

Компатибилистское опровержение аргумента манипуляции*

М.А. Секацкая

Аргумент манипуляции Дерка Перебума направлен на то, чтобы показать, что со-ответствия критериям свободы воли, выдвигаемым компатибилистами, недостаточно для наличия у агента моральной ответственности, поскольку некий агент может удовлетворять этим критериям и при этом не быть морально ответственным. Этот аргумент является одним из наиболее влиятельных инкомпатибилистских аргументов в современных дискуссиях о совместимости физического детерминизма и моральной ответственности.

В статье анализируется структура аргумента, критика этого аргумента со стороны философов-компатибилистов и ответ Перебума на эту критику. Демонстрируется, что две разные стратегии компатибилистской защиты, так называемый «мягкий» и «жесткий» ответы на аргумент манипуляции, частично, хотя и не полностью, ставят под сомнение посылки аргумента. В заключительной части статьи демонстрируется, каким образом совместное применение этих стратегий позволяет окончательно опровергнуть аргумент манипуляции.

КЛЮЧЕВЫЕ СЛОВА: аргумент манипуляции, Дерк Перебум, компатибилизм, моральная ответственность.

СЕКАЦКАЯ Мария Александровна – кандидат философских наук, старший преподаватель, кафедра философии науки и техники, Институт философии Санкт-Петербургского государственного университета.

maria.sekatskaya@gmail.com

Статья поступила в редакцию 11 июня 2018 г.

Цитирование: *Секацкая М.А.* Компатибилистское опровержение аргумента манипуляции // Вопросы философии. 2018. № 12. С. 185–196.

1. Введение

Если наши физические и психологические характеристики полностью детерминированы состоянием мира до нашего рождения и законами природы, можем ли мы быть обладать свободой воли? Можем ли мы считать себя и других людей морально ответственными за совершаемые действия? Классический компатибилизм Юма и его последователей, основанный на анализе смысла выражения «я мог поступить иначе», дает положительный ответ на оба вопроса. Несмотря на то, что некоторые инкомпатибилисты утверждали, что эпоха классического компатибилизма осталась позади, не все исследователи с этим согласны, что позволяет считать классический компатибилизм по-прежнему актуальной позицией [Васильев 2017]. Кроме того, в XX в. появляется новая версия компатибилистского анализа, основанная не на смысле слов обыденного языка, а на моральных интуициях относительно принципов приписывания ответственности. Приписывание ответственности, по убеждению таких философов как Питер Стросон, Гарри Франкфурт, Дэниэл Деннет, Джон Фишер, Р. Джей Уоллес

* Исследование выполнено за счет гранта Российского научного фонда (проект № 18-18-00222).

[Strawson 1962; Frankfurt 1971; Dennett 2003; Fischer 2006; Wallace 1994] и др., исходит не из логических, физических или метафизических истин относительно того, что представляют собой альтернативные возможности и могут ли они существовать в детерминированном мире, а из особенностей человеческой психологии и практик общественной жизни. Компатибилистские теории психологического типа утверждают, что детерминированный агент, обладающий определенными психологическими свойствами и находящийся в определенной ситуации, наделен такой степенью свободы воли, которой достаточно для приписывания этому агенту моральной ответственности, и дают, тем самым, положительный ответ как минимум на второй из приведенных выше вопросов о свободе воли.

Компатибилистские теории психологического типа получили широкое распространение в дискуссиях о свободе воли, имевших место в последние десятилетия, тем самым поставив инкомпатибилистов перед необходимостью высказать формальное возражение, не сводящееся к интуиции о том, что моральная ответственность и физический детерминизм никак не могут быть совместимы. Аргумент манипуляции американского философа Дерка Перебума [Pereboom 2001; Pereboom 2005; Pereboom 2014] представляет собой один из наиболее известных современных инкомпатибилистских аргументов. Он призван продемонстрировать, что компатибилистские критерии недостаточны для наличия у агента моральной ответственности, так как нет существенной разницы между детерминированным агентом, отвечающим компатибилистским критериям ответственности, и подобным ему манипулируемым агентом, не несущим ответственности. В данной статье я проанализирую и опровергну этот аргумент Перебума.

2. Четыре истории про профессора Плама

Пербум начинает с описания истории, вызывающей у читателя моральное негодование. История заключается в том, что некий профессор Плам, живущий в детерминированном мире, желает убить мисс Уайт, исходя из соображений эгоистического характера. Его желание убить не является навязчивым, находится в гармонии с его характером, соответствует его желаниям более высокого порядка — он хочет быть таким человеком, который действует на основании этого желания. Плам размышляет о своем желании, взвешивает за и против, и действует в соответствии с ним. Пербум указывает, что, поскольку Плам отвечает критериям моральной ответственности, выдвигаемым компатибилистскими теориями психологического типа (Пербум называет такие теории «интеграционистскими», поскольку они подразумевают определенную интеграцию желаний, целей, ценностей, способностей и характера агента), он должен быть признан компатибилистами ответственным за совершение данного преступления.

