

Федеральное государственное бюджетное образовательное учреждение  
высшего образования «Московский государственный университет имени  
М.В. Ломоносова»

Диссертационный совет по истории философии Д 501.001.38

*На правах рукописи*

Волков Дмитрий Борисович

Проблема свободы воли и моральной ответственности в аналитической  
философии конца XX — начала XXI вв.

Специальность 09.00.03 — история философии

Диссертация на соискание ученой степени

доктора философских наук

Научный консультант —

Доктор философских наук, профессор

Васильев Вадим Валерьевич

Москва 2017

# ОГЛАВЛЕНИЕ

<b>ВВЕДЕНИЕ</b> .....	5
I. Актуальность темы.....	5
II. Степень разработанности темы.....	7
III. Объект и предмет исследования .....	14
IV. Цели и задачи исследования .....	14
V. Основные теоретико-методологические принципы исследования .....	15
VI. Положения, выносимые на защиту .....	17
VII. Научная новизна работы.....	20
VIII. Теоретическая и практическая значимость исследования.....	22
IX. Апробация работы.....	23
X. Структура диссертации.....	24
<b>ГЛАВА 1. Современные решения проблемы ментальной каузальности</b> .....	25
I. Аргументы в пользу ментальной каузальности Д. Дэвидсона .....	25
1. Связь проблемы ментальной каузальности и проблемы свободы воли.....	25
2. Корреляция ментальных состояний и действий.....	27
3. Два типа возможных связей между ментальными состояниями и действиями.....	30
4. Доводы против каузальной связи .....	32
5. Доводы Д. Дэвидсона за каузальную связь.....	37
II. Проблема ментальной каузальности в современной аналитической философии .....	43
1. Аргумент исключения Дж. Кима против ментальной каузальности .....	43
2. Аргумент супервентности против каузального влияния одних ментальных состояний на другие .....	47
III. Решение проблемы ментальной каузальности в локальном интеракционизме .....	50
1. Ключевые положения локального интеракционизма .....	50
2. Аргумент каузальных траекторий за ментальную каузальность и против локальной супервентности .....	52
3. Возражения на Аргумент каузальных траекторий.....	62
IV. Решение проблемы ментальной каузальности в биологическом натурализме Дж. Сёрла.....	72
1. Ключевые положения биологического натурализма Дж. Сёрла. ....	72
2. Иллюстрации к каузальным связям в биологическом натурализме.....	76
3. Внутренние противоречия теории Дж. Сёрла .....	80
4. Недостатки проведенных аналогий и иллюстраций .....	82

V.	Решение проблемы ментальной каузальности в телеофункционализме Д. Деннета ....	83
1.	Ключевые положения телеофункционализма Деннета .....	83
2.	«История двух черных ящиков» как иллюстрация каузальной эффективности высокоуровневых свойств .....	90
3.	Интерпретация мысленного эксперимента.....	94
4.	Критика телеофункционализма и возможные ответы .....	97
VI.	Предлагаемое решение проблемы ментальной каузальности.....	111
1.	Аргумент обобщения Н. Блока против Аргумента исключения .....	111
2.	Контрпример к Аргументу исключения. Библиотека первых изданий .....	118
3.	Что не так с Аргументом исключения Дж. Кима. ....	123
VII.	Итоги первой главы.....	125
 <b>ГЛАВА 2. Современные теории свободы воли и детерминизм.....</b>		<b>128</b>
I.	Проблема соотношения свободы, детерминизма и индетерминизма.....	128
1.	Свобода воли и детерминизм .....	141
2.	Свобода воли и индетерминизм. Аргумент случайности.....	151
3.	Дилемма свободы воли: сочетание с детерминизмом и индетерминизмом.....	155
4.	Альтернативные позиции по вопросу свободы воли в конце XX — начале XXI вв. ....	156
II.	Решение проблемы свободы воли в либертарианстве Р. Кейна .....	158
1.	Этапы либертарианской кампании .....	158
2.	Свобода воли и детерминизм несовместимы .....	160
3.	Свобода воли и индетерминизм совместимы .....	173
4.	Возражения к теории Кейна .....	181
III.	Свобода воли и моральная ответственность в твердом инкомпатибилизме Д. Перебума .....	188
1.	Свобода воли и индетерминизм несовместимы .....	190
2.	Свобода воли и детерминизм несовместимы: Аргумент манипуляций.....	203
IV.	Обоснование суперкомпатибилизма .....	209
1.	Возражение на Аргумент манипуляций .....	209
2.	Возражение на Аргумент исчезающего агента.....	221
V.	Итоги второй главы .....	234
 <b>ГЛАВА 3. Теории тождества личности и моральная ответственность .....</b>		<b>236</b>
I.	Связь теории личности с проблемой моральной ответственности и свободы.....	236

II. Проблема единства и тождества личности .....	241
III. Современные подходы к решению вопроса о тождестве личности .....	244
1. Психологический подход.....	246
2. Биологический подход Э. Олсона.....	259
3. Субстанциональный подход и его сторонники .....	266
4. Критика классических подходов с помощью Аргумента пяти трансплантаций .....	278
IV. Теория личности Д. Деннета .....	283
1. Мысленный эксперимент «Где Я»? .....	283
2. Личность как результат эволюции.....	294
3. Критерии личности.....	298
V. Нарративный подход в теориях Д. Деннета и М. Шехтман .....	318
VI. Центр нарративной гравитации .....	333
1. Три аналогии, проясняющие, что такое личность.....	334
2. Эмпирические доводы в пользу концепции Центра нарративной гравитации .....	338
VII. Итоги третьей главы .....	355
<b>ЗАКЛЮЧЕНИЕ .....</b>	<b>358</b>
<b>БИБЛИОГРАФИЯ .....</b>	<b>366</b>

## ВВЕДЕНИЕ

### I. Актуальность темы

Проблема свободы воли является одной из великих философских загадок. Она вызывала интерес Августина и Декарта, Юма и Канта. Однако в последние несколько десятилетий в аналитической философии обсуждение этой темы разгорелось с новой силой. Современные аналитические философы считают свободу воли необходимым условием возможности моральной ответственности и именно в таком ракурсе исследуют её. Возобновление интереса отчасти связано с прогрессом в науке. Квантовая физика позволила философам переосмыслить концепцию каузальности, в частности ввести в оборот понятие вероятностной причины. Новейшие исследования мозга предоставили возможность лучше объяснить механизм принятий решений и выявить основания многих антисоциальных действий. Эти новые открытия заставили философов обновить представления о свободе воли и ответственности. Впрочем, прогрессу способствовали и независимые от эмпирических исследований внутренние философские открытия. Во второй половине XX в. были сделаны концептуальные шаги, которые обозначили новые направления дискуссий. Событиями, повлиявшими на развитие новых направлений, стали публикации статей П. Стросона «Свобода и негодование» (1962), Г. Франкфурта «Альтернативные возможности и моральная ответственность» (1969), работы П. ван Инвагена «Эссе о свободе воли» (1983). Эти публикации позволили с новой перспективой взглянуть на традиционную философскую проблему и обусловили возникновение современного компатибилизма — обновленной теории свободы и ответственности.

Однако наличие новых перспектив не означает консенсус. Текущее состояние философского дискурса интересно именно потому, что в нем представлены радикально отличающиеся позиции. Большинство аналитических философов считают моральную ответственность и свободу воли возможными (и совместимыми с каузальной замкнутостью физического

мира и детерминизмом), но среди философов встречается немало оппонентов, критикующих эту точку зрения. Эти мыслители используют новые аргументы и эмпирические данные для того, чтобы показать, что доминирующая позиция является проблематичной. Такими новыми аргументами являются Аргумент исключения Дж. Кима, Аргумент манипуляций Д. Перебума, Аргумент последствий П. ван Инвагена и другие. Подобные аргументы представляют собой угрозу для наиболее распространенной позиции и требуют внимательного анализа.

Более того, именно альтернативные позиции укрепляются благодаря новым научным данным, полученным, в частности, в квантовой механике и нейрофизиологии. Критика становится более опасной и обоснованной. В этом контексте попытка разобраться в дискуссии и оценить взгляды ключевых современных аналитических философов представляет собой очень интересную и актуальную историко-философскую задачу. Такая попытка предпринята в настоящей работе. В диссертации проанализированы основные позиции и аргументы современных мыслителей по вопросу свободы воли и моральной ответственности, определены наиболее перспективные решения, а также представлены и обоснованы собственные суждения автора по этому поводу. В фокусе исследования оказались также современные исследования в области ментальной каузальности и тождества личности. Согласно основной гипотезе данной работы не только детерминизм, но и каузальная замкнутость физического мира, а также проблема тождества личности являются препятствиями для наличия свободы воли и моральной ответственности. Я считаю, что для решения проблемы свободы воли и моральной ответственности необходимо рассмотреть эти три подпроблемы. Данное исследование должно подтвердить эту гипотезу, а также продемонстрировать, насколько современные аналитические философы близки к формированию теории, способной справиться с указанными трудностями.

## II. Степень разработанности темы

Специального и полного исследования, которое ставило бы своей целью анализ дискуссий в аналитической философии конца XX — начала XXI века по проблемам моральной ответственности и свободы воли и связанных с ними вопросов, в России пока не существует. Однако есть историко-философские работы, в которых рассматриваются теории отдельных мыслителей. Примером такого исследования служит монография Н.С. Юлиной «Философский натурализм: о книге Дэниела Деннета "Свобода эволюционирует"» (2007). В этой работе анализируется позиция в отношении свободы и моральной ответственности одного из самых ярких сторонников современного компатибилизма, Д. Деннета. Автор выявляет наиболее общие черты философского метода Деннета, который она называет «дарвиновским мышлением».

Важной историко-философской работой является книга В. Васильева «Трудная проблема сознания» (2009). В фокусе этого исследования — проблема сознания и ментальной каузальности. Автор анализирует теории трёх наиболее влиятельных философов: Дж. Сёрла, Д. Деннета и Д. Чалмерса, а также предлагает собственное решение проблемы ментальной каузальности и аргументы в его пользу. Другим источником по этой теме служит диссертация Кузнецова А.В. «Проблема ментальной каузальности в аналитической философии сознания». Эта работа защищена относительно недавно (26.10.2016) и представляет один из наиболее полных историко-философских обзоров по проблеме ментальной каузальности. В особенности для нашей темы релевантны вторая и третья главы, в которых рассматривается философия Дж. Кима, Д. Чалмерса, Д. Деннета и Дж. Сёрла. Еще одна историко-философская работа написана С.Ф. Нагумановой — «Материализм и сознание: анализ дискуссии о природе сознания в современной аналитической философии» (2011). Эта книга посвящена дискуссии о природе сознания в современной аналитической философии и защите репрезентативистской теории.

