отделения моделируемых репрезентаций о состоянии другого от ощущений причинно-следственной связи физического мира. Этот механизм разделения и был приписан передней парамедиальной коре [Leslie, 1994]. Другими словами, все события в мире, воспринимаемые человеком, сопровождаются перцептивным ощущением физической

каузальности — человек наблюдает некоторые события (действия другого человека), которые не сопровождаются этим ощущением.

С эволюционной точки зрения необходимо поставить вопрос о том, в чем преимущество Теории Разума. Ответ вполне ясен: если вам приходится взаимодействовать с сородичами, у которых есть убеждения, желания и намерения, полезно уметь предугадывать, что представляют собой их ментальные состояния. Исходя из этой идеи нейропсихиатром Хенриком Вальтером был проведен эксперимент с использованием мультфильмов и фМРТ. Он проверял гипотезу о том, что АРСС специфически активируется, когда задача Теории Разума включает социальное взаимодействие [Walter, 2004]. Было обнаружено, что АРСС не активируется в мультфильмах, изображающих намерение в действиях одного человека (например, когда он берет яблоко с явным намерением его съесть), но только в тех случаях, где речь идет о коммуникативных намерениях (например, когда человек хочет обменять яблоко на грушу у другого человека). Его функция заключается не только в отделении ментальных репрезентаций от физического казуально-замкнутого мира, но и в отделении и представлении ментальных репрезентаций при взаимодействии минимум двух сознаний. Появляется возможность сопоставлять свои намерения и намерения другого. Основываясь на этом, можно предположить, что происхождение интенциональной позиции имеет социальную природу. Так, у шизофреников есть проблемы с правильным различением собственного разума от разума других людей, они демонстрируют пониженную активацию АРСС [Walter et al., 2005]. Есть также эмпирические свидетельства того, что состояние мозга по умолчанию таково, что собственные ментальные состояния представлены только в соотношении с ментальными состояниями других [Gusnard, 2001].

Таким образом, ментальная каузальность является формой обоснования поведения других людей, в котором репрезентированные ментальные состояния являются элементами процессов мозга.

И тот факт, что у нас есть устойчивая интуиция, что ментальная каузальность не может быть сведена к физической каузальности, объясняется тем, что существуют две различные области мозга, STS и АРСС, которые автоматически активируются при наблюдении за «биологическими объектами», которые движутся определенным образом и (социально) взаимодействуют друг с другом. Параллельность этих двух процессов на нейрональном уровне и обусловливает наличие феноменальной ментальной каузальности.

Свобода воли и моральная ответственность. Нейрофилософский компатибилизм

Проблему свободы воли можно изучать и как проблему моральной философии. При таком подходе вопрос звучит не столько как: «Есть ли у нас свобода воли?», а скорее: «Какого рода свободную волю (какого рода компетентность) мы должны предположить, чтобы допустить моральную ответственность?». Ответственность — это нормативный вопрос, а не вопрос фактов. Свобода воли является необходимым условием моральной ответственности. Если и только если человек действует по собственной свободной воле, он несет моральную ответственность за свои действия[4]. Классический компатибилизм[5] здесь присоединяется к либертарианству в отношении утверждений об ответственности, за исключением того, что концепция свободы воли отличается. Основная проблема моральной ответственности в этой позиции заключается в том, что она влечет за собой идею конечной ответственности. Но точно так же, как неограниченная, неопределенная абсолютная свобода воли является концептуальной невозможностью в натуралистических рамках, идея конечной ответственности тоже невозможна в рамках натурализма. Под конечной ответственностью в данном случае имеется в виду строгое соответствие эпистемическому условию моральной ответственности в случае как компатбилистских, так и либертарианских трактовок свободы воли. Оба варианта сталкиваются с проблемой соответствия эпистемического условия моральной ответственности. Эти трудности хорошо иллюстрируются в дискуссии об аргументе удачи [см., напр.: Беседин, 2021].