Эти теории Пербум стремится опровергнуть посредством четырехступенчатого аргумента манипуляции, который должен продемонстрировать читателю, что Плам может не нести моральной ответственности, несмотря на то, что он соответствует выдвигаемым этими теориями критериям. Этот аргумент включает четыре вариации на тему изложенной выше истории Плама, между которыми, по утверждению Пербума, нет морально значимых отличий. Рассмотрим их по порядку.

Случай 1. Профессор Плам был создан нейроучеными, которые могут прямо манипулировать им, используя технологию, подобную радиосвязи, но он настолько похож на обычного человека, насколько это возможно с учетом данной истории. Предположим, что эти нейроученые «локально» манипулируют им таким образом, что он предпринимает процесс размышлений, в котором у него возникают и изменяются желания — [манипуляция состоит в том, что] нейроученые производят каждое из этих состояний, момент за моментом. Нейроученые манипулируют им в том числе посредством нажатия на ряд кнопок, до того, как он начнет размышлять о своей ситуации, тем самым каузально воздействуя на его мыслительный процесс таким образом, что он становится

рационально эгоистичным. Действия Плама не являются вынужденными — в том смысле, что он не действует на основе непреодолимых желаний, потому что нейроученые не вызывают в нем желаний такого типа. Также он не размышляет и не действует вразрез со своим характером, поскольку им часто манипулируют в рационально-эгоистической манере [Pereboom 2001, 112–113].

Из этого описания мы можем заключить, что Плам соответствует критериям свободы и ответственности, сформулированным Г. Франкфуртом и Р.Дж. Уоллесом [Frankfurt 1971; Wallace 1994]. Далее Перебум указывает, что Плам умеренно отзывчив к доводам рассудка (*is moderately reasons-responsive*): если бы Плам обнаружил доводы в пользу того, что убийство будет раскрыто, или еще какие-либо доводы, которые побудили бы его воздержаться от убийства, то он воздержался бы. Увы, он не обнаружил таких доводов и убил мисс Уайт. Тем не менее наличие способности прислушиваться к доводам разума демонстрирует, что Плам соответствует и критериям Дж. Фишера и М. Равизы [Fischer, Ravizza 1998].

Перебум утверждает, что Плам₁ соответствует требованиям интеграционистских теорий моральной ответственности, но тем не менее интуитивно не кажется агентом, обладающим моральным статусом. Если он прав, то данный сценарий представляет собой опровержение интеграционистских компатибилистских теорий моральной ответственности.

Конечно, компатибилист может возразить, что между Пламом₁ и обычными детерминированными агентами, отвечающими требованиям интеграционистских теорий, есть большая разница. Ведь в обычном случае нет никаких нейроученых-манипуляторов! Перебуму нужно показать, что между прямой локальной манипуляцией, описанной в Случае 1, и обычным физическим детерминизмом нет такой разницы, которая была достаточным бы основанием для проведения различия между моральным статусом обычного детерминированного агента и моральным статусом манипулируемого детерминированного агента. Для этого Перебум вводит еще два промежуточных случая.

Случай 2. Плам подобен обычному человеку, за исключением того, что он был создан нейроучеными, которые, хотя они и не могут контролировать его напрямую, запрограммировали его таким образом, чтобы при взвешивании мотивов к действию быть часто, но не исключительно рациональным эгоистом. В результате в тех обстоятельствах, в которых он себя обнаруживает в данный момент, он каузально детерминирован предпринять процесс размышления, умеренно отзывчивый к доводам разума (*moderately reasons-responsive process*), и обладать таким набором желаний первого и второго уровня, которые приводят к тому, что он убивает мисс Уайт [Pereboom 2001, 113–114].

Между первым и вторым случаем есть ряд отличий, но эти отличия, по утверждению Перебума, не должны влиять на оценку морального статуса Плама. Какая разница, манипулируют ли нейроученые действиями Плама непосредственно, в режиме реального времени, или они программируют их заранее, еще до рождения Плама? Так или иначе, поведение Плама регулируется факторами, над которыми у него нет контроля, и в этом причина интуитивной убежденности, что как в первом, так и во втором случае, Плам не несет моральной ответственности. Высказав эти соображения, Перебум переходит к описанию третьей ситуации:

Случай 3. Плам — обычный человек, за исключением того, что строгими воспитательными практиками у себя дома и в своем сообществе он детерминирован быть часто, хотя и не всегда, рационально эгоистичным (точно так же эгоистичным, как в случаях 1 и 2). Это воспитание происходило в раннем детстве, так что у Плама не было возможности помешать ему или

изменить те практики, которые детерминировали его характер. В нынешних обстоятельствах Плам тем самым детерминирован предпринять процесс размышления, умеренно отзывчивый к доводам разума, и обладать таким набором желаний первого и второго уровня, которые приводят к тому, что он убивает мисс Уайт [Pereboom 2001, 114].

