Этическая мысль
2021. Т. 21. № 1. С. 40–49
УДК 17.022; 17.026

Ethical Thought
2021, Vol. 21, No. 1, pp. 40–49
DOI: 10.21146/2074-4870-2021-21-1-40-49

Б.В. Фауль

Экстернализм в отношении моральной ответственности: модификация мысленного эксперимента А. Мелe*

Богдан Владимирович Фауль – аспирант, лаборант-исследователь. Санкт-Петербургский Государственный Университет. Институт Философии. Российская Федерация, 199034, г. Санкт-Петербург, Менделеевская линия, д. 5; e-mail: faulbogdan@gmail.com

В данной статье автор модифицирует мысленный эксперимент, проведенный А. Мелe в пользу экстернализма в отношении моральной ответственности. С его точки зрения, история агента частично определяет то, несет ли агент моральную ответственность за конкретные действия, либо же последствия действий. В оригинальном мысленном эксперименте рассмотрена ситуация, в которой личность не несет моральной ответственности за убийство по причине манипуляции, то есть по внешней по отношению к агенту причине. Теория А. Мелe была подвергнута критике со стороны А.В. Мерцалова, Д.Б. Волкова и В.В. Васильева на семинаре, организованным Московским Центом Исследования Сознания. Аргументация против концепции А. Мелe имела следующую форму: А. Мелe не демонстрирует, что историческое объяснение является наилучшим, поскольку имеется не менее убедительные конкурирующие объяснения, которые исключают экстернализм А. Мелe. В статье автор эксплицирует и анализирует объяснения, предложенные философами из Московского Центра Исследования Сознания: объяснение от тождества личности, объяснение от самоидентификации, объяснение от условия знания, объяснение от будущих состояний. Несмотря на то, что эти объяснения применимы к оригинальному мысленному эксперименту А. Мелe, они не способны объяснить отсутствие моральной ответственности в модифицированном мысленном эксперименте, предлагаемом автором статьи: объяснения от тождества личности и самоидентификации исключаются посредством постепенного изменения агентной структуры личности; объяснение от условий знания опровергается посредством

* Исследование выполнено при финансовой поддержке РФФИ, проект № 19-311-90083 «Метафизика существования личности во времени». Funding: the reported study was founded by RFBR, according to the research project No. 19-311-90083 «Metaphysics of personal identity over time».

включения знания о манипуляции в условия мысленного эксперимента; объяснение от будущих состояний исключается посредством удаления из мысленного эксперимента каких-либо состояний агента после убийства.

***Ключевые слова*:** свобода воли, моральная ответственность, тождество личности

В 2019 г. один из лидирующих мировых специалистов по проблеме свободы воли Альфред Меле написал книгу, посвященную моральной ответственности, заглавие которой можно перевести как «Агенты под манипуляцией: окно в моральную ответственность»[1]. В данный момент на русском языке нет публикаций, посвященных именно этой работе, однако она активно обсуждается в кругах русскоязычных философов. В недавнем времени один из центральных элементов работы Меле – интерпретация мысленного эксперимента с Чаком и Салли – подвергся разносторонней критике, которая была предложена на семинаре Московского Центра Исследования Сознания 14 апреля 2020 г. Однако, несмотря на оригинальность предложенной критики, она кажется недостаточной для отвержения позиции Меле. В данной статье я выступлю в защиту американского философа и предложу модифицированную версию критикуемого мысленного эксперимента, которая учитывает и отвергает возражения, озвученные А.В. Мерцаловым, Д.Б Волковым и В.В. Васильевым. Помимо упомянутых философов в семинаре участвовали и другие исследователи, однако их критика касалась других аспектов работы Меле, на которых я не имею возможности сконцентрироваться. Рассуждения в этой статье наметят теоретическое пространство для плодотворного обсуждения рассматриваемой темы в будущем.

Сначала я кратко опишу мысленный эксперимент и те цели, которых пытается достичь Меле. После этого я проведу реконструкцию возражений, которые были предложены коллегами на семинаре. Я сконцентрируюсь только на определенном *типе* возражений, структуру которого я кратко опишу. После этого я предложу модификацию мысленного эксперимента Меле, которая учитывает и отвергает возражения коллег.

Мысленный эксперимент

Меле защищает тезис, что моральная ответственность необходимо зависит от *истории агента*. Поскольку история агента является внешней по отношению к агенту, данная позиция является разновидностью *экстернализма в отношении моральной ответственности*. Тезис, который А. Меле пытается защитить, можно сформулировать следующим образом:

Т: История агента частично *определяет* то, является ли агент ответственным за конкретные действия или нет.