Почему классический компатибилизм без проблем отказывается от рассмотренных в начале статьи интуиций каузальности, связанных со свободой воли, но так упорно сопротивляется отказу от некоторых интуиций, связанных с моральной ответственностью? Боятся ли люди бесконечного регресса оправданий, когда мы начинаем объяснять каузальную историю того или иного поведения?

Согласно компатибилизму, наше моральное отношение к другим и приписывание им ответственности должно зависеть

---

[4] У нас есть привычка ожидать, что предметы всегда будут падать вниз, а не вверх, но нет ее в отношении действий других людей. Хотя это различие перестает быть столь непосредственно данным в опыте. Научное описание поведения людей уже порождает тенденцию говорить о человеке не как об агенте, но как о биомашине, от которой ожидают действий «по привычке».

[5] Под классическим компатибилизмом в настоящей статье понимается компатибилизм в духе Г. Франкфурта [см.: Маккенна, 2009].

от знания того, каким образом детерминированы люди. У нас есть основания оправдывать людей, если мы узнаем, что они детерминируются силами, которые не находятся под их компатибилистским условием контроля моральной ответственности. Это же является одной из причин, почему многие люди считают убедительным, что неблагополучное детство, ведущее к насилию, имеет хоть какую-то оправдательную силу. Существует множество исследований, демонстрирующих эту объяснительную регрессию наших моральных поступков [Сапольски, 2019]. Хотя мы уже знаем некоторые детерминанты, которые могут изменить наше моральное отношение к людям, мы также знаем (индуктивно), что их будет гораздо больше. Мы понимаем, что по практическим и принципиальным причинам мы никогда не узнаем всего, что касается реальных причин их поведения. Иногда этот факт интерпретируется как «аргумент от манипуляции» в пользу жесткого детерминизма, и в таком виде он не валиден в отношении свободы воли [Кузнецов, 2023: 228–230], но, с моей точки зрения, данный факт релевантен к вопросу об условиях моральной ответственности. Ввиду ограниченности знания мы будем несправедливы по отношению ко многим людям. Например, когда мы хвалим людей за то, что они сумели компенсировать ужасные жизненные события, мы должны поставить нашу похвалу в зависимость от того, какой у них тип серотонинового транспортера. То, что кто-то с «хорошими генами» справился с этой ситуацией, заслуживает меньшей похвалы, чем если бы это сделал кто-то с «плохими генами», потому что первому было гораздо легче это сделать. Интересен вопрос, что мы на самом деле восхваляем, когда кто-то справляется с трудной ситуацией. Вполне возможно, что то, за что мы хвалим, — это усилия, которые кто-то приложил для достижения цели [Kane, 1996]. То же самое верно и для других достижений: представьте, что кто-то бросает курить, потому что сигареты больше не кажутся ему вкусными, в то время как другой бросает, но с большим трудом, потому что сигареты ему все еще нравятся. Разве последний не заслуживает большей похвалы? И разве не так, что в конечном счете безуспешная, но требующая усилий борьба с курением заслуживает большей похвалы, чем отказ от курения того, кто больше не любит сигареты?

В обычной жизни мы учитываем такие факторы, насколько они нам известны. Действительно, мы ставим свою моральную оценку в зависимость от обстоятельств — насколько они нам известны. Однако, и это главный аргумент, существует множество факторов, влияющих на наш физический мозг, которые определяют, как люди справляются с морально значимыми ситуациями в жизни, о которых мы просто не знаем. Оценивая их достижения и недостатки,

мы никогда не сможем учесть все важные факторы, определяющие их жизнь. Поэтому нам будет трудно, если вообще возможно, быть по-настоящему справедливыми к другим. Итак, допустим все люди детерминированы, и это имеет последствия для нашей концепции моральной ответственности. Однако, в отличие от жесткого детерминизма, мы можем попытаться продумать такой нейрофилософский компатибилизм, который настаивает на центральной идее компатибилизма о том, что важно, как они детерминированы. Однако мы должны признать, что нам никогда не будут известны все значимые факторы. Это можно назвать эпистемическим аргументом в пользу такого рода нейрофилософского компатибилизма.