Среди российских историко-философских обзоров можно упомянуть также учебник «Этика» (2016), изданный под редакцией А.А. Гусейнова. Свободе воли в нем посвящен небольшой раздел, где проведен типологический анализ современных подходов. Авторы различают четыре ключевых позиции по вопросу свободы воли и моральной ответственности: теорию совместимости, либертарианскую концепцию, детерминистскую концепцию и концепцию Стросона. В целом этот обзор полезен в качестве введения в проблему, но не предназначен для глубокого исследования темы. Более глубокому знакомству с отдельными аспектами темы способствуют статьи Разина А.В. «Свобода и необходимость» (в ней рассматривается позиция Дж. Сёрла по вопросу свободы воли и рациональности) (2015г.), Секацкой М.А. «Свобода воли и предсказуемость» (об экспериментах Либета и его критике со стороны Д. Деннета, А. Миле, Дж. Тревена) (2016) и С. Левина «Физикализм, субъект и метафизическая свобода воли» (в ней рассматривается соотношение детерминизма и существования субъекта, в частности в либертарианстве Кейна) (2013).

Среди зарубежных историко-философских исследований по теме диссертации можно выделить сборник «Оксфордский справочник по свободе воли» (2005) под редакцией Р. Кейна. В нем содержится более двадцати статей с обзорами и анализом позиций по вопросам свободы воли и моральной ответственности в современной аналитической философии. Достоинством этого сборника является его репрезентативность. Редактор представил все наиболее значимые альтернативные мнения по этим вопросам, а также попытался отразить влияние научных открытий на современный философский дискурс.

В 2016г. вышла еще одна работа, которая тоже в большей степени может считаться историко-философским исследованием. Это книга «Свобода воли: введение в современную проблематику». В отличие от «Оксфордского справочника» эта книга состоит из текстов только двух авторов — М. МакКенны и Д. Перебума. Оба автора известны собственными теориями



свободы воли и моральной ответственности, но в данном издании их задачей является представить все ключевые современные достижения в этой области. Учитывая то, что МакКенна и Перебум занимают противоположные точки зрения (компатибилистскую и инкомпатибилистскую соответственно), им удалось сделать максимально объективный экскурс. Эта работа освещает следующие темы: понятие свободы воли в современной аналитической философии, концепцию детерминизма, связь моральной ответственности и свободы воли, современные аргументы в пользу совместимости и несовместимости свободы воли, моральной ответственности и детерминизма.

Преимущественно историко-философскими текстами можно считать также книги «Философия сознания» (1996) Дж. Кима, «Открывая сознание заново» (1992) и «Сознание: краткое введение» (2004) Дж. Сёрла. Книга Кима выдержала уже три издания; в ней представлен обзор современных позиций по разным аспектам психофизической проблемы. Автор анализирует различные философские подходы к проблеме сознания: субстанциональный дуализм, теорию тождества, функционализм и соответствующие им решения проблемы ментальной каузальности. В монографии философ пытается выяснить, какие ответы современные философы дают на два ключевых вопроса: (1) как соотносится ментальное и физическое и (2) каким образом это соотношение делает возможным ментальную каузальность.

Монографии Сёрла также направлены на изучение современной полемики о проблеме сознания. Его анализ современных теорий сознания и ментальной каузальности охватывает такие позиции, как дуализм свойств, бихевиоризм, теорию тождества, функционализм и элиминативный материализм. В основном автор концентрируется на материалистических теориях, так как они являются доминирующими в философии конца XX — начала XXI вв. Первая из указанных книг переведена на русский язык.

В целом в зарубежной аналитической философии преобладает проблемный и дискуссионный подход. Среди авторов, которые разработали новые концепции и участвовали в полемике по теме данного исследования

необходимо выделить Д. Дэвидсона, Дж. Кима, Дж. Сёрла, Д. Деннета, Д. Чалмерса. Н. Блока, П. Стросона, Р. Кейна, Дж. Фишера, Д. Перебума, Г. Франкфурта, А. Миле, П. ван Инвагена, Р. Суинбёрна, Дж. Перри, Д. Шумейкера, Э. Олсона, М. Шехтман и Д. Парфита. Здесь невозможно перечислить все их монографии и статьи на данную тему, поэтому попытаюсь выделить лишь самые, с моей точки зрения, важные.

В вопросе ментальной каузальности наиболее значимыми считаются исследования Дж. Кима. Корейско-американский философ занимался этой проблематикой на протяжении почти 50 лет. В конце 80-х гг. Ким был президентом Американской философской ассоциации. С 1987г. является заслуженным профессором университета Браун, а также редактором ежеквартального философского журнала «Нус». Помимо уже упомянутой книги «Философия сознания», значимым вкладом Кима в рассматриваемую проблематику служат работы «Сознание в физическом мире» (1998) и «Физикализм или что-то около того» (2005). В этих текстах философ привлекает внимание к проблеме сочетания ментальной каузальности и каузальной замкнутости физического мира.

В вопросе о соотношении моральной ответственности, свободы воли и детерминизма наиболее значимым, с моей точки зрения, является вклад трех современных философов: Р. Кейна, Д. Перебума и Д. Деннета. Эти философы представляют характерные противоположные позиции. Р. Кейн — один из редких сторонников либертарианской теории. Эта теория нашла отражение, в частности, в его монографии «Значение свободы воли», которая стала победителем Гамильтоновской премии в 1996г. Основной заслугой этого философа служит разработка концепции совместимости индетерминизма в квантовом мире и моральной ответственности. Оппонентом Кейна является Д. Перебум — эксперт в области проблем свободы воли и моральной ответственности, философии религии и истории философии, в частности философии И. Канта. Перебум занимает противоположную по отношению к Кейну позицию. По его мнению, не только детерминизм, но и

индетерминизм не совместим с моральной ответственностью и свободой воли. Перебум является профессором университета Корнелл и редактором Стэнфордской энциклопедии философии. Ключевыми проблемно-ориентированными работами философа на интересующую нас тему являются книги «Жизнь без свободы воли» (2001) и «Свобода воли, агентность и смысл жизни» (2014).

«Пространство для свободы» (1984) и «Свобода эволюционирует» (2003) являются проблемно-ориентированными работами Д. Деннета, посвященными той же теме, что и данная диссертация. Их автор защищает компатибилистскую теорию. Книга «Пространство для свободы» написана философом на основе его лекций в Оксфорде в 1983г. В этой монографии Деннет пытается объяснить, что детерминизм не представляет угрозы для свободы и ответственности. Окончательно объяснить и рассеять метафизические страхи американский философ пытается в более поздней работе «Свобода эволюционирует». Эта книга — наиболее полное объяснение Деннетом феноменов свободы и ответственности с натуралистических, эволюционных позиций. Ее цель — показать, что свобода — не фундаментальное свойство объектов, как масса или заряд, а появившаяся в ходе эволюции способность личностей.

Работы Д. Деннета важны и в контексте проблемы тождества личности, в первую очередь — его статья «Личность как центр нарративной гравитации» (1992). Совсем недавно вышла новая книга Деннета «От бактерии до Баха и обратно: эволюция разума» (2017). В ней доработана теория личности как результат биологической культурной эволюции. В этом тексте философ представляет нарративный подход к проблеме тождества личности. Но наиболее значимый вклад по этому вопросу внёс Д. Парфит. Это британский философ, занимавшийся вопросами этики, рациональности и тождества личности. В его главной работе «Причины и личности» (1984) представлены идеи, ставшие ядром современной дискуссии. Парфит был лауреатом международной премии Рольфа Шока в 2014г. за новаторство в

области решения проблемы тождества личности, заботы о будущих поколениях и структурирования этических теорий. Академическая карьера Парфита проходила в Оксфордском университете и, к сожалению, прервалась его смертью 1 января 2017г.

В современной российской философии также есть проблемно-ориентированные исследования. Прежде всего необходимо отметить вклад Д.И. Дубровского. Он был одним из тех, кто стал развивать аналитическую философию еще в СССР и привлек внимание к проблеме сознания. В 1962 г. Дубровский защитил кандидатскую диссертацию на тему «Об аналитико-синтетическом характере отражательной деятельности мозга». Во второй половине 60-х он опубликовал несколько статей в журнале «Вопросы философии» и защитил докторскую диссертацию «Философский анализ психофизиологической проблемы» (1969г.). В этих работах, как и в последовавших монографиях («Информация, сознание, мозг», 1980; «Проблема идеального», 1983, 2002; «Сознание, мозг, искусственный интеллект», 2007), Д.И. Дубровский защищал информационный подход к объяснению «субъективной реальности». В этом направлении Дубровский предлагал и решение проблемы ментальной каузальности. По его мнению, информационные каузальные объяснения оказываются необходимыми, когда предметом исследования служат самоорганизующиеся системы (биологические, социальные, технические). При этом информационная причинность не противоречит физической причинности. Как пишет философ, физическая причинность сохраняет значение, если не претендует на роль универсального средства объяснения всех явлений действительности. Информационный подход развивается Дубровским и в одной из его недавних работ, книге «Проблема "сознание и мозг": теоретическое решение» (2015).

Среди других российских авторов, использующих проблемный подход, — Д. Иванов. Книга Иванова «Природа феноменального сознания» (2013) проясняет природу феноменального аспекта сознания и затрагивает вопросы ментальной каузальности, обсуждаемые в первой главе настоящей

диссертации. Особенно важными для целей проводимого исследования являются две книги В. Васильева «Сознание и вещи» (2014) и «В защиту классического компатибилизма» (2016). В первой работе представлена теория локального интеракционизма — одно из оригинальных решений проблемы ментальной каузальности. Вторая книга содержит развернутую защиту классической позиции в отношении свободы воли, разработанную ещё Д. Юмом. Оппонентами автора на страницах этой книги являются современные аналитические инкомпатибилисты, такие как Р. Кейн и Д. Перебум. Основными выводами российского философа служат утверждения, во-первых, о том, что свобода воли и моральная ответственность совместимы с детерминизмом и, во-вторых, что свобода воли основана на свободе действия как необходимом условии моральной ответственности. Нужно отметить, что работа «В защиту классического компатибилизма» является уникальной, так как это — первая отечественная книга на данную тему, выполненная в современной аналитической традиции.