Интуитивно кажется, что случай 3 отличается от двух предыдущих случаев, так как в нем нет нейроученых, которые манипулировали бы Пламом. В соответствии с критериями интеграционистских теорий и в соответствии с повседневной интуицией кажется, что Плам₃ несет моральную ответственность за свой поступок. Но Перебум полагает, что компатибилисты не имеют права выносить такое суждение о Пламе₃, пока они не продемонстрируют, в чем именно заключаются важные с моральной точки зрения отличия между вторым и третьим случаем: «В обоих случаях Плам удовлетворяет всем критериям каузальных интеграционистских теорий, так что разница в моральной оценке не может быть основана на разнице в соответствии данным критериям. Каузальная детерминация факторами, над которыми у Плама нет контроля, кажется наиболее вероятным обоснованием отсутствия у него моральной ответственности во втором случае, и, кажется, что мы вынуждены сказать, что и в третьем случае он не несет моральной ответственности по той же причине» [Pereboom 2001, 115].

Четвертый случай, который рассматривает Перебум, представляет собой описание обыкновенного детерминированного агента:

Случай 4. Физикалистский детерминизм истинен, а Плам – обычный человек, появившийся на свет и воспитанный при обычных обстоятельствах, который часто, но не всегда, бывает рационально эгоистичен (настолько же, как в случаях 1–3). Убийство Уайт является результатом предпринятого Пламом процесса размышлений, умеренно отзывчивого к доводам разума, Плам проявляет такую же (как Пламы из предыдущих трех случаев. – *М.С.*) организацию желаний первого и второго уровня, и не действует исходя из непреодолимого желания [Pereboom 2001, 115].

Плам соответствует критериям интеграционистских теорий, но, пишет Перебум, у нас нет оснований считать его морально ответственным, если мы не найдем принципиального отличия, которое заблокировало бы перенос суждений об отсутствии моральной ответственности между случаями от первого до четвертого.

Перебум делает следующий вывод: «Лучшее объяснение интуитивной убежденности в том, что Плам не несет моральной ответственности в первых трех случаях, заключается в том, что его действие является результатом детерминистического каузального процесса, вызванного факторами, над которыми у него нет контроля. Поскольку Плам точно так же каузально детерминирован в Случае 4, мы должны заключить, что и здесь Плам по той же причине не несет моральной ответственности. Обобщим: если действие является результатом детерминистического каузального процесса, вызванного факторами, над которыми у агента нет контроля, то агент не является морально ответственным за это действие» [Pereboom 2001, 116].

Поскольку то, что верно относительно Плама в случае 4, верно и относительно любого агента в детерминированном мире, четырехступенчатый аргумент манипуляции, если компатибилистам не удастся найти в нем фатальных недостатков, опровергнет перечисленные Перебумом интеграционистские теории, т.е. несколько наиболее влиятельных версий психологического компатибилизма. В следующем параграфе я проанализирую стратегии защиты со стороны компатибилистов и ответные меры Перебума.

3. «Мягкий ответ» компатибилистов на аргумент манипуляции

В статье «Критика “четырёхступенчатого аргумента” Перебума против компатибилизма» [Mele 2005] американский философ Альфред Миле оспаривает то, что вывод о

несовместимости физического детерминизма и моральной ответственности представляет собой «наилучшее объяснение» того факта, что Плам не кажется нам ответственным в первом и втором случаях манипуляции. Для того чтобы продемонстрировать это, он модифицирует первый случай, называя его «случай 1а». В сценарии Миле все происходит точно так же, как в первом сценарии Перебума, за исключением того, что механизм, посредством которого нейроученые манипулируют поведением Плама, не детерминистический: «Существует очень большая вероятность, что каждое нажатие кнопки приведет к тому результату, которого хотят ученые, но есть и маленькая вероятность того, что каждое нажатие кнопки введет Плама в ступор» [[Mele 2009, 76]. Миле полагает, что Плам, которым манипулируют при помощи недетерминистического механизма, ничуть не более ответственен, чем Плам, которым манипулируют при помощи детерминистического механизма. Следовательно, из первого сценария нельзя сделать вывод о том, что именно детерминизм несовместим с моральной ответственностью.

Между первыми тремя сценариями и последним, четвертым, сценарием, есть большая разница. В двух первых фигурируют манипулирующие поведением Плама нейроученые, в третьем – особые и необычные условия воспитания, а в четвертом сценарии нет ничего, выходящего за рамки нашей повседневной жизни. Перебум настаивает, что мы должны сделать вывод об отсутствии ответственности в четвертом случае, так как он, так же как и предыдущие три случая, детерминистичен. Миле полагает, что из того факта, что все эти сценарии у Перебума являются детерминистскими, компатибилистам не следует делать вывод, что именно детерминизм, а не какие-то иные аспекты описываемых ситуаций, снимают с Плама ответственность, как это утверждает Перебум:

Они должны считать вывод Перебума не более убедительным, чем вывод, что наилучшим объяснением того, почему машина Скарлет была повреждена, является влажность, в следующих трех случаях, когда на эту машину падает влажный объект: (Случай 1) На машину Скарлет упала большая влажная свинцовая труба, в результате машина была повреждена; (Случай 2) На машину Скарлет упал большой влажный гаечный ключ, в результате машина была повреждена; (Случай 3) На машину Скарлет упал большой влажный металлический подсвечник, в результате машина была повреждена. (В каждом случае объект упал с высоты в десять метров.)