Эпистемический аргумент в пользу нейрофилософского компатибилизма похож на аргумент от эпистемического индетерминизма в дебатах о свободе воли — только он направлен назад, а не вперед во времени. Эпистемический индетерминизм утверждает, что по принципиальным причинам мы не можем предсказать самих себя, так что мы субъективно свободны, даже если должны быть объективно детерминированы. Аналогично я утверждаю, что наше знание о том, как кто-то детерминирован, обязательно ограничено (потому что мы не всеведущи), и поэтому мы не должны основывать наши представления о моральной ответственности только на фактическом компатибилистском контроле. Разница заключается в том, что эпистемический аргумент в пользу свободы воли подразумевает ссылку на самореференцию, тогда как эпистемический аргумент в пользу пересмотра моральной ответственности — нет.

Нейрональные механизмы морали. Нейрофилософский элиминативизм

Говоря о нейрофилософском компатибилистском подходе к моральной ответственности, я концентрировался на вопросе о том, что существуют скрытые факторы, объясняющие наше поведение, которые имеют отношение к нашему моральному суждению об агенте. То есть я сосредоточился на субъекте, который оценивается как ответственный. Однако есть и другая возможность — сосредоточиться не на агенте, правонарушителе, а на судье, карателе. Каковы психологические и нейронные механизмы, лежащие в основе морального осуждения и наказания?

Понимание того, что мозг имеет две различные системы для оценки и приписывания физической и ментальной каузальности, соответственно, объясняет, почему кажется столь очевидным, что ментальная каузальность является чем-то категорически отличным от физической каузальности.

Это иллюстрирует широко известное исследование Дж. Хайдта, из которого следует, что наши суждения обычно основаны на быстрых, автоматических и интуитивных оценках, в которых эмоциональные механизмы играют важную роль [Haidt, 2001].

Существует несколько способов изучения морального познания и суждений с помощью нейровизуализации. Например, людям могут быть представлены визуальные сцены, содержащие моральные проступки, и аналогичные сцены, очень похожие, но не изображающие моральных проступков. Другой подход заключается в сравнении различных типов оценок. Например, были проведены исследования, посвященные различиям между семантическими решениями (суждение о том, является ли предложение семантически правильным) и простыми моральными решениями (суждение о том, является ли предложение морально правильным). Другой подход заключается в том, чтобы ставить людей перед моральными и неморальными дилеммами в сканере и искать различия. После создания таких парадигм исследователи могут попытаться разделить их на несколько компонентов. Например, можно изучить влияние телесных повреждений на моральные суждения [Heekeren et al., 2005]. Обобщим[6] все это.

Во-первых, одним из основных результатов почти всех исследований морального восприятия и суждений была активация уже знакомого нам набора областей из Теории Разума. Если мы вспомним, что зоны мозга, относящиеся к Теории Разума, отвечают за восприятие намеренной каузальности и социального взаимодействия, это добавляет новый смысл: моральные суждения — это феномен исключительно социальной сферы.

Во-вторых, можно показать, что в процесс вовлечена обработка эмоциональной информации. Например, миндалина играет роль в восприятии морального проступка или когда испытуемым предъявляются моральные дилеммы, требующие убить нескольких человек, чтобы спасти многих [Greene et al., 2004].

Островковая доля, область, которая, как известно, участвует в восприятии боли и отвращения, активируется, когда люди сталкиваются с несправедливыми предложениями и отказываются от них, и ее активация коррелирует с количеством несправедливости [Sanfey et al., 2003]. Островковая доля также участвует в сопереживании, например, когда вы видите, что ваш близкий человек страдает от боли [Singer et al., 2004].

При возникновении морального чувства активируются те же самые нейроны, что отвечают и за чувство тошноты, которое мы

[6] Обзор по нейронным механизмам морали см. в статье: [Greene, Haidt, 2002].

испытываем от пищи. По всей видимости, нейроны этой области мозга и взяли на себя функцию морального чувства, когда она возникла в процессе социальной эволюции.