Истинность «Т» автоматически влечет истинность экстернализма в отношении моральной ответственности. Отмечу, что аргументация в пользу «Т»

[1] *Mele A.R.* Manipulated Agents: A Window to Moral Responsibility. Oxford; New York, 2019.

уже давно развивается Альфредом Меле, самое позднее – с 2006 г.[2] В новой же работе в защиту «Т» предлагается мысленный эксперимент с Чаком и Салли[3]. Суть мысленного эксперимента можно изложить так: Чак является плохим человеком и склонен убивать. Салли является святой, и убивать никогда не хотела, и даже не могла захотеть. Однако в одну из ночей команда психологов заменяет жизненные ценности Салли жизненными ценностями, которые присущи Чаку. После того, как Салли просыпается, она хочет убить своего соседа и, в конечном счете, убивает его. То же самое делает Чак, убивая другого человека. В следующую ночь психологи возвращают Салли ее «святые жизненные ценности». Несет ли Салли моральную ответственность за убийство?

Меле исходит из очевидности того, что Чак ответственен за убийство, а Салли – нет, поскольку она находилась под манипуляцией. Меле демонстрирует, что в данной ситуации выполняются другие компатибилистские условия моральной ответственности за убийство: Салли и Чак хотят убить человека, разделяют схожий набор ценностей, считают себя источниками своих действий и пр. В результате, единственная морально релевантная разница между ними – это их история, где история Салли содержит манипуляцию, а история Чака – нет. Поскольку это единственное релевантное различие между этими случаями, оно и *объясняет то, почему Чак ответственен за убийство, а Салли – нет*. Следовательно, «Т» истинно, как и экстернализм в отношении моральной ответственности.

Данная интерпретация мысленного эксперимента не показалась убедительной для некоторых коллег из МЦИС. Они предполагают, что помимо *истории* Чака и Салли существуют и другие, *не менее убедительные объяснения* разницы в моральной ответственности за похожие действия. Коллеги озвучили следующие конкурирующие объяснения:

(а) Объяснение от тождества личности
(б) Объяснение от самоидентификации
(в) Объяснение от условия знания
(г) Объяснение от будущих состояний.

Назовем то объяснение, которое предлагает Меле – *историческим объяснением*. Таким образом, структуру рассматриваемых возражений можно описать следующим образом*: А. Меле не демонстрирует, что историческое объяснение является наилучшим, поскольку имеется не менее убедительное конкурирующие объяснение (а)/(б)/(в)/(г). Конкурирующее объяснение исключает историческое объяснение, следовательно, «Т» ложно.*

Моя цель в данной работе скромна, я хочу сконцентрироваться лишь на этом возражении и продемонстрировать, что мысленный эксперимент А. Меле можно модифицировать так, что все конкурирующие объяснения (а), (б), (в) и (г) будут исключены. Это продемонстрирует, что возражения, которые имеют указанную структуру, и объяснения, которые сводятся к (а)/(б)/(в)/(г) – несостоятельны. После рассмотрения всех конкурирующих объяснений по отдельности, я предложу модифицированный мысленный

---

2 *Mele A.R.* Free Will and Luck. Oxford; New York, 2006. P. 167–172.

3 *Mele A.R.* Manipulated Agents. P. 19–22.

эксперимент и продемонстрирую, каким образом этим экспериментом исключаются конкурирующие объяснения. Еще раз отмечу, что данная работа претендует лишь на предварительный анализ.

(а) Объяснение от тождества личности

Объяснения (а) и (б) были выдвинуты А.В. Мерцаловым. Сконцентрируемся на реконструкции (а). На первый взгляд, кажется, что в тот момент, когда манипуляторы заменяют агентную структуру Салли и делают ее ценностным двойником Чака, Салли перестает быть тем человеком, которым она была. Иными словами, из-за вмешательства манипуляторов нарушается важное условие тождества личности – *психологическая связанность*. Под психологической связанностью я буду иметь в виду конкретный вид психологической связанности, которую предполагает Мерцалов: связанность *черт характера* и *ценностей*. Иными словами, теория Мерцалова, если я корректно понимаю его позицию, имеет следующий вид: X тождественен Y в том случае, если *ценности* и *черты характеров* X и Y находятся в отношении R. Я предполагаю, что в его подходе личность способна пережить *изменение ценностей и черт характера*. Отношение R, таким образом, не является отношением нумерического или качественного тождества, но является более «мягким» типом отношения – *отношением связанности*. Теория Мерцалова, таким образом, представляет собой разновидность теории психологической связанности с той разницей, что в отношении *связанности* должны находиться определенные психологические черты X и Y: *ценности X и Y и черты характеров X и Y*[4].