<...> В Четвертом случае из хроник Скарлет на ее машину упала большая влажная губка. Пикок делает вывод, что и в этом случае машина Скарлет была повреждена. Но она ошибается. Причиной нанесенных повреждений были твердость и вес падающих объектов, а не их влажность. Подобным же образом, несмотря на то, что говорит Перебум, в описанных им случаях манипуляция, а не детерминистическая каузальность, влияют на интуицию [Mele 2005, 79–80].

Сам Миле считает, что аналогом твердости и веса, т.е. фактором, снимающим моральную ответственность в приведенных Перебумом трех первых случаях, является особенность жизненной истории Плама – тот факт, что у него никогда не было возможности изменить свои убеждения и ценности в зависимости от результатов собственных действий. Эта особенность не распространяется на четвертый случай, следовательно, аргумент не работает, заключает Миле.

Похожую критику аргумента Перебума предлагает и Джон Фишер, который в статье «Ответственность и манипуляция» [Fischer 2004] пишет, что Плам в случае 1 не несет ответственности за убийство Уайт потому, что его ценности, убеждения, желания и характер были сформированы специфическим образом, который не имеет места в случае 4. Критика Миле и Фишера основывается на позиции исторического компатиби-

лизма — отстаиваемой ими версии компатибилизма, в соответствии с которой жизненная история агента имеет непосредственное отношение к его моральному статусу.

Эта критика направлена на то, чтобы продемонстрировать, что Плам₁ из аргумента Перебума не удовлетворяет компатибилистским критериям — между ним и обычными агентами в детерминированном мире есть значительная разница, из-за которой их моральный статус различен. Майкл Маккенна, американский философ компатибилистского направления, в статье «Жесткий ответ на четырехступенчатый аргумент манипуляции Перебума» [McKenna 2008] называет такую стратегию компатибилистской защиты «мягким ответом» (soft-line reply). Суть этой стратегии в том, чтобы показать, что Плам из одного из трех первых сценариев не удовлетворяет компатибилистским критериям моральной ответственности, следовательно, обобщение от морального статуса Пламов 1–3 к моральному статусу Плама₄ неправомерно.

4. «Жесткий ответ» компатибилистов на аргумент манипуляции

В вышеупомянутой статье [McKenna 2008] Маккенна указал, что предложенный Перебумом четырехступенчатый аргумент манипуляции можно рассматривать как специфическую модель аргументации против любых компатибилистских теорий психологического типа. Аргументы такого типа берут за стартовую точку критерии тех или иных компатибилистских теорий и демонстрируют, что этим критериям может соответствовать агент, которым манипулируют таким образом, что он интуитивно кажется лишенным моральной ответственности. Некоторые конкретные аргументы такого рода можно опровергнуть, пользуясь стратегией «мягкого ответа», т.е. показав, что агенты, описанные в них, на самом деле не соответствуют компатибилистским критериям. Но эта стратегия неудовлетворительна, потому что инкомпатибилист может так модифицировать свой аргумент, чтобы учесть данное возражение. По мнению Маккенны, нет принципиальных оснований считать, что невозможно сформулировать такой аргумент манипуляции, в котором агент удовлетворял бы компатибилистским критериям ответственности и при этом интуитивно представлялся бы нам лишенным ответственности, следовательно, любой конкретный мягкий ответ всегда может быть опровергнут инкомпатибилистом.

Поэтому более надежным будет «жесткий ответ» (hard-line reply), который можно рассмотреть как четырехшаговый алгоритм:

1. Рассмотреть данный предложенный инкомпатибилистом аргумент манипуляции. Если агент, описываемый в этом аргументе, очень далек от удовлетворения стандартам компатибилистской теории, отвергнуть этот аргумент.

2. Если агент удовлетворяет компатибилистским критериям, то перейти к пункту три. Если агент совсем немного не дотягивает до того, чтобы удовлетворять компатибилистским критериям, дополнить описание агента таким образом, чтобы обеспечить соответствие, и перейти к пункту 3.

3. Привлечь внимание читателя к сходству между этим манипулируемым агентом и обычным агентом в детерминистском мире.

4. Напомнить читателю о случаях манипуляции в обычной жизни.

В соответствии с этой стратегией, компатибилисту следует признать, что Плам, или некий очень похожий на Плама манипулируемый агент, является подходящим кандидатом на роль морально ответственного агента, и использовать стратегию Перебума в обратном направлении. Вместо того чтобы начинать с локально манипулируемого Плама₁, действительно вызывающего интуицию непригодности на роль того, кто достоин морального осуждения, нам следует начать с Плама₄, и, согласившись с Перебумом в том, что между соседними случаями нет морально значимых отличий, прийти к Пламу₁ и заключить, что и он, несмотря на первоначальные наши интуиции, тоже несет ответственность — ведь между манипуляцией и детерминацией нет существенной разницы, как утверждает сам Перебум. Чтобы убедиться, что детерминированный агент может представляться нам ответственным, Маккенна предлагает подумать о случаях, в которых некий опыт полностью изменяет последующую жизнь агента, предопределяя

его ценности и поведение. Так, пережитая в детстве болезнь и смерть любимого человека может навсегда преобразить представления человека о жизни — так же, как их может преобразить рождение ребенка, религиозное обращение и т.д. Мы не считаем, что агент обязательно должен быть способен выбрать свои ценности для того, чтобы быть способным действовать ответственно, исходя из имеющихся ценностей.