Помимо этого, есть дорсальное полосатое тело — область, относящаяся к системе вознаграждения человеческого мозга, она активна у людей, которые наказывают других людей [de Quervain et al., 2004]. Желание наказать и отомстить действительно является очень сильным импульсом, если человеку причинили вред или, тем более, если причинили вред близкому родственнику. Месть является важным вопросом в психологии наказания, и существует большая дискуссия о роли возмездия против консеквенциализма в оправдании наказания. Понимание внутренней работы морального мозга будет иметь значение для этой дискуссии.

Третий интересный момент, вытекающий из этих исследований, заключается в том, что иногда ни зоны мозга, относящиеся к Теории Разума, ни эмоциональные центры не задействованы.

Например, если люди принимают несправедливое денежное предложение по рациональным соображениям (потому что если они его отвергнут, то ничего не получат), активируются латеральные участки префронтальной коры, связанные с более рациональными способностями и когнитивным контролем. Другая область префронтальной коры, ее передняя часть, которая связана с когнитивной, а не эмоциональной обработкой информации, показала активность в исследованиях моральной дилеммы у тех испытуемых, которые решают, что они должны пожертвовать жизнью нескольких человек, чтобы спасти жизни многих, то есть которые принимают утилитарное решение [Greene et al., 2004].

Мы можем сказать, что утилитарный тип рассуждений вытесняет эмоции и моральное чувство при вынесении моральных суждений. Логично предположить, что все нормативные этические теории обладают таким свойством, давая людям иные, рационалистические схемы рассуждений. Это подменяет наши суждения, которые мы называем моральными. Те суждения, которые высказывает человек в моменте возникновения ситуации морального выбора, и те, что он высказывает, вспоминая этическую нормативную теорию, которой хотел бы следовать, имеют совершенно разную физическую природу. Тут есть новый вопрос: а какие из двух совершенно разных типов суждений следует считать моральными суждениями? А если оба типа, то на каком основании мы их объединяем? Только лишь потому, что они относятся к одной внешней ситуации морального выбора? Этот вопрос требует отдельной исследовательской проработки.

Если мы хотим, чтобы этические принципы были эффективными, мы должны учитывать, как работает мораль. Более того,

я убежден, что эти выводы имеют значение для формирования нашего мышления по вопросам морали. Например, это может привести к пониманию того, что эмпирическая обоснованность моральных суждений, которые обычно возникают быстро и автоматически, объясняется не столько тем, что моральная оценка обоснована (рационально), сколько тем, что она основана на быстрых, автоматических и когнитивно непроницаемых механизмах (которые являются, по существу, эмоциональными). Заметим, что это не означает ни того, что эмоциональные механизмы не могут привести к обоснованным суждениям, ни того, что они не полезны. Напротив, я считаю, что вышеупомянутые исследования убедительно подтверждают тезис о том, что эмоциональные механизмы являются необходимым ингредиентом для моральных суждений и поведения, и эта роль недооценивается рационалистическими подходами к морали. Эмоции, как известно, являются отличным решением проблемы моральной мотивации, то есть проблемы того, почему осознание правоты должно приводить к совершению правильных поступков.

Другой интересный вопрос заключается в том, чем моральная оценка отличается от других видов оценки. Одним из способов эмпирического подхода к этому является проведение экспериментов, в которых различные компоненты моральных суждений разделяются. Однако это непросто. Например, утверждается, что попытка отделить эмоциональные факторы от моральных суждений невозможна, поскольку эмоциональные факторы (опять же) могут быть необходимым ингредиентом для моральных суждений. Аналогичная проблема возникает, если мы пытаемся отделить социальное суждение от морального. Кроме того, трудно сконструировать сценарии социального взаимодействия, которые не были бы запятнаны моралью. Может быть, моральное суждение — это не что иное, как социальное суждение, включающее в мозгу области Теории Разума и эмоции?

Так, американский когнитивный нейробиолог М. Газзанига утверждает, что свобода действий и моральная ответственность являются социальными понятиями и поэтому не могут находиться в мозгу, они появляются только в социальном взаимодействии [Gazzaniga, 2005].