В заключение этого обзора я также хотел бы отметить проблемно-ориентированные статьи других российских авторов. Это публикации Д. Разеева о философии сознания и тождестве личности (в частности, статьи «Ложные воспоминания и тождество личности», «О двух уровнях эпистемологии сознания», «Сознание как феномен природы», доклад «Свобода воли: иллюзия или реальность»), С. Левина на тему свободы воли и моральной ответственности (статья «Коллективное сознание и моральная ответственность»), а также о нарративной теории тождества личности (статья «моральная ответственность и личность корпораций: нарративный подход), Д. Чирвы о свободе воли и тождестве личности (статьи «"Правильная" постановка вопроса о свободе: деятельностный индетерминизм», «Одинокое животное. Биологический подход к тождеству личности», «Попытка преодоления психофизического дуализма в современных дискуссиях о тождестве личности»), М. Секацкой о свободе воли и тождестве личности (статьи «Интенциональность, свобода воли и моральная ответственность»),

«Пересадка мозга и тождество личности: альтернативная интерпретация одного мысленного эксперимента», «Моральная ответственность без свободы воли») и А. Беседина по свободе воли («Может ли детерминизм нами манипулировать», «Натуралистический аргумент П. Стросона в пользу моральной ответственности»). Эти исследователи ведут активную работу в интересующей нас области в составе Центра феноменологии и философии сознания при СПбГУ и Центра исследования сознания при МГУ.

### III. Объект и предмет исследования

Объектом исследования диссертационной работы являются теории моральной ответственности и свободы воли в современной аналитической философии. Предметом исследования служат лежащие в их основе аргументы, направленные на доказательство или опровержение возможности свободы воли и моральной ответственности. В центре исследования — вопросы о том, как возможна моральная ответственность и что об этом думают современные аналитические философы, а также вопрос о возможности свободы воли, который рассматривается в диссертации в контексте трех проблем: каузальной замкнутости физического мира, детерминизма и тождества личности.

### IV. Цели и задачи исследования

Целью данного исследования является выявление наиболее значимых теорий моральной ответственности и свободы воли в аналитической философии конца XX в. — начала XXI в., определение наиболее аргументированной позиции по этим вопросам, а также представление и обоснование собственных суждений, способных укрепить эту позицию. Поставленная цель предполагает решение следующих исследовательских задач:

- определение связи между проблемами моральной ответственности и свободы воли в современной аналитической философии;

- определение основных подпроблем, которые стоят перед современными философами, разрабатывающими теории моральной ответственности и свободы воли;
- выявление основных работ, написанных на эту тему, определение основных представителей ключевых позиций и специфики их взглядов;
- систематизация позиций по указанным проблемам;
- реконструкция этих позиций: выявление внутренних связей, ключевых аргументов, суждений, доводов, мысленных экспериментов, которые лежат в основе философских позиций по указанным вопросам;
- сравнительный анализ этих позиций и их сопоставление для выяснения их сильных и слабых сторон;
- выявление наиболее обоснованной и достоверной теории моральной ответственности и свободы воли;
- представление собственных аргументов, подтверждающих ее достоверность и помогающих укрепить ее слабые стороны.

#### V. Основные теоретико-методологические принципы исследования

Данная работа представляет собой исследование в области истории современной аналитической философии. Поэтому ключевой используемый метод в исследовании — *историко-философский*. Этот метод предполагает фактографическое воспроизведение истории философской мысли, в частности выделение основных источников и подробное их изложение. Аналитический обзор охватывает как основные положения философских теорий, так и их аргументативную базу. Фактографическое изложение и *текстологический анализ* материала составляют основу настоящего исследования, однако применение историко-философского метода этим не ограничивается. Он предусматривает также аналитическую интерпретацию философских теорий, их историческую реконструкцию, выявление этапов разработки теорий, определение их контекста, выделение существенных и

вторичных элементов, концептуальный анализ понятий. Такой подход применен по отношению к современным теориям свободы воли и моральной ответственности.

В работе также использован метод *сравнительного анализа*. В данном исследовании современные теории моральной ответственности и свободы воли не только реконструированы и проинтерпретированы, но и сопоставлены между собой. В результате сопоставления выделены их ключевые общности и отличия, слабые и сильные стороны, представлена их систематизация. Сравнительный анализ позволил представить философские теории моральной ответственности и свободы воли в контексте метафизических позиций, в котором происходило их формирование (в частности, теорий сознания), и тем самым объективно отразить их историко-философский смысл. Все это достаточно характерно для большинства историко-философских исследований. Однако в данной работе также есть специфика. Она заключается в превалировании *проблемного подхода*.

В фокусе исследования находятся именно философская проблема существования свободы и моральной ответственности, а также способы ее решения. Это обуславливает специфический отбор материала и структуру исследования. В главах диссертации отсутствуют биографические данные философов, чьи теории разбираются. Такое содержание могло бы отвлечь от основной линии повествования. Практически не освещены исторические социокультурные предпосылки и субъективные факторы, которые могли повлиять на формирование позиции мыслителей. Несмотря на то, что эта информация могла бы быть полезной, она не имеет непосредственного отношения к цели исследования. Аспекты творчества исследуемых мыслителей представлены избирательно. Основное внимание уделено тем местам трудов, которые имели непосредственное отношение к предмету исследования. Материал представлен не в хронологической, а в логической последовательности. Такой подход обеспечил целостность и практическую направленность работы.



Большая часть внимания уделена аргументативной основе положений изучаемых теорий моральной ответственности. В частности, разобраны и проанализированы многочисленные примеры (как собственного авторства, так и предлагаемые другими философами), подкрепляющие или опровергающие те или иные философские выводы. Значительное место посвящено анализу мысленных экспериментов — специальных гипотетических ситуаций, используемых для доказательства или иллюстрации позиций. Это часто применяемый в аналитической философии приём. В данном исследовании он использован активно, но не без дополнительной осторожности. Мысленные эксперименты позволяют получить новое знание о предмете исследования без новых эмпирических наблюдений. Во многих случаях их использования в философии это новое знание, как предполагается, появляется вследствие раскрытия интуиций, почерпнутых в предыдущем опыте. Однако в слишком фантастических сценариях интуиции могут быть ошибочными. Поэтому анализ гипотетических ситуаций по возможности сопровождается анализом самого аргумента: посылок и следствия, вывода в целом.

#### VI. Положения, выносимые на защиту

1. Решение проблемы свободы воли и моральной ответственности предполагает решение вопроса об их совместимости с детерминизмом, а также проблем ментальной каузальности и тождества личности.
2. Наиболее убедительной теорией действия является стандартный каузализм, предложенный Д. Дэвидсоном. На ней основывается «тезис о ментальном»: утверждение о том, что ментальная каузальность является необходимым условием свободного и ответственного действия.
3. Ключевым аргументом в современной аналитической философии против ментальной каузальности является Аргумент исключения, разработанный Дж. Кимом.

4. Наиболее значимыми подходами к решению проблемы ментальной каузальности являются локальный интеракционизм, биологический натурализм и телеофункционализм.
5. Общим для локального интеракционизма и биологического натурализма является утверждение нередуцируемости ментальных состояний.
6. Уязвимость локального интеракционизма обнаруживается в аргументе локальных траекторий. Этот аргумент можно критиковать с помощью Аргумента-двойника.
7. Слабая сторона биологического натурализма Сёрла — в его слишком широком определении причины. В теории Сёрла это понятие неоправданно включает другие важные философские отношения, такие как реализация и супервентность.
8. Наиболее правдоподобным является решение проблемы ментальной каузальности в телеофункционализме Деннета. Это решение в целом поддерживается Аргументом обобщения, а также собственным мысленным экспериментом автора диссертации — «Библиотекой первых изданий».
9. Предлагаемое Деннетом решение позволяет отыскать уязвимость в аргументе Дж. Кима. Эта уязвимость — в принципе каузальной замкнутости физического: он предполагает слишком узкую интерпретацию физических свойств.
10. В современной аналитической философии условием свободы воли считается соблюдение двух принципов: принципа альтернативных возможностей и принципа автономии. Аргументы, направленные против возможности свободы воли, атакующие совместимость их с каузальным детерминизмом, построены на этих принципах. Такими аргументами являются Аргумент последствий в формулировке П. ван Инвагена и Аргумент манипуляций в формулировке Д. Перебума. Если

принять хотя бы один из этих аргументов, придется занять инкомпатибилистическую позицию.

11. Наиболее значимыми инкомпатибилистскими теориями свободы воли и моральной ответственности являются теории Р. Кейна и Д. Перебума.
12. Новизна теории Кейна заключается в использовании современных эмпирических знаний о физическом мире и, в частности, устройстве мозга для укрепления либертарианской позиции.
13. Инкомпатибилистские теории не являются до конца удовлетворительными. Так, позиция Кейна уязвима для встречного аргумента — Аргумента исчезающего агента Д. Перебума. Этот аргумент принадлежит традиционному типу Аргументов случайности. Новизна этого аргумента в том, что он показывает, что выбор даже между желаемыми альтернативами не спасает моральную ответственность, но делает ее еще более проблематичной. Твердый инкомпатибилизм Д. Перебума также не лишен существенных недостатков. Автор диссертации показывает, что Аргумент манипуляции Перебума рождает противоречивые интуиции и может быть интерпретирован иначе.
14. Наиболее значимыми компатибилистскими и полукompatibiliстскими теориями свободы воли и моральной ответственности являются теории Д. Деннета и Дж. Фишера.
15. Анализ аргумента манипуляций Д. Перебума демонстрирует, что необходимым основанием моральной ответственности является существование протяженной во времени целостной личности.
16. Попытки решения проблемы единства личности во времени в современной аналитической философии представлены теориями, основанными на психологическом, биологическом, субстанциональном и нарративном подходе. Наиболее распространен психологический подход.