Таким образом, у нас имеется две Салли: Салли(1) и Салли(2). Салли(1) является святой, а Салли(2) является ценностным двойником Чака. Салли(2) может убить, а Салли(1) не может. Далее манипуляторы снова возвращают «личность Салли», назовем ее Салли(3). Тождественны ли Салли(1) и Салли(3)? Это, в действительности, не принципиально, поскольку суть возражения следующая: Салли(3) не может нести ответственность за Салли(2), поскольку это разные личности. Личности несут моральную ответственность только за самих себя. Если это так, это могло бы объяснить то, почему мы считаем, что Салли не несет моральной ответственности за то, что убила соседа. Таким образом, нарушение отношения тождества между Салли(2) и Салли(3) является *конкурирующим объяснением того*, почему мы считаем, что Салли(3) не несет моральной ответственности за убийство.

---

[4] Если в теории Мерцалова отношение R – это отношение *тождества*, то мы умираем намного чаще, чем нам кажется: жизнь prima facie одной личности в действительности является жизнью *нескольких* личностей. С точки зрения этого подхода вы можете умереть в любой момент, изменив, например, некоторые важные для вас убеждения. Один из рецензентов справедливо указал на то, что теория Мерцалова может оказаться именно такой теорией. Если это так, то критика Меле Мерцаловым зависит от очень спорных (как мне представляется в данный момент) предположений в дискуссии о тождестве личности. К сожалению, я не имею достаточно пространства для рассмотрения теории, в которой R – это отношение тождества. Критика, высказанная в данной статье, основана на предположении, что R – это *отношение связанности*.

Данное объяснение основывается на предположении истинности того, что психологическая связанность является *необходимым условием тождества личности*. Получается, что такое объяснение неэффективно для тех теорий, в которых отрицается необходимость *психологической связанности* для диахронического тождества. Однако данное возражение можно модифицировать.

Некоторые философы проводят различие между *отношением тождества* и *отношением значимости*[5]. Если возможно диахроническое тождество личности, и тождество транзитивно, то одна личность в момент времени Т1 может быть тождественна только одной личности в момент времени Т2. Отношение значимости может быть устроено совсем иначе, оно включает в себя:

- *Отношение моральной ответственности* – если А в Т1 связан отношением моральной ответственности с Б в Т2, то Б несет моральную ответственность за те действия или последствия действий, за которые А несет моральную ответственность.
- *Психологическое отношение* – если А в Т1 связан психологическим отношением с Б в Т2, то Б радуется или сожалеет о том, что претерпел или совершил А таким образом, будто Б и есть А.
- *Прагматическое отношение* – если А в Т1 связан прагматическим отношением с Б в Т2, то А беспокоится о Б таким образом, будто А и есть Б.

К сожалению, в современной литературе суть этого отношения не прояснена однозначно и эти формулировки – это лучшее, что я смог сделать для этой работы. Данное различение кажется мне крайне продуктивным, поскольку позволяет перевести дискуссию в менее метафизически нагруженную область. Интересно, что даже у А. Меле есть раздел работы, в которой он концентрируется на тождестве личности (вслед за М. Варгасом[6]), хотя в этом нет необходимости.

Является ли отношение значимости зависимым от отношения тождества? Возможно, но не обязательно. Например, при случае рассечения личности на две части, две новые личности могут нести моральную ответственность за поступки или последствия поступков исходной личности. Это означает, что отношение значимости *может* являться отношением *одного ко многим*, тогда как *отношение тождества – это отношение одного к одному с необходимостью*.

Все эти дополнительные пояснения нужны для того, чтобы можно было переформулировать возражение с использованием *отношения значимости*, а не *отношения тождества*, делая возражение применимым к большему количеству теорий тождества личности. Необходимость психологической связанности для отношения тождества не является настолько общепринятым среди современных философов, как необходимость психологической связанности для *отношения значимости*. Таким образом, имеется более универсальное конкурентное объяснение:

Быть может Салли(1), Салли(2) и Салли(3) не находятся в отношении тождества. Однако это совершенно не принципиально, поскольку нарушается куда более важное для моральной ответственности отношение – *отношение*

---

5 *Sider T.* Four-Dimensionalism: An Ontology of Persistence and Time. Oxford, 2001. P. 144.