5. Анализ успешности жесткого и мягкого ответов

Степень успешности жесткой и мягкой стратегий компатибилистского ответа является предметом оживленных дискуссий. В статье «Мягкое разрешение четырехступенчатого аргумента Перебума» [Demetriou 2010] Кристин Димитриу предлагает способ демонстрации того, что не только конкретный аргумент Перебума, но и любой аргумент манипуляции, в котором ментальная жизнь агента не играет решающей каузальной роли в его действиях, не будет успешен, так как он лишает агента моральной ответственности, но вовсе не по причине того, что работает подобно детерминизму, якобы не позволяющему агенту контролировать исход своих действий, а потому, что вообще разрушает связь между агентом и действиями, которые он «осуществляет». В сценариях, подобных сценарию с Пламом₁, агент вообще не действует — действуют нейроученые, нажимающие на кнопку. Эта особенность первого сценария не переносима на последующие сценарии, следовательно, полагает Димитриу, обобщенная мягкая стратегия такого рода однозначно опровергает не только аргумент Перебума, но и все аргументы, похожие на него. Жесткая же стратегия, по мнению Димитриу, безуспешна в аргументах манипуляции именно потому, что такие манипулируемые агенты, о которых идет речь, принципиально не удовлетворяют требованиям, постулируемым компатибилистскими теориями моральной ответственности. Таким образом, пункт 3 невозможно выполнить.

Дмитрий Волков в статье «Что доказывают манипуляции с “Аргументом Манипуляции” Д. Перебума?» [Волков 2015] изложил собственную версию критики аргумента манипуляции, которую можно рассматривать как синтез мягкой и жесткой стратегии. Д.Б. Волков пишет, что, несмотря на то, что приведенные Перебумом четыре случая имеют между собой много общего, в них присутствуют и существенные различия, и для того чтобы определить, что именно представляется нам тем решающим фактором, который снимает с Плама моральную ответственность, необходимо действовать, как «экспериментаторы в лаборатории», последовательно внося различные изменения в описываемые сценарии:

В Сценариях аргумента Манипуляций есть несколько ключевых параметров. На первый взгляд это: (1) тип манипуляторов; (2) время манипуляции; (3) компатибилистические критерии свободного агента; (4) характер действия; (5) последовательность сценариев; (6) тип детерминации. Изменяя эти параметры, можно попытаться выяснить, что будет происходить с интуицией о моральной ответственности. Представьте, что каждый из этих параметров — тумблер на панели моральной ответственности. Мастер выключатель выключает все тумблеры. При выключении мастер выключателя, электрическая лампочка гаснет — появляется интуиция, что моральная ответственность исчезает. Следует выяснить, какой из обычных тумблеров отвечает за отключение лампочки, какой из параметров отвечает за интуицию об отсутствии моральной ответственности. Перебум, считает, что этим тумблером является тумблер каузального детерминизма. [Волков 2015, 936].

Последовательно экспериментируя с каждым из тумблеров, Волков приходит к выводу, что ни один из них не является необходимым и достаточным условием для наличия моральной ответственности. Волков полагает, что настоящий источник для приписывания моральной ответственности — это целостность личности, осуществляющей поступок. Наш анализ морального статуса Плама изменится, если дополнить описание первого сценария

поступка Плама сведениями о его последующей жизни — раскаялся ли он в содеянном или, наоборот, не только убил Уайт, но и продолжил дальнейшую аморальную жизнь, полную преступлений. Важно также, продолжали ли нейрочеловеки в дальнейшем воздействовать на Плама. Таким образом, Волков полагает, что аргумент Перебума не достигает своей цели: «Аргумент Манипуляций внушает интуиции об отсутствии моральной ответственности не из-за детерминизма, а из-за информационных лагун, допущенных в мысленном эксперименте. Именно эти лагуны не позволяют судить о Пламе как о самостоятельной ответственной личности. А ведь для моральной ответственности нужен не только виновный поступок и злой умысел, но и целостная аморальная личность, которой можно приписать поступок и когерентную личностную историю. Детерминизм не противоречит существованию таких целостных личностей. А, значит, Аргумент Манипуляций не работает против компатибилизма» [Волков 2015, 941].

Если Волков прав, то в том виде, в котором Перебум сформулировал свой аргумент манипуляции в книге «Жизнь без свободы воли» [Pereboom 2001], этот аргумент не работает, так как не удовлетворяет историческому компатибилистскому критерию ответственности по причине недостаточности описания — мы не считаем Плама ответственным, потому что мы не знаем, станет ли его поступок составной частью его последующей истории и каково будет его отношение к этому поступку. Таким образом, по отношению к данной версии аргумента манипуляции работает мягкий ответ. Однако, если Плам будет описан таким образом, чтобы мы смогли вынести суждение о его будущем характере, то он будет удовлетворять историческому компатибилистскому критерию ответственности и будет, как следует из аргументации Волкова, подходящим кандидатом для приписывания моральной ответственности. Следовательно, если Плам будет описан в соответствии с историческим компатибилистским критерием ответственности, то аргумент будет опровергнут при помощи жесткого ответа.