17. Самую существенную проблему для психологического подхода представляет проблема ветвления личности в том виде, в котором она сформулирована Б. Вильямсом в «Аргументе редупликации» и Д. Парфитом в мысленном эксперименте «Телетранспортер». Попытки преодоления этой проблемы современными теоретиками тождества личности в основном являются неудовлетворительными.
18. Наиболее удачным является решение этой проблемы в нарративном подходе. Нарративный подход и концепция личности как «Центра нарративной гравитации» позволяют разобраться в наиболее трудных гипотетических ситуациях и прояснить понятие личности.
19. Нарративный подход (в частности, в том виде, в котором его представляют М. Шехтман и Д. Деннет) позволяет снять требование нумерического тождества во времени и заменить его требованием связанности, что успешно обеспечивается текстом. Нарративный подход отличается от классических подходов тем, что он заменяет вопрос реидентификации на вопрос характеристики.
20. Согласно нарративному подходу единый агент — это допущение, результат интерпретации, теоретическая сущность. Личность как абстрактная, фиктивная сущность наделяется фантазийными характеристиками, такими как полная рациональность, целостность, неделимость, уникальность. Для конкретного человека эти характеристики недостижимы. И человек всегда обладает ими только «отчасти».

## VII. Научная новизна работы

Диссертация посвящена исследованию теорий моральной ответственности и свободы воли конца XX — начала XXI вв. В отечественной историко-философской науке нет работ, которые бы раскрывали весь спектр позиций на эту тему. И как я уже отмечал выше, даже в зарубежной истории философии такие работы представляют собой редкость. Таким образом, сам объект исследования обуславливает новизну.

Чаще всего в монографиях зарубежных философов в центре внимания находится собственная позиция авторов, а позиции их современников служат мишенью для критики. Настоящая работа в этом отношении отличается. В ней репрезентативно представлены основные направления решений проблем свободы воли и моральной ответственности. Выделены ключевые фигуры дискуссий, основные положения их теорий и центральные аргументы, обнаружены их взаимосвязи, основные точки противоречий и слабые места.

Кроме того, новизну работы определяет ракурс, в котором рассматривается проблема моральной ответственности. В данной диссертации она представлена в тесной связи с проблемами ментальной каузальности, свободы воли и тождества и единства личности. Такой ракурс позволяет провести обзор достаточно широкой области современной аналитической философии и познакомиться со значительным кругом ключевых фигур в этой сфере. Более того, он позволяет оценить эффективность основных методов аналитической философии на разнообразном материале.

Большинство источников, на которые опирается исследование, представляют собой книги, изданные в период с 1980 по 2016г. Большая часть этих работ не освещалась в отечественной истории философии. Почти все эти тексты существуют только на английском языке и недоступны читателям, которые им не владеют. Поэтому диссертация может служить также источником знакомства с этими новыми работами.

Кроме того, в работе представлены собственные мысленные эксперименты и аргументы автора. Они позволяют укрепить защищаемые позиции и по-новому взглянуть на анализируемые теории и их основные положения. Новыми, в частности, являются Аргумент четырехмерного агента, Аргумент-двойник. Автором представлены также собственные мысленные эксперименты: Закоренелый радиоуправляемый преступник, Община джайнов, Библиотека первых изданий. Они позволяют усилить отстаиваемую позицию.

В совокупности перечисленные особенности составляют новизну диссертационного исследования.

### VIII. Теоретическая и практическая значимость исследования

Теоретическая значимость работы состоит в том, что диссертация создает полномасштабную картину дискуссий о проблеме свободы воли и моральной ответственности в современной аналитической философии. В диссертации сочетаются как историко-философский, так и проблемный подходы. Данное исследование знакомит читателей с современным англоязычным философским дискурсом по обозначенным проблемам. В нем определяется круг наиболее значимых современных мыслителей в данной области. На базе обширного материала проводится демонстрация работы ключевых философских методов и приемов аналитической философии, в частности построения аргументов, концептуального анализа и мысленных экспериментов. Теоретическая значимость диссертации состоит также в том, что она позволяет российским исследователям получить широкое представление о степени разработанности указанных проблем в современной аналитической философии.

Практическая значимость диссертации заключается в том, что ее материалы могут использоваться для чтения общих курсов по современной зарубежной философии, для специализированных курсов, посвященных аналитической философии, и курсов по истории философии, направленных на изучение проблем свободы воли и моральной ответственности. Результаты, полученные в диссертации, также могут быть использованы для составления учебных пособий по современной аналитической философии, энциклопедических статей и обзоров. Кроме того, диссертация может служить в качестве путеводителя по современной аналитической философии для издательств, которые намереваются переводить и публиковать современных англоязычных философов на русском языке. Кроме академической значимости, диссертация также имеет и другое практическое

значение. Ряд проблем, освещенных в диссертации, имеет непосредственное отношение к юридическим и уголовным вопросам. В частности, выводы диссертанта могут быть отчасти использованы при вынесении наказаний за уголовные преступления психопатам и другим личностям с психическими особенностями. Также выводы диссертации могут использоваться на практике для обновления юридических норм в отношении преступлений, совершенных в различных психических состояниях. Эти моменты представляют ценность работы для специалистов. Однако данная работа имеет и более широкую практическую значимость. Аналитическая философия в целом и данная работа в частности представляют собой пример анализа понятий и аргументов, лежащих в основе критического мышления. Мне кажется, нет более применимой способности, чем способность рационально и непротиворечиво рассуждать.

#### IX. Апробация работы

Материалы данной диссертации использовались автором в ходе выступлений на различных конференциях, семинарах, коллоквиумах, летних школах, дискуссиях. В частности, автор выступал на международной конференции «Философия сознания: аналитическая традиция. Третьи Грязновские чтения» (6–7 ноября 2009г., МГУ); на круглом столе «30 лет Китайской комнате», посвященном аргументу Китайской комнаты (15 марта 2010г., МГУ); на круглом столе «Мыслимость, возможность и крипкеанская критика тождества сознания и тела», проведенном проф. Сомсом, посвященном аргументам Крипке (26 мая 2010г., МГУ); на пресс-конференции с участием Д. Деннета, посвященной презентации монографии Волкова Д.Б. «Бостонский зомби. Теория сознания Д. Деннета» (8 июля 2012г.); телемосте «Философский факультет МГУ — кафедра философии SUNY», посвященном вопросам тождества личности (20 мая 2014г., МГУ); международной конференции «Проблемы сознания и свободы воли в аналитической философии», состоявшейся 12–19 июня 2014г. в Гренландии, в совместном докладе с проф. Р. Хауэллом «Либертарианская теория Р.

Кейна»; летней школе в Москве, проводимой проф. Дж. Фишером «Свобода воли и моральная ответственность» (11–25 августа 2014, МГУ); круглом столе «Новейшие тенденции в аналитической философии» (26 февраля 2015г., МГУ); круглом столе «Новейшие тенденции в аналитической философии: Почему я — это я?», посвященном проблеме тождества личности (15 апреля 2016г., МГУ); дебатах «Свобода воли: новые повороты старых дискуссий» (совместно с В.В. Васильевым), проводимых ИФ РАН (21 апреля 2016г.); летней школе в Риге под руководством проф. Перебума «Свобода воли и сознание» (11–24 июля 2016); коллоквиуме «Наука о самоконтроле и новые аргументы в дебатах о свободе воли» (21–22 сентября 2016, НИУ ВШЭ). Часть материалов была представлена студентам философского факультета МГУ во время чтения спецкурса, посвященного современной аналитической философии. Некоторые материалы диссертации были также включены в состав монографии автора «Бостонский зомби. Деннет и его теория сознания», опубликованной в 2012г.

#### Х. Структура диссертации

Диссертация состоит из введения, трех глав, заключения и библиографического списка. Первая глава посвящена проблеме ментальной каузальности как необходимому условию свободного действия. Она состоит из семи частей. Вторая глава посвящена проблеме свободы воли и детерминизму и содержит пять частей. В третьей главе разбирается проблема единства личности в связи с проблемой моральной ответственности, она содержит семь частей.



## ГЛАВА 1. Современные решения проблемы ментальной каузальности

### I. Аргументы в пользу ментальной каузальности Д. Дэвидсона

#### 1. Связь проблемы ментальной каузальности и проблемы свободы воли

В начале первой главы я представлю аргументы в пользу «стандартного каузализма»<sup>1</sup> и «тезиса о ментальной каузальности». Стандартным каузализмом называется сочетание двух утверждений: (1) событие может являться действием, если имеет специфические причины; (2) характерные объяснения действий — это каузальные объяснения. Тезисом о ментальной каузальности я называю утверждение о том, что действенность ментальных свойств необходима для возможности свободного и ответственного действия. Стандартный каузализм и тезис о ментальной каузальности демонстрируют зависимость проблемы свободы воли и ответственности от подпроблемы ментальной каузальности.

Стандартный каузализм является наиболее убедительной теорией действия. Аргументы в пользу стандартного каузализма и тезиса о ментальной каузальности разработаны Д. Дэвидсоном в 60-х гг. XX в. В статье 1963 г. «Действия, мотивы и причины» он предложил новые доводы в пользу утверждения, что мотивы являются причинами действия. Эти доводы были направлены против влиятельных философов XX в., таких как Л. Витгенштейн, Г. Райл, Э. Энском. Почти сразу после выхода статьи рассуждения философа получили значительную поддержку и «определили стандарт для работ в области теории действий»<sup>2</sup>.

Дэвидсон показал, что ментальные содержания являются логически независимыми от реализующих действий, поэтому могут служить их причиной. Дэвидсон также доказал, что каузальная действенность ментальных состояний наилучшим образом обосновывает их

<sup>1</sup> Mele A. Causation, Action, and Free Will // Beebe H., Hitchcock C., and Menzies N. (eds.). The Oxford Handbook of Causation. Oxford, New York: Oxford University Press, 2009. P. 534.

<sup>2</sup> A Companion to the Philosophy of Action. O'Connor T. and Sandis C. (eds.). Oxford: Blackwell, 2010. P. 598.

объяснительную способность (*explanatory force*). В совокупности с другими соображениями, в том числе с «принципом элеатов»<sup>3</sup> и тезисом о ментальной каузальности, эти рассуждения представляют достаточное основание для утверждения наличия ментальной каузальности. Однако, как будет показано во *второй части* первой главы, наличие ментальной каузальности может быть подвергнуто критике. Эта критика усилилась в последние несколько десятилетий.