6 *Vargas M.* Building Better Beings: A Theory of Moral Responsibility. Oxford; New York, 2013. P. 300.

значимости. То есть по той причине, что была нарушена психологическая связанность (а следовательно, и нарушено отношение значимости), Салли(3) не несет моральной ответственности за Салли(2), вне зависимости от того, являются ли Салли(2) и Салли(3) одной и той же личностью, или нет.

(б) Объяснение от самоидентификации

С точки зрения этого конкурирующего объяснения, если бы мы спросили Салли(3), считает ли она сама, что она была Салли(2) – то есть той Салли, которая убила человека – то она ответила бы, что не считает. То есть, иными словами, с точки зрения *самой* Салли ее психологическая связанность была нарушена. Таким образом, это возражение концептуально связано с предыдущим модифицированным возражением. В этом случае, *самоидентификация* является необходимой для того, чтобы между двумя агентами А и Б имелось *отношение значимости*. Получается, что невыполнение условия самоидентификации влечет нарушение связанности отношением значимости, и мы снова получаем конкурирующее объяснение.

(в) Объяснение от условий знания

Данное возражение было выдвинуто В.В. Васильевым. Оно предполагает, насколько я могу судить, что Салли должна *знать*, манипулируют ли ею или нет. Для демонстрации этой идеи предлагался следующий мысленный эксперимент: предположим, что ученые запрограммировали Салли таким образом, что при попадании в нее спиртного она захотела убивать. Этой же ночью ученые делают Салли инъекцию спирта. На утро Салли сообщают, что если сегодня она увидит человека, то захочет его убить и, скорее всего, убьет, поскольку ее запрограммировали определенным образом. Салли проигнорировала это и вышла в город, после чего встретила человека на улице и убила его. Данный мысленный эксперимент предполагает, что несмотря на манипуляцию и на определенную каузальную историю, Салли все равно ответственна за смерть человека по той причине, что она *знала* об этой манипуляции. Таким образом, в исходном мысленном эксперименте Меле мы не считаем Салли морально ответственной по той причине, что *она не знает о манипуляции*. Иными словами, не выполняется необходимый критерий моральной ответственности – *достаточное знание обстоятельств совершения действия*, что объясняет то, почему в мысленном эксперименте Меле Салли не несет моральной ответственности.

(г) Объяснение от будущих состояний

Последнее объяснение было предложено Д.Б. Волковым. Реконструкция данного объяснения представляется наиболее сложной. С точки зрения Меле предыдущая история Салли объясняет, почему Салли не ответственна за убийство, а Чак ответственен, хотя они и являются ценностными двойниками. Д.Б. Волков рассматривает следующую теоретическую возможность: *будущие*

состояния Салли и Чака могли бы объяснить причину, по которой имеется разница в моральной ответственности между Чаком и Салли. Чак *сохранит* свои ценности, тогда как манипуляторы вернут Салли ее исходные ценности. Получается, что имеется психологическая разница между Салли *во время* убийства и Салли *после* убийства, что, на первый взгляд, идентично ситуации наличия разницы между Салли *до* убийства и Салли *во время* убийства. Кажется, что нет хороших причин считать, что *объяснение из будущего* хуже, чем *объяснение из прошлого*. Более того, если манипуляторы не возвращают Салли ее исходное «святое» состояние, то, как считает Волков, корректно говорить о моральной ответственности Салли за убийство соседа, когда она является ценностным двойником Чака. Соответственно, у нас в очередной раз имеется конкурирующее объяснение.

Модификация мысленного эксперимента

Чак является убийцей, Салли имеет совершенные моральные качества. Манипуляторы решили сделать из Салли ценностного двойника Чака и заставить ее убить соседа. Однако в этот раз они делают это иначе: на протяжении двух лет каждую ночь они меняли психологию Салли незначительным образом. Более того, Салли об этом знала в подробностях и даже знала, какие конкретные фрагменты ее психологического устройства подвергались изменению, однако она этого не хотела и пыталась противиться этим изменениям. Через два года Салли стала ценностным двойником Чака и совершила убийство. Салли была ценностным двойником Чака ровно один день. По случайному стечению обстоятельств этой ночью вселенная взрывается и перестает существовать.