При таком сценарии мы встаем на путь элиминации моральной ответственности. Примером подобной элиминативной стратегии является работа Б. Уоллера «Против моральной ответственности» [Waller, 2011]. Уоллер утверждает, что, хотя люди в целом отвечают требованиям компатибилизма к моральной ответственности, мы тем не менее не несем моральной ответственности; никто никогда не заслуживает похвалы или наказания. Это, конечно,

контринтутивное утверждение, противоречащее здравому смыслу или по крайней мере привычной социальной практике. Тем не менее, по его мнению, моральная ответственность несовместима с научным натурализмом и морально неоправданна, поскольку заставляет нас игнорировать причины человеческого поведения и зачастую вредные результаты привлечения людей к моральной ответственности. Компатибилисты принимают как данность укоренившуюся в культуре привычку приписывания моральной ответственности. Чем внимательнее мы изучаем каузальный контекст атрибуции моральному агенту ответственности, чем дальше мы уходим «в лес» по цепи причинно-следственных связей, тем менее значимы характеристики и способности человека в решении. Это все больше затрудняет даже интуитивное возложение вины исключительно на агента, а именно это и делает атрибуция моральной ответственности. То, что агент может даже с настойчивостью одобрять свои плохие черты и определять себя через них — через расточительность, эгоизм, нечестность, жестокость, — это просто результат определенной каузальной истории, а не основание для возложения на него вины за то, что он стал таким. Моральная ответственность основана на деонтологических аргументах о благе и противоречит консеквенционалистким подходам, поскольку они оказываются неспособными обосновать выгоду от страданий. Компатибилист порицает агента за то, что он «заслуживает», что зачастую может противоречить благу как всего человечества, так и самого компатибилиста. Подход без моральной ответственности, а значит, и осуждения был бы направлен на поиск решений оптимального взаимодействия между людьми.

Таким образом, под нейрофилософском элиминативизмом в объяснении моральной ответственности понимается элиминация моральной ответственности как процесса приписывания агентом моральной ответственности и интерпретация агента как обусловленного причинно-следственными связями мозга, функционирующего в рамках жесткого детерминизма.

В качестве возражения к элиминативистскому аргументу «от мозга», онтологически выраженному во фразе «я есть мозг», предъявляется претензия в сведении к онтологическому абсурду. Так, утверждается, что «в конечном счете мы оказываемся в ситуации, когда любые утверждения о существовании становятся невозможными» [Кузнецов, 2023: 231]. Но я утверждаю, что это не так, элиминативисты не обязательно настаивают на такой «грубой» редукции, в данном случае происходит подмена понятий. Утверждение «я есть мозг» от элиминативистов не является онтологическим утверждением, но преследует своей целью устранение избыточной объяснительной сущности — понятия «Я» — из наших

концептуальных, натуралистически ориентированных объяснительных схем. Таким образом, это утверждение является эпистемологическим. В вопросе об онтологии натуралист может остановиться и на уровне нередуктивного натурализма, постулируя онтологический плюрализм [Northoff, 2004]. Отношение между онтологией и эпистемологией не определяется онтологическими предпосылками, напротив, эпистемология первична [ibid.].

Но даже если мы потребуем от элиминативиста постулировать какую-то натуралистическую онтологию, например физикализм, это не сведет «аргумент от мозга» к абсурду. Это и показывает стратегия «слоистой» онтологии Дж. Ким [Kim, 1998]. Как заметил Кузнецов, «Я» реализовано мозгом, мозг — клетками [Кузнецов, 2023: 231], и далее мы доходим до фундаментальных физических частиц, которые и являются первичной субстанцией в физикализме. Биологический, физико-химический и психологический слои описания реальности составляют различные слои одной физической реальности. И эти слои ограничены эпистемологическим уровнем, а не онтологическим (они «не существуют» в строгом смысле). Понятия «мозг», «клетка» не являются здесь сущностями, «вещами» и объяснены как функциональные процессы, которые из эпистемологической необходимости были категориально обозначены через концептуальные понятия, отсылающие к некоторому функциональному свойству сложно организованной физической материи (например, совокупность химических процессов клеточной жизнедеятельности в биологическом понятии «клетка»). Сама же необходимость выделять подобного рода понятия не требует постулирования за ними онтологических сущностей, а может объясняться, например, через наблюдаемые функциональные эмерджентные свойства в сложных системах. Стол массой 10 килограммов обладает свойством, которым не обладает ни один из его микрокомпонентов. Такой стол обладает причинно-следственной силой, которой нет ни у одного из его микрокомпонентов [Kim, 1998]. Такие эмерджентные свойства вполне обыденны. Они проявляются, если существует определенная конфигурация элементарных физических частиц. И эта конфигурация сама по себе является эмерджентным свойством.