Наиболее грозным орудием критики каузальной эффективности ментальных свойств является Аргумент исключения. Первые версии этого аргумента были разработаны Ч. Тейлором и Н. Малкольмом в 60-х гг. XX в. В конце 80-х этот аргумент был актуализирован американским метафизиком Дж. Кимом. Аргумент исключения Кима основывается на «тезисе о физикализме» — утверждении о том, что все события в качестве своих причин имеют только другие физические события. Из Аргумента исключения следует невозможность ментальной каузальности. А этот вывод в сочетании с тезисом о ментальной каузальности ведет к отрицанию возможности свободного и ответственного действия. Такое следствие крайне контринтуитивно. Поэтому многие современные философы не согласны с Аргументом исключения и пытаются найти решение выявленной Кимом проблемы. Наиболее значимыми, по моему мнению, подходами к решению проблемы ментальной каузальности являются локальный интеракционизм, биологический натурализм и телеофункционализм.

После анализа этих подходов я также представлю собственное дополнение к телеофункционалистскому решению проблемы ментальной каузальности. Итогом главы будет обоснование «ментального каузализма» — конъюнкции трех утверждений: (1) действие — это событие, определенное ментальной каузальностью; (2) ментальная каузальность является

---

<sup>3</sup> Colyvan M. Can the Eleatic Principle Be Justified? // Canadian Journal of Philosophy. 1998. Vol. 28. No. 3. P. 313–336.

необходимым условием свободного действия; (3) ментальная каузальность возможна в условиях физикализма. Подчеркну, что ментальный каузализм, который я обосновываю в этой главе, представляет только необходимые, но не достаточные условия свободного действия. Другие условия свободного действия и свободы воли будут разобраны во второй и третьей главе.

## 2. Корреляция ментальных состояний и действий

Среди всего множества процессов и событий можно выделить те, что являются действиями. Особенность действий в том, что они предполагают наличие агента, действующего активного субъекта. Действия каким-то образом связаны с агентом и его психическими состояниями. Но какова эта связь, и чем различаются события в жизни агента и его действия? Д. Дэвидсон задается подобными вопросами в эссе «Агентность»: «Какие события в жизни личности демонстрируют агентность? Какие события являются его поступками и действиями в его истории? Что выделяет его действия?»<sup>4</sup> Сегодня я проснулся рано утром, разбуженный звуком скрипки, на которой кто-то занимался, умылся, побрился, — продолжает Дэвидсон, — налил себе чашку кофе, споткнулся о ковер и разлил кофе по неуклюжести. Часть из того, что случилось — мои действия, а часть — просто события из моей жизни. Так, пробуждение и спотыкание — события, а остальное — действия. Но чем они отличаются?

Кажется, действиями являются те события, которые объясняются внутренними состояниями агента. И кажется, что внутренние состояния агента объясняют действия. Это утверждение разделяют многие аналитические философы, например А. Макинтайр. В монографии «После добродетели: исследование теории морали» он пишет: «Определить событие как действие означает дать ему такое описание, чтобы это событие разумно и логично проистекало из намерений, желаний, целей, переживаний агента-

---

<sup>4</sup> Davidson D. *Essays on Actions and Events*, Oxford: Clarendon Press, 1980. P. 51.

человека»<sup>5</sup>. От других событий действия отличаются тем, что у них есть описание, логически проистекающее из психических состояний агента.

Вот в чайнике закипела вода. Кто-то поставил его на плиту и включил конфорку. Процесс разогревания чайника — не только ряд физических событий, но и действие агента. Конечно, нагревание и кипячение проходят по законам термодинамики. Но в отличие от других событий и процессов, скажем, вспышки сверхновой или извержения вулкана, кипение воды в чайнике — преднамеренное событие; оно соотносится с ментальными состояниями агента, его внутренним осознанным содержанием, намерениями и убеждениями. Кто-то захотел чая, и ему потребовалась горячая вода.

Можно, конечно, предположить, что действия отличаются в другом: что они состоят из движений тела живых существ, или движений предметов, которые пространственно соединены с телом. Но искать контрпример долго не придётся. Скажем, рефлекторные движения рук или ног явно нельзя отнести к действиям. Вздрагивания, кашель и храп не являются действиями, как и пробуждение или спотыкание<sup>6</sup>. А полёт радиоуправляемого самолёта, скорее всего, — это действие, хотя сам пилот остаётся на земле. Убийство с помощью выстрела тоже, скорее всего, следует считать действием, хотя движение пули, непосредственно попадающей в жертву, не является движением тела агента.

Действия от процессов также отличаются в возможности телеологического объяснения. Они предполагают направленность событий в соответствии с намерениями, желаниями и убеждениями субъектов. Так, в случае с убийством направленность может соотноситься с желанием отомстить, войти в историю или получить коммерческую выгоду.

---

<sup>5</sup> MacIntyre A. *After virtue: A Study in Moral Theory*. Notre Dame, Indiana: University of Notre Dame Press, 2007. P. 209.

<sup>6</sup> Davidson D. *Essays on Actions and Events*, Oxford: Clarendon Press, 1980. P. 45.

Соотношение желаний, убеждений и действий агента аналитический философ Дж. Ким выражает с помощью следующего принципа:

*(A) Если агент S желает что-то и убежден, что действие A обеспечит ему реализацию желания оптимальным образом, S будет делать A<sup>7</sup>.*

В формулировке Кима неизбежной оказалась ссылка на оптимальность. Без неё прогноз включал бы в себя бесчисленное множество возможных действий, ведь одного и того же результата можно добиться разными способами. Сузить список помогает понятие оптимального действия — это реализация намерения с наименьшим количеством усилий в кратчайшие сроки и в соответствии с навыками агента. Но в данном случае эта оптимальность относительная. Для того, чтобы сделать выбор образа действий, достаточно, чтобы сам агент считал способ действия оптимальным, не видел других лучших альтернатив. При этом объективно могут существовать более оптимальные способы реализации желания.

Однако с одним этим разъяснением принцип все еще не будет эффективно работать. Он выглядит слишком упрощенным, ведь у агента редко бывает только одно желание. Нужно сделать поправку на то, что желание не входит в конфликт с другими желаниями и убеждениями. Скажем, желание выпить чашку горячего чая могло бы вступить в конфликт с другим, более сильным желанием — успеть на работу — и убеждением, что время выходить уже наступило. Тогда вода бы не закипела в чайнике, и он остался бы пустым и холодным. На применение обозначенного принципа накладываются и другие ограничения. Вот вы захотели выпить чашку чая, идёте на кухню, а дверь в кухню закрыта на ключ и ключа нет. Принцип работает только в рамках допустимых возможностей. Конечно, на кухню можно было бы попасть и через окно или даже выломав дверь. Но здесь, скорее всего, опять возникнет конфликт желаний. Вряд ли желание чая будет

---

<sup>7</sup> В таком виде этот принцип приводит Дж. Ким в книге “Философия сознания”. Kim J. *Philosophy of Mind*. Cambridge, Massachusetts: MIT Press, 2006. P. 175.

сильнее, чем желание сохранить собственную жизнь или имущество неприкосновенными. Поэтому принцип нужно усовершенствовать так:

*(Б) Если агент S желает что-то и убежден, что действие A обеспечит ему реализацию желания оптимальным образом, и при этом реализация желания физически возможна и не противоречит другим его более сильным желаниям, S будет делать A.*

Помимо физических ограничений и конфликтующих желаний, препятствовать действию A могут также привычки. S, например, захотел утром горячего чая, но в чашку плеснул по привычке кофе. Действие может не последовать, даже если нет препятствий, в том случае, если само желание не проявляется достаточно сильно. Или не произойти потому, что оно противоречит представлению агента о собственном долге. Это представление может складываться из морального чувства, обычаев и практики. Агент может взять на себя обязательства, и, несмотря на наличие противоречащего желания, эти обязательства соблюдать. Подобные случаи можно было бы также внести в качестве исключений в указанный базовый принцип. Но чтобы избежать дополнительных усложнений, следует просто принять, что в принципе (Б) подразумевается желание достаточно сильное, чтобы инициировать действие, а оптимальный способ реализации не противоречит нормативным представлениям агента. В таком виде принцип, видимо, корректно фиксирует корреляцию между ментальными состояниями (убеждениями и желаниями) и действиями. Но остаётся не до конца ясным, на чем основана эта корреляция. Как именно желания и убеждения связаны с действием?

### 3. Два типа возможных связей между ментальными состояниями и действиями

Из принципа (Б) мы видим, что желания и убеждения по направленности коррелируют с действиями. Они предвещают результат действий. Кроме того, ментальные содержания участвуют в объяснении, рационализации самого действия. Почему человек вскипятит воду? Потому что хотел выпить

чай. Почему он совершил убийство? Потому что он надеялся попасть в историю. Но тот факт, что объяснение действия предполагает указание на эти состояния, не даёт нам однозначности в понимании подчинительного союза «потому что». Дж. Ким замечает, что можно выявить как минимум два значения этого союза и, соответственно, два типа отношений. Иногда подчинительный союз обозначает (1) каузальную связь, т.е. указывает на причину, как, например, в предложении: «здание обрушилось потому, что случилось сильное землетрясение». Здесь союз связывает два различных события, одно из которых было достаточным и необходимым основанием совершения второго. И эта каузальная связь подпадает под один из общих законов. Второе значение не предполагает каузальной связи и не является частным случаем общего закона. Союз «потому что» в таком случае разъясняет первое событие посредством расшифровки его значения: «Она стала профессиональной художницей потому, что рисует картины и выставляет их на выставках» или «Он стал совершеннолетним потому, что ему исполнилось восемнадцать лет». В этих предложениях речь не идет о двух различных состояниях или событиях. Художник и человек, профессионально рисующий картины — синонимичные понятия. Наступление совершеннолетия и достижение восемнадцати лет — тоже одно событие, и слово «потому» здесь просто разъясняет это событие. Разъяснение здесь фактически указывает на (2) логическую связь.