Несет ли Салли моральную ответственность за убийство человека? Я считаю, что нет. В конце концов она была святой, и если бы не действия манипуляторов, то она бы никогда не убила. Мне кажется, что интуиции людей, в целом, будут против того, что Салли ответственна за убийство. Это связано с тем, что манипуляция проводилась против ее воли, фундаментальные черты ее характера подвергались *насильственному изменению*. Это отличается от случая, в котором у человека постепенно меняется структура личности под воздействием факторов, которые находятся *в пределах контроля личности и проистекают из ее воли*. Например, когда человек медленно становится алкоголиком, он принимает решения «выпить» в конкретные моменты времени, и до определенной стадии он способен *контролировать* эти действия. Изменение фундаментальной структуры его личности, таким образом, частично основано на его волевых актах и лишь *отчасти имеет манипулятивный характер*. Мы склонны наделять алкоголика ответственностью за его действия именно по этой причине: необходимым каузальным условием «алкоголической» структуры личности были *свободные волевые акты* этой личности. В случае с Салли все происходило *против ее воли*, а потому она не несет ответственности за убийство. Иными словами, пример с алкоголиком – это пример *частичной* манипуляции, тогда как пример с Салли – это пример *тотальной* манипуляции.

Каким образом мы можем это объяснить? Давайте рассмотрим альтернативные объяснения и продемонстрируем причины, по которым они невозможны

в новом мысленном эксперименте. Сначала я начну с объяснений (г) и (в), а потом перейду к объяснениям (а) и (б).

Очевидно, что объяснение из будущего в прошлое в данной ситуации невозможно, поскольку никаких будущих состояний у Салли нет – будущего попросту не существует. К тому же, условия *знания* полностью выполняются: Салли знает о том, что ею манипулируют, однако все равно она не несет моральной ответственности в тот день, когда она становится ценностным двойником Чака. Сконцентрируемся на этом подробнее.

Объяснение Волкова, как минимум, не является универсальным, поскольку возможны ситуации, в которых данное объяснение невозможно в силу того, что будущего не существует, и нет тех сущностей и метафизических отношений, которые должны иметься в объяснении от будущего. Если будущего нет, то нет никакого объяснения того, почему Салли не ответственна за убийство, кроме *исторического объяснения*. Если бы объяснение Волкова было универсальным и исчерпывающим, то отсутствие будущего создало бы в реальности ситуацию *неопределенности* того, является ли Салли ответственной за убийство соседа или нет, однако это не так – Салли *не несет моральной ответственности* за то, что убила соседа. Получается, что объяснение от будущих состояний не отвергает «Т».

С точки зрения Васильева, насколько я могу понять, в оригинальном мысленном эксперименте *не* выполняется *условие знания*, что объясняет то, почему Салли не несет моральную ответственность за убийство. Если это так, то представляется верным, что *если бы она знала о манипуляции, то она несла бы моральную ответственность за убийство*. Однако в модифицированном мысленном эксперименте мы видим, что это не так. Салли знает, что ею манипулируют, но ответственности все равно не несет. Знает Салли о манипуляции или нет – она все равно не несет моральной ответственности за убийство. Это делает критерий знания нерелевантным по крайней мере в ситуации модифицированного мысленного эксперимента, что делает знание о манипуляции *не необходимым для объяснения*. Недостаточным для объяснения знание о манипуляции является по той причине, что оно не может существовать *независимо от наличия самой манипуляции*. Однако манипуляция – это как раз и есть то, на чем основывается историческое объяснение. Получается, что критерий знания о манипуляции всегда идет в одном комплекте с историческим объяснением, либо же не присутствует в объяснении вообще. Выходит, что историческое вмешательство манипуляторов всегда релевантно даже при условии истинности критерия, предложенного Васильевым[7]. В результате, «Т» остается истинным.

---

[7] Хотелось бы обратить внимание читателя на то, что в данной аргументации я не демонстрирую, что критерий знания всегда недостаточен для объяснения. Речь идет о критерии «знания о манипуляции», где *знание* о манипуляции не может существовать *без самой манипуляции*, поскольку *знание*, что Х, предполагает *истинность* Х. Таким образом, критерий знания о манипуляции может быть «добавлен» к объяснению того, почему Салли не несет моральной ответственности. Однако этот критерий сам по себе не исключает того объяснения, которое предлагает Меле.