Таким образом, нейрофилософский элиминативизм в объяснении моральной ответственности сводится к одному из двух вариантов:

1) элиминация всей концепции моральной ответственности, и тогда перед нами встает необходимость разработки альтернативной системы наказания [см., напр.: Pereboom, 2001] и преодоления устойчивой интуиции моральной ответственности;

2) элиминация онтологических оснований моральной ответственности, а именно сохранение моральной ответственности как

эмерджентного функционального свойства и полная редукция к онтологическому натурализму или при нередуктивном натуралистическом подходе — к онтологическому плюрализму.

* * *

Целью данной статьи было исследовать, какие возможны натуралистические подходы к концепции моральной ответственности с нейрофилософской точки зрения. В дискуссиях о моральной ответственности нейронаука подрывает возможность либертарианской и классической компатибилисткой трактовки моральной ответственности как действия человека по собственной свободной воле с условием полного контроля, при котором он несет моральную ответственность за них. Кроме того, при таком понимании моральной ответственности становится проблематичным справедливое приписывание моральной ответственности агенту. Мы должны признать, что нам никогда не будут известны все значимые каузальные факторы, определяющие поведение человека. Это можно назвать эпистемическим аргументом в пользу необходимости разработки нейрофилософского компатибилизма в отношении моральной ответственности.

Вместе с тем в свете последних исследований стали яснее и эмпирические основания процесса приписывания моральной ответственности, что позволяет задать и концептуальные рамки того, какими способами возможно натуралистическое объяснение моральной ответственности. Были выявлены две основные объяснительных стратегии, одна из которых также делится на две вариации:

1) нейрофилософский компатибилизм;
2) нейрофилософский элиминативизм;
   a. элиминация всей концепции моральной ответственности (сильный элиминативизм);
   b. элиминация онтологических оснований моральной ответственности (слабый элиминатвизм).

Данные стратегии требуют дальнейших уточнений и пересмотра по мере появления все новых нейронаучных исследований, касающихся проблемы моральной ответственности.

Strategies of Naturalistic Explanation of Moral Responsibility: Compatibilism and Eliminativism

Ivan G. Snetkov
Visiting Lecturer, Postgraduate Student of the School of Philosophy and Cultural Studies.
National Research University Higher School of Economics.

НАУЧНЫЕ ИССЛЕДОВАНИЯ

11 Myasnitskaya Str., 101000 Moscow, Russian Federation.
ORCID: 0000-0003-4314-3346
i.snetkov@yandex.ru

Abstract. The article explores the intuition of moral responsibility from a naturalistic perspective. The aim of the article was to analyse what possible naturalistic approaches to the concept of moral responsibility are possible from a neurophilosophical point of view. The aim is to revisit the concept of moral responsibility in order to remove the libertarian and classical compatibilist intuition and move towards a naturalistic explanation. This study shows that the intuition of moral responsibility is based on the fact that the human brain is endowed with two different parallel systems of perception of causality. With this empirical evidence in mind, two possible strategies for a naturalistic explanation of moral responsibility are proposed, conceptualised as neurophilosophical compatibilism (NC) and neurophilosophical eliminativism (NE). The first NC strategy is defended by an epistemic argument which argues that knowing exactly how someone is determined is fundamentally limited, and so we should not base our ideas about moral responsibility on actual compatibilist control alone. The second strategy of NE is divided on the strength of the elimination of moral responsibility: either the whole concept of moral responsibility or only its ontological reality is eliminated. In the first option, we are faced with the need to develop an alternative system of punishment, law and overcoming the persistent intuition of moral responsibility. In the second scenario, the preservation of moral responsibility as an emergent functional property is accepted and a total reduction to ontological naturalism or, in the case of an unreductive naturalistic approach, to ontological pluralism is achieved.