Современных философов, которые считают, что между действиями и ментальными состояниями, объясняющими их, существует каузальная связь, называют *каузалистами*. Их противников — *антикаузалистами* (*noncausalist*). Последние убеждены, что между действиями и объясняющими их ментальными состояниями не существует каузальной связи, хотя может признаваться какая-то другая связь, например, логическая<sup>8</sup>. До публикации Дэвидсоном статьи «Действия, мотивы и причины» (1963) доминирующей

---

<sup>8</sup> Ginet K. Reasons Explanations of Action: Causalist Versus Noncausalist Accounts // Oxford Handbook on Free Will. Ed. Kane R. Oxford University Press, 2005. P. 386-387.

позицией среди аналитических философов был антикаузализм. После публикации Дэвидсона ситуация радикально поменялась. О результате изменения А. Миле пишет: «Лидирующим подходом в отношении как понимания, что такое действие, так и объяснения или определения действий является подход, опирающийся на каузальность. То, что может называться *стандартным каузализмом*, выражается в сочетании двух тезисов: (1) событие может являться действием в зависимости от того, что служит его причиной; (2) характерным объяснениями действий являются каузальные объяснения»<sup>9</sup>. Таким образом, Дэвидсон стал основоположником современного каузализма. Он представил аргументы, показывающие, что мотивы, объясняющие действия, являются их причинами. Эти доводы будут рассмотрены в ближайшее время. Но сначала следует рассмотреть аргументы в пользу антикаузализма, считавшиеся наиболее правдоподобными до 1963г. Любопытно, что Дэвидсон косвенным образом, внес вклад в развитие и этой позиции.

#### 4. Доводы против каузальной связи

##### А. Трудности с доказательством психофизических законов

Традиционной моделью каузальности является номологическая модель<sup>10</sup>. Её принимают многие современные философы. Модель предполагает, что каждый случай причинной цепи представляет собой частный случай общих законов. Т.е. событие *a* является причиной события *b* в том случае, если событие *a* подпадает под тип событий *F*, а *b* — под тип *G*, и существует закон, связывающий события типа *F* и *G*. Нагревание воды в чайнике до определённой температуры, к примеру, является причиной её кипения, т.к. существует общая закономерность парообразования в жидкости при

---

<sup>9</sup> Mele, A. Causation, Action, and Free Will // Beebe H., Hitchcock C., and Menzies N. (eds.). The Oxford Handbook of Causation. Oxford, New York: Oxford University Press, 2009. P. 534.

<sup>10</sup> К примеру, её принимали К. Гемпель и Д. Дэвидсон. «Согласно как Гемпелю, так и Дэвидсону, каузальность — это вопрос номологической достаточности: С является причиной Е, если и только если существует закон, который соединяет события, подобные С, с событиями, подобными Е». Psillos S. Regularity Theories // Beebe H., Hitchcock C., and Menzies C. (eds.). Oxford Handbook of Causation. Oxford, New York: Oxford University Press, 2009. P. 129.



определенных температурах и давлении, выраженная в законах термодинамики. Аналогичная ситуация должна быть и в нашем случае. Но получится ли найти подобную регулярность между ментальными событиями и действиями, обнаружить психофизический закон? Кажется, что такие законы обнаружить, проверить или фальсифицировать проблематично.

Обычно гипотеза о наличии регулярности проверяется путем выделения предполагаемого действующего фактора, многократного воспроизведения его и наблюдения результатов. Если исходное событие является причиной, то ожидаемый результат последует всегда. Но в ситуации с убеждениями и желаниями это сделать практически нельзя: невозможно полностью воспроизвести исходные условия и изолировать ментальную причину. Если действие уже осуществилось, нельзя исключать, что его причиной могли быть иные желания или убеждения. А если результата не последовало в случае воспроизведения желания, вполне вероятно, что появилось другое конфликтующее желание или субъект передумал. Ментальные состояния (желания в том числе) можно всегда совместить с отсутствием действия или с совсем другим действием. Поэтому с ментальными состояниями невозможно проверить наличие регулярности. Вот, к примеру, некто бросился бежать из-за того, что испугался. Страх и желание выжить, возможно, были причиной бегства. Но разве обязательно, что если этот некто испугается в следующий раз, он побежит опять? Вполне вероятно, в следующий раз он замрёт на месте (чтобы его не обнаружили или потому, что потерял способность двигаться) или, наоборот, бросится в атаку на противника.

Тогда, может быть, наличие психофизических законов можно подтвердить на своем примере, с помощью интроспекции? Ведь когда я что-то желаю, я это делаю всегда. Достоверность этого суждения, однако, тоже не очевидна. Возможно, желание присутствовало раньше и повлекло действие, но, возможно, по действию я сделал вывод о собственном желании,

т.е. имела место инверсия. В оценках собственных состояний всегда имеется возможность новой интерпретации. Путем интроспекции такие законы проверить тоже трудно.

Но, может, объективно идентифицировать ментальное состояние и время его возникновения удастся в будущем? Может, это временное ограничение, и оно будет преодолено с развитием науки, появлением инструментов для точного сканирования мыслей? К примеру, в будущем появится возможность «увидеть» и выделить «желание горячего кофе» (или его нейрофизиологический коррелят) и проследить всю последующую каузальную цепочку вплоть до движения рук и разогревания чайника? Существует, однако, аргумент в пользу того, что психофизические законы вообще не могут существовать. Его предложил Д. Дэвидсон. Вывод, который получается в результате этого рассуждения, он называет выводом об «аномальности ментального»<sup>11</sup>.

#### *В. Аргумент психофизической аномальности*

Отличительной чертой ментальных содержаний является рациональность и последовательность. Невозможно допустить, чтобы субъект был убежден в противоречащих вещах: например, в том, что Москва находится на расстоянии 700 км от Петербурга и на расстоянии 1000 км от Петербурга. Даже когда субъект делает противоречивые высказывания, мы находим им непротиворечивые интерпретации. Так, высказывание «я хочу поехать в Москву и не хочу туда ехать» мы склонны объяснять как сменяющиеся во времени состояния или как желание увидеться с родственниками в Москве, но нежелание стоять в бесконечных московских пробках. Высказывание «Мне нравится и не нравится Хёрст» мы не будем интерпретировать как свидетельство противоречащих убеждений, но как выражение того, что «некоторые работы Хёрста ценителю современного искусства нравятся, а некоторые — нет». Итак, требования логической

---

<sup>11</sup> Davidson D. *Mental Events // Essays on Actions and Avents*. Oxford, New York: Oxford University Press, 2001. P. 207-226.

непротиворечивости и последовательности являются обязательными для ментальных содержаний. Но подобных ограничений не существует для физических событий. Из этого следует, что области ментального и физического не могут быть связаны между собой строгими законами. А если они не связаны законами, следовательно, согласно номологической модели каузальности, они не могут стоять в каузальной связи.

Для доказательства допустим обратное: предположим, что мы установили, что состояние мозга X связано с убеждением У, т.е. выявили такой-то психофизический закон. Тогда мы могли бы устанавливать и воспроизводить каждое отдельное убеждение без учета его соотношения с другими желаниями и убеждениями, даже противоречащими ему. Но это невозможно. Вероятно, мы столкнулись с противоречием, а значит, психофизические законы невозможны. Может быть, это является угрозой и для стандартного каузализма?

Ранее я указывал, что альтернативой каузальной связи может быть отношение тождества, концептуальное отношение. Для того, чтобы опровергнуть каузализм, необходимо показать, что именно такая связь присутствует в отношении ментальных состояний и производимых ими действий.

### *С. Аргумент логической связи*

Можно попытаться показать, что ментальные содержания находятся в концептуальной связи с действиями, и таким образом продемонстрировать, что объяснения с помощью ментальных содержаний не указывают на два различных следующих друг за другом события (причину и действие), а просто расшифровывают, рационализируют действие. В этом заключается суть Аргумента логической связи (АЛС). Считается, что «это самый значимый аргумент против каузализма»<sup>12</sup>. Его отстаивали Г. Райл и Э.

---

<sup>12</sup> Stoecker R. Davidson // O'Connor T. and Sandis C. (eds). A Companion to the Philosophy of Action. Oxford: Blackwell, 2010. P. 603.

Энском<sup>13</sup>. Аргумент удобней продемонстрировать на примерах. Человек издали машет руками. Если спросить, почему он это делает, он ответит: «я размахиваю руками, чтобы привлечь внимание». Очевидно, что в этом случае движения руками и привлечение внимания — это не разные события, а одно и то же, только с разными описаниями. Второе описание является объяснением первого потому, что указывает на значение первого действия, а не на его предшествующие состояния. (Это описание из более широкой перспективы, схватывающее более общий паттерн поведения.) Аналогичным образом, желания и убеждения, фигурирующие в нашем случае, возможно, являются не предшествующими событиями, а значениями действий. И логически они неотделимы от самого действия, в то время как причина должна быть логически независимой от следствия. Вот еще пример: гость за столом просит вас передать соль. Вы протягиваете ему солонку. Почему вы это делаете? Очевидно, вы поняли просьбу и хотите передать гостю солонку. Но желание передать солонку, разве оно отделимо от действия? Нельзя сказать, что вы передаёте солонку без желания это делать (ведь никто вас не заставляет передавать соль, и это действие — не автоматизм), но вряд ли можно выделить желание в отдельный процесс. Если бы это был отдельный процесс, были бы способы его отдельно зафиксировать. Но как узнать о наличии подобного желания в отрыве от действия? Представляется абсурдным искать какое-то особенное ментальное состояние, «тягу передать соль». Каким образом эта тяга могла бы проявляться, кроме как в передаче соли? Некоторым из тех, кто размышлял над подобными вопросами, ответ кажется очевидным. По их мнению, действие и желания действия логически связаны. И все же это далеко не окончательный вывод. Ключевым оппонентом этого аргумента является Д. Дэвидсон.