Рассмотрим объяснения (а) и (б), предложенные Мерцаловым. С этого момента под возражением (а) я буду понимать модифицированное возражение через концепт *отношения значимости,* которое я назову (а*). Таким образом, (а*) предполагает следующее: тот факт, что Салли не несет моральной ответственности объясняется *нарушением отношения значимости*. Однако в данном мысленном эксперименте отношение значимости не нарушается, поскольку манипуляция совершается *слишком медленно* для нарушения психологической связанности. Этот аргумент апеллирует к нашим *интуициям*: изменения являются слишком долгими для того, чтобы нарушалась психологическая связанность. Однако есть еще один аргумент, основанный на *метафизических* соображениях.

Если Салли(1) не является психологически связанной (в том смысле, в котором уточнялось выше) с Салли(2) (хотя Салли(1) может быть тождественной Салли(2)), то должен быть момент, в котором эта связанность нарушается. Получается, что на протяжении двух лет должен быть день, в котором истинно утверждение, что *сегодняшняя* Салли не отвечает за действия или последствия действий *вчерашней* Салли. Однако данная граница не будет иметь никакого объяснения, в силу чего она проводится, к примеру, на триста семьдесят четвертом дне, а не на шестисотом? Это проведение границы должно быть мотивировано чем-то помимо желания отвергнуть аргумент Меле, и этого объяснения в данный момент нет. Я не утверждаю, что это объяснение невозможно, но, как мне представляется, для каждой независимо мотивированной теории прерывания психологической связанности возможно построение ситуации с Салли, в которой психологическая связанность будет сохраняться. Оставлю это для последующей дискуссии, если она возникнет. Объяснение (б), как было продемонстрировано, концептуально связанно с объяснением (а). В данном же мысленном эксперименте на протяжении всего времени Салли идентифицирует себя как Салли и не видит никакого психологического разрыва. Таким образом, условие (б) также не является удовлетворительным. В результате, оба возражения Мерцалова не опровергают «Т».

Заключение

Таким образом, если моя реконструкция возражений коллег верна, то предложенный модифицированный мысленный эксперимент создает трудности для их возражений. Конкурирующие объяснения, предложенные коллегами, либо не объясняют то, что мы имеем в модифицированном мысленном эксперименте, либо делают *вклад* в объяснение, который дополняет объяснением Меле, но не отвергает его. Таким образом, та разновидность экстернализма в отношении моральной ответственности, которую защищает Меле, не опровергается рассмотренными аргументами.

Externalism about Moral Responsibility: Modification of A. Mele's Thought Experiment

Bogdan V. Faul

Saint-Petersburg State University, Institute of Philosophy. 5 Mendeleevskaya line Str., Saint-Petersburg, 199034, Russian Federation; e-mail: faulbogdan@gmail.com

Tthe author modifies A. Mele's thought experiment for externalism about moral responsibility, which suggests that the agent's history partially determines whether the agent is morally responsible for particular actions, or the consequences of actions. The original thought experiment constructs a situation in which the individual is not morally responsible for the killing because of manipulation, that is, for a reason external to the agent. A. Mele's theory was criticized by A.V. Mertsalov, D.B. Volkov, and V.V. Vasiliev at the seminar organized by the Moscow Center for Consciousness. The arguments against A. Mele's theory had the following structure: A.A. Mele does not show that the historical explanation is the best explanation, because there are competing explanations, no less convincing, which are incompatible with A. Mele's externalism. The author explicates and analyzes the explanations offered by philosophers from the Moscow Center for Consciousness: the explanation from identity, the explanation from self-identification, the explanation from the condition of knowledge, the explanation from future states. Although these explanations apply to Mele's original thought experiment, they cannot explain the absence of moral responsibility in the modified thought experiment proposed by the author: the explanations from identity and self-identification are excluded by the gradual change in the agent structure of personality; the explanation of knowledge conditions is refuted by including knowledge of manipulation in the conditions of the thought experiment; the explanation of future states is excluded by removing relevant future states from the thought experiment.

***Keywords*:** free will, moral responsibility, personality identity

Список литературы / References

Mele, A.R. *Manipulated Agents: A Window to Moral Responsibility.* Oxford; New York: Oxford UP, 2019. 184 pp.

Mele, A.R. *Free Will and Luck.* Oxford; New York: Oxford UP, 2006. 223 pp.

Sider, T. *Four-Dimensionalism: An Ontology of Persistence and Time.* Oxford: Oxford UP, 2001. 255 pp.

Vargas, M. *Building Better Beings: A Theory of Moral Responsibility*. Oxford; New York: Oxford UP, 2013. 360 pp.