В 2014 г. Перебум учел возражения, сделанные сторонниками мягкого ответа, в книге «Свобода воли, деятельность и смысл жизни» [Pereboom 2014] и сформулировал новую версию четырех историй о профессоре Пламе. Новый Плам, который будет подробно проанализирован в следующем параграфе, кажется, действительно соответствует критериям ответственности, выдвигаемым критиками предыдущей версии аргумента.

Успешность жесткого ответа на данный момент тоже спорна. Перебум в статье «Жесткий ответ на аргумент манипуляции» [Pereboom 2008] оспорил правомерность диалектической стратегии Маккенны. Перебум пишет, что жесткий ответ подразумевает компатибилистский подход к ответу на вопрос об истинности двух положений:

1. Какую изначальную установку рационально занять относительно моральной ответственности обычных детерминированных агентов в повседневном мире?
2. Какова степень убедительности случаев детерминистической манипуляции (подобных случаям 1 и 2)?

Если читатель, впервые встречающийся с аргументом манипуляции, уже занимает компатибилистскую позицию и не считает, что детерминизм представляет собой хотя бы кажущуюся угрозу для моральной ответственности, то жесткий ответ сработает. Однако аргумент манипуляции нацелен именно на такого читателя, который еще не имеет укоренившейся компатибилистской установки, потому что его аргументативная сила основана на базовой, претеоретической интуиции. Что более убедительно — последовательное рассмотрение аналогии случаев, находящихся между манипуляцией и детерминизмом, по результатам которого читатель решает, что детерминизм похож на манипуляцию в достаточной степени для того, чтобы быть причиной признания неответственности агента, или наоборот, последовательное рассмотрение аналогии случаев, находящихся между детерминизмом и манипуляцией, по результатам которого читатель решает, что детерминизм похож на манипуляцию в достаточной степени для того, чтобы признать даже искусственных агентов, манипулируемых нейрочеловеками, достойными кандидатами на статус морально ответственных агентов? Перебум полагает, что вторая опция убедительна только для того, кто уже является компатибилистом, так как

подобное следствие для незаинтересованного наблюдателя представляет собой опровержение компатибилистской теории.

Маккенна возражает на это, что значимость интуиций, вызываемых фантастическими сценариями, подобными сценарию с нейроучеными и радиоуправляемым Пламом, не так велика, как значимость сценариев из повседневной жизни, поскольку наши моральные интуиции вырабатываются именно для оценки обыденных ситуаций [McKenna 2014], однако вопрос о том, что является более контринтуитивным, компатибилистская интуиция или фантастические примеры, трудно решить однозначно. Таким образом, на настоящий момент успешность стратегии жесткого ответа оказывается под вопросом: те, кто принимают компатибилистские критерии, считают аргумент опровергнутым, те, кто не принимают их, считают аргумент успешным [Mickelson 2016]. Дискуссии продолжаются с неослабевающей силой, в том числе в отечественной философии — например, в коллективной статье «Несостоятельность аргументов манипуляции и исчезающего агента для проблемы свободы воли и моральной ответственности» [Беседин и др. 2017].

6. Создание нового Плама

В книге «Свобода воли, деятельность и смысл жизни» [Pereboom 2014] Перебум модифицирует первый случай своего аргумента манипуляции, чтобы опровергнуть критику со стороны Миле, Фишера и Деметриу, проанализированную выше. Теперь про Плама не говорится, что он был создан нейроучеными, вызывающими каждое его действие нажатием кнопки. О Новом Пламе нам известно следующее:

Случай 1. Команда нейроученых обладает возможностью манипулировать нейрональными состояниями Плама в любой момент при помощи использования технологии, похожей на радио. В этом конкретном случае они манипулируют им путем нажатия на кнопку до того момента, как он начнет рассуждать о сложившейся ситуации, и они знают, что это нажатие вызовет в нем нейрональное состояние, представляющее собой эгоистичный мыслительный процесс, который, как известно нейроученым, детерминированно приведет к решению Плама убить Уайт. Плам не убил бы Уайт, если бы нейроученые не вмешались, поскольку в этом случае его размышления не были бы в достаточной степени эгоистическими для того, чтобы привести к этому решению [Pereboom 2014, 76].

Далее описание случая Нового Плама₁ аналогично описанию случая Плама₁ — его желание убить Уайт не является навязчивым, соответствует его желаниям второго порядка и т.д. В результате всей этой операции Плам убивает Уайт — несет ли он моральную ответственность за свой поступок? Перебум полагает, что интуиция читателя должна склониться к отрицательному ответу. После этого Перебум излагает три последующих случая манипуляции, аналогичные тем, что были изложены в предыдущей версии аргумента, и утверждает, что компатибилисты должны указать на принципиальное различие между этими случаями.