---

<sup>13</sup> Подробно аргумент разбирается в статье Дж. Оттена «Возвращение к жизни аргумента логической связи». Otten J. Reviving the Logical Connection Argument // Canadian Journal of Philosophy. 1977. No.4. P. 725-743

## 5. Доводы Д. Дэвидсона за каузальную связь

В статье «Действия, мотивы и причины» Дэвидсон защищает положения, которые до него считались несовместимыми. Первым из них является утверждение, что объяснения действий с помощью ссылок на ментальные состояния являются каузальными объяснениями. Вторым положением является утверждение о том, что у ментальных состояний, объясняющих действия, есть рационализирующая функция, то есть что ментальные состояния раскрывают мотивы действий. Для целей доказательства объясняющие ментальные состояния Дэвидсон называет «основными мотивами» (primary reasons). Философ утверждает: «1. Для того, чтобы понимать, как какой-либо мотив рационализирует действие, необходимо и достаточно, чтобы мы понимали (как минимум в самом общем виде) из чего состоит основной мотив. 2. Основной мотив действия является его причиной»<sup>14</sup>. Составом основных мотивов или объясняющими ментальными состояниями автор называет пару из а) стремления или склонности и б) соответствующего убеждения. Примером такой пары в случае действия «кипячение воды» служит желание чая и убеждение, что кипячение воды позволит приготовить чай. Дэвидсон последовательно защищает оба своих утверждения.

Первое утверждение Дэвидсон обосновывает с помощью анализа серии примеров. Так, он прибегает к следующему примеру: я переключил выключатель, включил свет и осветил комнату, не зная, что тем самым информировал грабителя о том, что я дома. Философ замечает, что включение света — это действие, но информирование грабителя — нет. У действия есть способ рационализации с помощью основного мотива: желания осветить комнату и убеждения, что переключение выключателя приведет к этому. Информирование грабителя не имеет такой рационализации. Дэвидсон замечает, что иногда для рационализации используется только мотивирующее желание. Он считает, что в таких случаях мотивирующее

---

<sup>14</sup> Davidson D. *Essays on Actions and Events*, Oxford: Clarendon Press, 1980. P. 4.

убеждение просто подразумевается. Знание основного мотива рационализирует действие, так как оно показывает его телеологическую направленность. «В свете основного мотива действие предстает соответствующим определенным качествам, краткосрочным или долгосрочным целям, характеристикам агента, и агент представляется в качестве Рационального Животного»<sup>15</sup>. Но на каком основании автор считает, что основные мотивы являются причинами действий?

Ссылка на основной мотив является способом объяснения действия. Это объяснение рационализирует действие, показывая, что оно соответствует интенциям агента. Но это не означает, что объяснение с помощью ссылки на основной мотив является только рационализацией. Более того, есть основание считать, что объяснение с помощью ссылки на основной мотив является также каузальным объяснением. Таким основанием служит гипотетическая возможность случаев, в которых есть основной мотив, рационализирующий действие, есть действие и, тем не менее, это действие выполняется по другой причине. «Сущностью отношения между действием и основным мотивом, который его объясняет, служит то, что основной мотив обосновывает поступок для агента»<sup>16</sup>.

Для примера возьмем случай, когда человек А случайно получает перевод средств, предназначенный человеку Б. А сообщает банковским служащим о том, что произошла ошибка<sup>17</sup>. Основным мотивом является желание вернуть деньги тому, кому они предназначались. Однако может быть такая ситуация, что существует еще один основной мотив: желание не попасть под налоговую проверку, которая обнаружит незаконное поступление средств. Оба желания, или обе пары желаний и убеждений могут быть рационализацией действия А, но только один из основных мотивов может быть причиной. Именно этот мотив будет основанием

---

<sup>15</sup> Davidson D. *Essays on Actions and Events*, Oxford: Clarendon Press, 1980. P. 8.

<sup>16</sup> Davidson D. *Essays on Actions and Events*, Oxford: Clarendon Press, 1980. P. 9.

<sup>17</sup> Lepore E., Ludwig K. *Introduction // The Essential Davidson*. Oxford: Oxford University Press, 2006. P. 12.

поступка. Так Дэвидсон обосновывает второе положение. Для обоснования стандартного каузализма, впрочем, необходимо ответить на критику с помощью уже обозначенных аргументов. Во-первых, нужно показать, в чем заключается слабое место Аргумента логической связи. И, во-вторых, необходимо продемонстрировать, как соотносится стандартный каузализм с собственными выводами Дэвидсона об отсутствии психофизических законов.

Дэвидсон предлагает несколько ответов на АЛС. Во-первых, он указывает на то, что, например, истинность высказывания «я хотел передать соль» вовсе не влечёт за собой истинность высказывания «я передал соль». Я мог хотеть, но не передать. Между этими событиями нет необходимого следования. Во-вторых, что более важно, событие, которое определяет истинность высказывания «я передал соль», не является интенциональным объектом в высказывании «я хотел передать соль». Ведь если я передал соль, то сделал это каким-то определённым образом, в какой-то конкретный момент. А желание могло быть реализовано почти бесконечным числом конкретных физических действий. Я мог бы передать соль сидя или стоя, правой или левой рукой, из рук в руки или бросив солонку гостю через весь стол. Учитывая такие различия желаний и реализующих действий, оправдано утверждать их логическую независимость и допустимость использования первых для объяснения вторых<sup>18</sup>.

Дэвидсон считает, что логическая зависимость не существует между желаниями и действиями, но может существовать между некоторыми их описаниями. Как, например, в случае с высказыванием «причина пожара стала причиной пожара». Но всегда можно дать другие описания. Можно найти описания ментального без логической связи с действием. Как, например, в предложении «разожженный костер стал причиной пожара». Стандартной языковой практикой является описание события с помощью указания на причину. Но «описывать событие в терминах его причины — не

---

<sup>18</sup> Davidson D. Action, Reasons, and Causes // The Essential Davidson. Oxford: Oxford University Press, 2006. P. 25.

то же самое, что объединять событие и его причину... объяснение с помощью такого описания не исключает каузальное объяснение<sup>19</sup>. То есть можно «получение травмы» описать как «лучевой ожог». Но, по мнению Дэвидсона, из практики подобных описаний не следует то, что они указывают на одно и то же событие. Таким же образом, по его мнению, из отсутствия психофизических законов не следует, что между основными мотивами и действиями нет каузальной связи.

Дэвидсон, как я уже писал, склоняется к тому, что каузальная связь должна быть частным случаем общего закона. Он также обращает внимание, что каузальная связь между мотивами и действиями в силу специфики мотивов действительно не может быть частным случаем какого-то закона. Но он предполагает, что мотивы могут быть описаны с помощью каких-то других дескрипций, и эти дескрипции могут быть описаниями частного случая закона. «Законы, существование которых необходимо, если мотивы являются причинами действий... не должны быть сформулированы в терминах рационализаций. Из того, что причины класса событий (действий) попадают в определенный класс (мотивов) и существует закон, объясняющий каждое высказывание о каузальности, не следует, что существует закон, соединяющий события, классифицированные как мотивы, с событиями, классифицированными как действия — классификация может быть в неологических, химических или физических терминах»<sup>20</sup>.

В качестве аналогии Дэвидсон предлагает рассмотреть метеорологический прогноз. Стандартные описания погоды включают в себя очень общие описания: «завтра во второй половине дня будут осадки, порывистый ветер, переменная облачность». Эти описания не могут фигурировать в законах метеорологии, так как не являются достаточно точными и простроены на других концепциях. Однако это общее описание

---

<sup>19</sup> Davidson D. *Essays on Actions and Events*, Oxford: Clarendon Press, 1980. P. 13.

<sup>20</sup> Davidson D. *Essays on Actions and Events*, Oxford: Clarendon Press, 1980. P. 17.



может быть переформулировано с большей точностью и с помощью других концепций, и тогда оно будет частным случаем общего закона. Таким образом, Дэвидсон снимает противоречие между Аргументом психофизической аномальности и стандартным каузализмом и обосновывает последний. Стандартный каузализм, в свою очередь, обосновывает необходимость ментальной каузальности в качестве основы действия. Впрочем, есть и другие аргументы в пользу неё.

Без ментальной каузальности невозможны объяснение и рационализация многих социальных практик, в частности применение моральной ответственности. Люди часто считают других людей виновными в каких-то действиях или, наоборот, благодарят за действия других. Они сердятся, возмущаются или испытывают благодарность в отношении других. За свои поступки люди также испытывают гордость или чувство вины. Эти реакции определяются в современной аналитической философии как проявление моральной ответственности и результат признания свободы воли. Однако эти реакции были бы неуместными, если бы ментальная каузальность отсутствовала. Современный метафизик Дж. Ким пишет: «Практика применения моральных категорий очевидным образом предполагает, что наши ментальные состояния имеют каузальный эффект в физическом мире»<sup>21</sup>. Иначе на каком основании требовать от субъектов соответствия норме поведения, если поведение не связано с их ментальными состояниями? На каком основании благодарить или требовать ответственности, если желания и убеждения не играют каузальной роли? Как может быть свободен агент, если свобода проявляется в действии, но действие каузально не определено его ментальными состояниями: желаниями и убеждениями? И как агент может нести ответственность за действия, если он является пассивным наблюдателем по отношению к ним?

---

<sup>21</sup> Kim J. *Physicalism, or Something Near Enough*. Princeton: Princeton University Press, 2008. P.9.

Да и само понятие агента, разве не станет оно пустым без ментальной каузальности, ведь агент — это тот, кто имеет способность поступать в соответствии с собственным разумом. Если нет агента, нет и действия, ведь, как мы заметили в самом начале, связь с агентом и его ментальными состояниями отличает просто процессы от действий. Отрицая ментальную каузальность, возможно, придется вообще отрицать разум и ментальное, ведь, само существование чего-то трудно представить без наличия у него каузальной силы. Это утверждается в так называемом принципе элеатов. Согласно этому принципу, «нечто является реально существующим только в случае, если оно может быть участником каузального процесса»<sup>22</sup>. Поэтому вопрос о ментальной каузальности представляется в большей степени не эмпирическим вопросом, существует она или нет, а философским: как возможно её существование.

Относительно важности ментальной каузальности категорически высказался другой современный аналитический философ, Дж. Фодор: «Я не полностью убежден, насколько важно, что ментальное — это физическое; и ещё меньше в том, что важно, чтобы мы могли это доказать. Однако если неверно то, что мое желание каузально отвечает за моё действие, щекотка за чесотку, мое убеждение за мои выражения... если все это неверно, тогда все, в чем я убежден, — ложь, и это конец света»<sup>23</sup>. Фодор справедливо заметил, что от ментальной каузальности зависит истинность человеческого знания. Ведь переход от восприятия к убеждению и от одного убеждения к другому невозможно помыслить без каузальных связей между ментальными состояниями. С Фодором согласны и большинство российских философов. Так, по данным опроса Московского Центра Исследования Сознания, 92% считают сознание каузально эффективным<sup>24</sup>. С этим согласен и я. Однако это

---

<sup>22</sup> Принцип сформулирован по статье Colyvan M. Can the Eleatic Principle Be Justified? // Canadian Journal of Philosophy. 1998. Vol. 28. No. 3. P. 313–336.