Keywords: moral responsibility, neurophilosophy, free will, neuroscience, naturalism, compatibilism, eliminativism, causality, anterior paramedial cortex, epistemic argument.

For citation: Snetkov I.G. Strategies of Naturalistic Explanation of Moral Responsibility: Compatibilism and Eliminativism // Chelovek. 2023. Vol. 34, N 3. P. 93–110. DOI: 10.31857/S023620070026107-4

Литература/References

Беседин А.П. Клещи удачи Нила Леви против компатибилистских теорий моральной ответственности // Вестник РГГУ. Сер. Философия. Социология. Искусствоведение. 2021. № 3. С. 26–37.

Besedin A.P. Kleshchi udachi Nila Levi protiv kompatibilistskih teorij moral'noj otvetstvennosti [Neil Levy's Ticks of Fortune Against Compatibilist Theories of Moral Responsibility]. *Vestnik RGGU. Ser.: Filosofiya. Sotsiologiya. Iskusstvovedenie*. 2021. N 3. P. 26–37.

Волков Д.Б. Решение проблемы ментальной каузальности в биологическом натурализме Дж. Серла // Философская мысль. 2017. № 2. С. 1–12.

Volkov D.B. Reshenie problemy mental'noj kauzal'nosti v biologicheskom naturalizme Dzh. Syorla [Solving the Problem of Mental causality in Biological Naturalism by J. Searle]. *Filosofskaya mysl'*. 2017. N 2. P. 1–12.

Кузнецов А.В. Когнитивные исследования и проблема ментальной каузальности // Вопросы философии. 2014. № 3. С. 133–142.

Kuznetsov A.V. Kognitivnye issledovaniya i problema mental'noj kauzal'nosti [Cognitive research and the problem of mental causality]. *Voprosy philosofii*. 2014. N 3. P. 133–142.

Кузнецов А.В. Химера натурализма и свобода воли // Эпистемология и философия науки. 2023. Т. 60, № 1. С. 221–240.

Kuznetsov A.V. Himera naturalizma i svoboda voli [Chimera naturalism and free will]. *Epistemologiya i filosofiya nauki*. 2023. Vol. 60, N 1. P. 221–240.

Логинов Е.В. Скептицизм и моральная ответственность // Философский журнал. 2023. Т. 16, № 1. С. 160–176.

Loginov E.V. Skepticizm i moral'naya otvetstvennost' [Skepticism and Moral Responsibility]. *Filosofskii zhurnal*. Vol. 16, N 1. P. 160–176.

Маккенна М. Компатибилизм // Стэнфордская философская энциклопедия: переводы избранных статей / под ред. Д.Б. Волкова, В.В. Васильева, М.О. Кедровой. 2009 [Электронный ресурс]. URL: http://philosophy.ru/compatibilism/ (дата обращения: 14.02.2023).

McKenna M. Compatibilism. *Stenfordskaya filosofskaya entsiklopediya: perevody izbrannykh statei* [The Stanford Encyclopedia of Philosophy], D.B. Volkov, V.V. Vasilyev, M.O. Kedrova (eds). Winter 2009 Edition [Electronic resource]. URL: http://philosophy.ru/compatibilism/ (date of access: 14.02.2023).

Сапольски Р. Биология добра и зла: Как наука объясняет наши поступки / пер. с англ. Ю. Аболиной, Е. Наймарк. М.: Альпина нон-фикшн, 2019.

Sapolsky R. *Biologiya dobra i zla: Kak nauka ob"yasnyaet nashi postupki* [Behave: The Biology of Humans at Our Best and Worst]. Moscow: Alpina non-fiction Publ., 2019.

Юм Д. Трактат о человеческой природе / пер. с англ. С.И. Церетели // *Юм Д.* Соч.: в 2 т. Т. 1 / вступ. ст. А.Ф. Грязнова; примеч. И.С. Нарского. М.: Мысль, 1996.