Можно возразить, что в описании Нового Плама₁ по-прежнему недостаточно информации о том, насколько вмешательство нейроученых сказывается на характере нашего героя в дальнейшем — это будет возражение в стиле исторической версии мягкого ответа. (Возражение сделано Дмитрием Волковым в частной корреспонденции.) Допустим, однако, что оно никак на нем не скажется — вероятно, именно такое допущение в наибольшей степени соответствует целям Перебума, который хочет продемонстрировать, что новый аргумент манипуляции свободен от проблем, стоявших перед старым аргументом.

Однако в этом случае Новый Плам₁, как мне кажется, уже настолько похож на обычного агента, что жесткая стратегия ответа становится успешной. Мы не можем утверждать, что сам по себе детерминизм лишает агентов ответственности — ведь в

таким случае мы будем заранее подразумевать, что компатибилизм ложен. Такая тактика не годится для опровержения детерминизма. Рассмотрим в таком случае, есть ли у нас основания считать Нового Плама₁ ответственным, если убрать за скобки детерминизм. На первый взгляд, он кажется ответственным. В самом деле, Новый Плам₁ не робот, он во всем подобен обычным агентам в детерминированном мире, за исключением того, что им в данном случае манипулировали нейрочуждые. Мы можем решить, что он является ответственным, пока не узнаем, что на самом деле ответственные нейрочуждые. Но ведь в третьем и четвертом случаях нет нейрочуждых, и Плам ведет себя в соответствии с компатибилистскими критериями ответственности, интуитивно представляющимися релевантными для вынесения суждения о моральной ответственности. Это уже дает компатибилистам возможность утверждать, что новый аргумент Перебума нарушает принцип отсутствия морально значимых различий между случаями. Инкомпатибилист будет настаивать, что нейрочуждые лишь подчеркивают отсутствие контроля, а не лишают агента контроля, но компатибилист, пользуясь аргументами в стиле Маккенны [McKenna 2014], имеет право указать на тот факт, что повседневные интуиции могут быть легко отвлечены от релевантных соображений о соответствии компатибилистским критериям нерелевантными, но пугающими соображениями о всевластии нейрочуждых, сразу вызывающих в воображении образы антиутопии и бессилия человека перед продвинутыми технологиями.

Еще более убедительным представляется следующее соображение. Предположим, что Пламом не манипулировали нейрочуждые путем воздействия на его мозг, а манипулировал некий человек, хорошо знающий его характер, мировоззрение и цели. Этот человек поговорил с Пламом и заронил в нем желание убить Уайт. Сам человек не является нейрочуждым и может лишь надеяться на то, что его коварный замысел удастся. Однако нейрочуждые, отслеживающие ментальную жизнь Плама, могут с уверенностью предсказать, что Плам действительно убьет Уайт. Нейрочуждые не вмешиваются, и Плам убивает Уайт. Назовем этот случай случаем «Плам без кнопки». Несет ли «Плам без кнопки» ответственность за свой поступок? Учитывая, что нейрочуждые ничего не делали, я не вижу оснований снимать с Плама ответственность — ведь детерминизм не может быть таким основанием, как уже было отмечено. Следовательно, компатибилисты имеют полное право считать Нового Плама₁ ответственным за свой поступок. Можно возразить, что этот случай не аналогичен случаю Нового Плама₁, в котором нейрочуждые вмешались в процесс принятия решения, так сказать, «изнутри», воздействуя на его мозг. Манипуляция путем разговора не то же самое, что манипуляция на нейрональном уровне. (Возражение сделано Дмитрием Волковым в частной корреспонденции.) Но такое возражение не может быть сделано сторонником аргумента манипуляции, так как оно покажет, что даже новая версия аргумента манипуляции уязвима по отношению к мягкой стратегии защиты — ведь успех аргумента манипуляции зависит от того, что мы признаем, что между тем случаем, в котором есть нейрочуждые, воздействующие на мозг, и тем случаем, в котором их нет, отсутствует морально значимая разница!

7. Заключение

Изначально аргумент манипуляции был сформулирован Дерком Перебумом, чтобы при помощи повседневных интуиций продемонстрировать несовместимость физического детерминизма и моральной ответственности. В ответ на мягкую стратегию критики со стороны ряда философов-компатибилистов, указавших на то, что профессор Плам из первого случая не соответствовал компатибилистским критериям ответственности, Перебум был вынужден сформулировать новую версию аргумента. Однако новая версия, как я показала в статье, уязвима перед жесткой стратегией ответа. Таким образом, обе версии четырехступенчатого аргумента манипуляции Дерка Перебума оказываются опровергнуты.

Источники— Primary Sources

- Demetriou, Kristin (2010) “The Soft-Line Solution to Pereboom’s Four-Case Argument”, *Australasian Journal of Philosophy*, Vol. 88, № 4, pp. 595–617.
- Dennett, Daniel (2003) *Freedom Evolves*, Viking, New York.
- Fischer, John Martin (2004) “Responsibility and Manipulation”, *The Journal of Ethics*, Vol. 8. № 2, pp. 145–177.
- Fischer, John Martin (2006) *My way: essays on moral responsibility*, Oxford University Press, Oxford.
- Fischer, John Martin, and Mark Ravizza (1998) *Responsibility and Control: A Theory of Moral Responsibility*, Cambridge University Press, Cambridge, New York.
- Frankfurt, Harry (1971) “Freedom of the Will and the Concept of a Person”, *The Journal of Philosophy*, Vol. 68, № 1, pp. 5–20.
- McKenna, Michael (2008) “A Hard-Line Reply to Pereboom’s Four-Case Manipulation Argument”, *Philosophy and Phenomenological Research*, Vol. 77, № 1, pp. 142–159.
- McKenna, Michael (2014) “Resisting the Manipulation Argument: A Hard-Liner Takes It on the Chin”, *Philosophy and Phenomenological Research*, Vol. 89, № 2, pp. 467–484.
- Mele, Alfred (2005) ‘A Critique of Pereboom’s “Four-Case Argument” for Incompatibilism’, *Analysis*, Vol. 65, № 1, pp. 75–80.
- Mele, Alfred (2009) *Effective Intentions: The Power of Conscious Will*, Oxford University Press, Oxford, New York.
- Mickelson, Kristin (2016) “The Manipulation Argument”, Griffith, Meghan ed., *Routledge Companion to Free Will*, Routledge-Taylor & Francis, New York, pp. 166–178.
- Pereboom, Derk (2001) *Living without Free Will*, Cambridge University Press, Cambridge, U.K., New York.
- Pereboom, Derk (2005) “Defending Hard Incompatibilism”, *Midwest Studies in Philosophy*, Vol. 29, № 1, pp. 228–247.
- Pereboom, Derk (2008) “A Hard-Line Reply to the Multiple-Case Manipulation Argument”, *Philosophy and Phenomenological Research*, Vol. 77, № 1, pp. 160–170.
- Pereboom, Derk (2014) *Free Will, Agency, and Meaning in Life*, Oxford University Press, Oxford, New York.
- Strawson, Peter (1962) “Freedom and Resentment”, *Proceedings of the British Academy*, Vol. 48, pp. 1–25.
- Wallace, R. Jay (1994) *Responsibility and the Moral Sentiments*, Harvard University Press, Cambridge, Mass.

Ссылки — References in Russian

- Беседин и др. 2017 — *Беседин А.П. и др.* Несостоятельность аргументов манипуляции и исчезающего агента для проблемы свободы воли и моральной ответственности // *Вопросы Философии*. 2017. № 6. С. 29–38.
- Васильев 2017 — *Васильев В.В.* В защиту классического компатибилизма: Эссе о свободе воли. М.: ЛЕНАНД: URSS, 2017.
- Волков 2015 — *Волков Д.Б.* Что доказывают манипуляции с «Аргументом Манипуляции» Д. Перейбума? // *Философия и Культура*. 2015. Т. 6. С. 933–942.

Voprosy Filosofii. 2018. Vol. 12. P. 185–196

A Compatibilist Refutation of the 'Four Case' Manipulation Argument*

Maria A. Sekatskaya

Pereboom's 'Four Case' manipulation argument is intended to show that compatibilist criteria of moral responsibility are not sufficient, because there can be agents who meet these criteria, but lack moral responsibility. This argument is one of the most influential incompatibilist arguments in contemporary debate on compatibility of moral responsibility and physical determinism.

I review the structure of the argument, the criticism of the argument by several compatibilist philosophers, and Pereboom's defense against this criticism. After that I show that two

* Research has been conducted with financial support from Russian Science Foundation grant (project No. 18-18-00222).

different strategies of compatibilist criticism, the so called ‘soft’ and ‘hard’ replies, make the premises and the conclusion of the argument less plausible, although neither of them is sufficient to refute the argument. In the final part of the article I demonstrate how the combination of these strategies refutes Pereboom’s ‘Four Case’ manipulation argument.

KEY WORDS ‘Four Case’ manipulation argument, Derk Pereboom, compatibilism, moral responsibility.

SEKATSKAYA Maria A. – CSc in Philosophy, senior lecturer, Institute of Philosophy, Saint-Petersburg State University.

maria.sekatskaya@gmail.com

Received at June, 11 2018.

Citation: Sekatskaya, Maria A. (2018) ‘A Compatibilist Refutation of the “Four Case” Manipulation Argument’, *Voprosy Filosofii*, Vol. 11 (2018), pp. 185–196.

DOI: 10.31857/S004287440002600-5

References

Besedin, Artem P., Volkov, Dmitry B., Kuznetsov, Anton V., Mertsalov, Andrey V., Sekatskaya, Maria A. (2017) “Failure of the Manipulation Arguments and Disappearing Agent to the Problem of Free Will and Moral Responsibility”, *Voprosy Filosofii*, Vol. 6 (2017), pp. 29–38 (in Russian).

Vasilyev, Vadim V. (2017) *In Defense of Classical Compatibilism: An Essay on Free Will*, LENAND, Moscow (in Russian).

Volkov, Dmitry B. (2015) “What Do the Manipulations with the Manipulation Argument Show?”, *Philosophy and Culture*, Vol. 6, pp. 933–942 (in Russian).