Hume D. *Traktat o chelovecheskoĭ prirode* [A Treatise on Human Nature], transl. from English by S.I. Tsereteli. Hume D. *Soch.: v 2 t.* [Essays: in 2 vols.]. Vol. 1, introductory article by A.F. Gryaznov; notes by I.S. Narsky. Moscow: Mysl' Publ., 1996.

Blakemore S.J. et al. How the Brain Perceives Causality: An Event-Related fMRI Study. *Neuroreport*. 2001. Vol. 12, N 17. P. 3741–3746.

Castelli F. et al. Autism, Asperger Syndrome and Brain Mechanism for the Attribution of Mental States to Animated Shapes. *Brain*. 2002. Vol. 125. Pt. 8. P. 1839–1849.

Danto A. Naturalism. *The Encyclopedia of Philosophy*. Editor Stone, 2008. P. 2.

de Quervain D.J., Fischbacher U., Treyer V., et al. The neural basis of altruistic punishment. *Science*. 2004. Vol. 305, N 5688. P. 1254–1258.

Gallagher H.L., Frith C.D. Functional Imaging of "Theory of Mind". *Trends Cogn. Sci*. 2003. Vol. 7, N 2. P. 77–83.

Gallagher H.L., Jack A.I., Roepstorff A., Frith C.D. Imaging the intentional stance in a competitive game. *Neuroimage*. 2002. Vol. 16, N 3. Pt. 1. P. 814–821.

Gazzaniga M.S. *The Ethical Brain*. Washington, DC: Dana Press, 2005.

Greene J., Haidt J. How (and where) does moral judgment work? *Trends Cogn Sci*. 2002. Vol. 6, N 12. P. 517–523.

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, N 2. P. 389–400.

Gusnard D.A., Raichle M.E., Raichle M.E. Searching for a baseline: functional imaging and the resting human brain. *Nature Neuroscience Reviews*. 2001. Vol. 2, N 10. P. 685–694.

Haidt J. The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychological Review*. 2001. Vol. 108, N 4. P. 814–834.

Heekeren H.R., Wartenburger I., Schmidt H., Prehn K., Schwintowski H.P., Villringer A. Influence of bodily harm on neural correlates of semantic and moral decision-making. *Neuroimage*. 2005. Vol. 24, N 3. P. 887–897.

Kane R. *The Significance of Free Will*. New York: Oxford University Press, 1996.

Kim J. *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. Cambridge (Mass.): MIT Press, 1998.

Leslie A.M. Pretending and believing: issues in the theory of ToMM. *Cognition*. 1994. Vol. 50, N 1–3. P. 211–238.

Northoff G. What is neurophilosophy? A methodological account. *Journal for General Philosophy of Science.* 2004. Vol. 35. P. 91–127.

Pereboom D. *Living without free will*. Cambridge, U.K.; New York: Cambridge University Press, 2001.

Sanfey A.G., Rilling J.K., Aronson J.A., Nystrom L.E., Cohen J.D. The neural basis of economic decision-making in the Ultimatum Game. *Science.* 2003. Vol. 300, N 5626. P. 1755–1758.

Singer T., Seymour B., O'Doherty J., Kaube H., Dolan R.J., Frith C.D. Empathy for pain involves the affective but not sensory components of pain. *Science*. 2004. Vol. 303, N 5661. P. 1157–1162.

Waller B. *Against Moral Responsibility*. Cambridge, MA: MIT Press, 2011.

Walter H. et al. A Parametric Theory-of-Mind Design to Investigate Social Cognition in Schizophrenia. Poster presented at the *International Congress of Schizophrenia Research*, Savannah, Georgia, USA, 2005.

Walter H., Adenzato M., Ciaramidaro A., Enrici I., Pia L., Bara B.G. Understanding intentions in social interaction: the role of the anterior paracingulate cortex. *Journal of Cognitive Neuroscience*. 2004. Vol. 16, N 10. P. 1854–1863.

Wegner D. *The Illusion of Conscious Will*. Cambridge, Mass.: MIT Press, 2002.