<sup>23</sup> По книге Kim J. Philosophy of Mind. Cambridge MA., 2006. P. 181.

<sup>24</sup> Материалы опроса философского сообщества готовятся к публикации Московским центром исследования сознания.

согласие ведет к проблеме сочетания ментальной каузальности и физикализма. В современной философии эта проблема выявляется в связи с *Аргументом исключения*.

## II. Проблема ментальной каузальности в современной аналитической философии

### 1. Аргумент исключения Дж. Кима против ментальной каузальности

Аргумент исключения, судя по количеству комментариев и ответных статей, — один из наиболее значимых аргументов современной аналитической философии в области ментальной каузальности. Первые версии его были разработаны Ч. Тэйлором (1964)<sup>25</sup> и Н. Малкольмом (1968)<sup>26</sup>. В конце 80-х он был актуализирован американским метафизиком Дж. Кимом, а затем подробно разработан им же в статье «Каузальность вниз, эмерджентизм и нередуктивный физикализм» в 1992 г. Как пишет Кузнецов А. в своей диссертации, «Ким — философ, вокруг которого строятся дискуссии о проблеме ментальной каузальности»<sup>27</sup>. В наиболее поздней монографии «Физикализм или что-то около того» Ким представляет обновленную версию аргумента, вписанную в более широкую теорию отношения ментального и физического<sup>28</sup>.

Рассуждение Кима основано на совмещении нескольких вполне правдоподобных принципов. Это три утверждения: принцип каузальной исключительности (ПКИ), принцип каузальной замкнутости физического мира (ПКЗ) и принцип отсутствия систематической сверхдетерминации (ПОСС). Принцип каузальной исключительности фиксирует, что в стандартной ситуации причина события может быть только одна:

---

<sup>25</sup> Taylor C. *Explanation of Behavior*. London: Routledge and Kegan, 1964.

<sup>26</sup> Malcolm N. *The Conceivability of Mechanism* // *Philosophical review*, 77, 1968. P. 45-72.

<sup>27</sup> Кузнецов А. В. *Проблема ментальной каузальности в аналитической философии сознания*. Диссертация на соискание степени кандидата философских наук: 09.00.03 — История философии. М., 2016. С. 52

<sup>28</sup> Kim J. *Physicalism, or Something Near Enough*. Princeton: Princeton University Press, 2008. P.12.

*(ПКИ) Если событие  $e$  имеет достаточную причину  $c$  в момент  $t$ , никакое событие в  $t$ , отличающееся от  $c$ , не может быть причиной  $e$  (кроме случаев свехдетерминации).*

Согласно этому принципу утверждается, что достаточная причина у события должна быть одна, кроме специальных случаев свехдетерминации. Под случаями свехдетерминации имеются в виду случаи, когда событие могло произойти по любой из нескольких не связанных между собой причин. К примеру, смерть от расстрела стрелковым подразделением — это свехдетерминированное событие, т.к. попадание каждой пули в жизненно важные органы является смертельным и пули двигаются независимо друг от друга. Принцип каузальной исключительности — для многих это установка по умолчанию. Она может даже казаться аналитической истиной. Большая часть исследователей среди *relata* причинно-следственной связи насчитывают одну достаточную причину и одно следствие. Если причина достаточная, то она будет одна.

Второй важный принцип — это принцип каузальной замкнутости физического мира (ПКЗ). На этом принципе базируется научное знание, в частности понятие физического закона. И несмотря на то, что существуют отдельные любопытные попытки оспорить этот принцип, он все-таки представляет правдоподобное положение. Согласно ПКЗ:

*(ПКЗ) Каждое физическое событие в момент  $t$  имеет в качестве своей достаточной причины другое физическое событие в момент  $t$ .*

Ссылка на конкретную отметку по временной шкале нужна Киму для того, чтобы исключить конфликт причин, разнесенных во времени. Одно и то же событие может быть результатом нескольких следующих друг за другом событий. В этом виде, согласно ПКЗ, все процессы и действия — вспышка сверхновой, извержение вулкана, взлёт беспилотника и кипение воды в чайнике — в равной степени обусловлены предшествующими физическими

событиями. Чтобы объяснить любые события, достаточно только физических причин, в природе нет места чудесам.

Впрочем, этот принцип можно немного адаптировать, ведь кто-то может допускать беспричинные события или случайности. Вместо того чтобы спорить по этому поводу, здесь можно просто несколько ослабить этот принцип, представив его в таком виде:

*(ПКЗ<sup>1</sup>) Если физическое событие имеет причину в момент  $t$ , оно имеет достаточную физическую причину в момент  $t$ .*

В таком виде принцип допускает беспричинные события (как и предыдущий принцип), утверждает лишь наличие физических причин у событий с причинами. И он не должен вызывать сомнения.

И, наконец, последний интересующий нас принцип — принцип отсутствия систематической сверхдетерминации

*(ПОСС) Не существует систематической сверхдетерминации физических событий.*

Я уже разъяснил, что имеется в виду под сверхдетерминацией, и привёл пример, когда она имеет место. Но подчеркну, что сверхдетерминация — это особенный случай. Систематическая, регулярная сверхдетерминация представляется невероятной. Одно дело — единичный случай наступления смерти в результате расстрела стрелковым подразделением, а другое — это утверждать, что всегда, когда кто-то погибает от пуль, стреляющих как минимум двое. Второе кажется невероятным.

Итак, что получается из соединения этих принципов? По мнению Кима, ПКЗ, ПОСС и ПКИ в сочетании образуют серьезную трудность для интерпретации ментальной каузальности при условии, что ментальные состояния не отождествляются с физическими процессами. Ход рассуждений

философа таков. Допустим, ментальное событие  $M$  является причиной физического события, действия  $P^*$ . Обозначим это  $M \rightarrow P^*$ . При этом, согласно ПКЗ, у события  $P^*$  есть также достаточная причина  $P$  — другое физическое событие, некоторое состояние мозга, одновременное с событием  $M$ . Т.е.  $P \rightarrow P^*$ . Если  $M \neq P$ , то у  $P^*$  две причины: ментальная и физическая. Т.е.  $P$  — сверхдетерминированное событие. Иными словами, каждое из этих предшествующих событий достаточно для реализации события  $P$ . Но, согласно ПКЗ, физическая причина, предшествующее состояние мозга, обязательна, а ментальная — нет: согласно условию, мы имеем дело с физическим событием, у которого есть причина, а значит, у него должна быть достаточная физическая причина (согласно ПКЗ), тогда как наличие у него ментальной причины не обязательно. Значит, событие  $P^*$  произошло бы, даже если бы  $M$  отсутствовало. Следовательно, ментальные события не вносят никаких изменений в ход дел. Они эпифеноменальны, т.е. фактически не имеют каузальной силы. События были бы такими же, если бы ментальных состояний не было. «Очевидный вывод этого аргумента в том, что каузальное действие ментального на физическое является иллюзией; оно никогда не происходит. Это эпифеноменализм, как минимум в отношении физических событий. Это не исключает влияния одних ментальных событий на другие ментальные события. Но если ментальные события, как убеждения и интенции, никогда каузально не влияют на движения тела, это лишает агента возможности существования...»<sup>29</sup>.

Конечно, чтобы попытаться спасти ментальную каузальность, можно допустить тождество  $M$  и  $P$ :  $M = P$ . Но это жертва очень значительная. На первый взгляд, мы спасем ментальную каузальность. Но при этом придется отрицать особый статус ментального и его специфические свойства, в том числе возможность множественной реализации. А это — тоже общепризнанный тезис. Ментальные свойства могут быть реализованы в

---

<sup>29</sup> Kim J. *Philosophy of Mind*. Cambridge, Massachusetts: MIT Press, 2006. P. 217.

разных случаях и у разных индивидов разными физическими состояниями. К примеру, ментальное свойство «быть болью» может быть реализовано возбуждением различных нервных окончаний и активацией разных нейронных сетей. Таким образом, Аргумент исключения фактически ставит перед нами дилемму: либо ментальная каузальность невозможна и ментальные состояния в лучшем случае эпифеноменальны, либо они тождественны физическим процессам, а значит, специфических ментальных свойств не существует. Обе альтернативы крайне нежелательны. Эта дилемма и представляет суть проблемы ментальной каузальности.

## 2. Аргумент супервентности против каузального влияния одних ментальных состояний на другие

Впрочем, в приведенных рассуждениях все еще присутствует открытый ход. Выше был рассмотрен случай каузального влияния ментального на физическое. Аргумент исключения приводит к выводу, что это влияние невозможно. Но, может быть, ментальное хотя бы способно влиять на ментальное, т.е. одни ментальные состояния могут быть причиной других? Если одни ментальные состояния влияют на другие, справедливым будет, по крайней мере, сказать, что убеждения могут основываться на воспоминаниях, желания — на убеждениях, а выводы — на умозаклучениях. Это хотя бы отчасти даст основания к утверждению существования и действенной силы разума.

К сожалению, и здесь нас ожидает препятствие — еще один аргумент Кима — Аргумент супервентности (АС). Он строится на уже знакомых принципах ПКЗ, модификации ПКИ и Тезисе о супервентности (ТС). Этот дополнительный тезис звучит так:

*(ТС) Все, что происходит в ментальной жизни, полностью определено физическими процессами.*

Тезис о супервентности представляет собой ядро физикалистских позиций, он разделяется практически всеми, кто придерживается доминирующих сегодня философских взглядов. С ним согласны сторонники редукционистского физикализма (например, последователи теории тождества), функционалисты и даже эмерджентисты. В основе ТС лежит представление о том, что полностью идентичные в физическом смысле системы будут полностью идентичны и в плане ментальных состояний. ТС указывает на то, что ментальное не является независимым и не существует автономно от физического, а, наоборот, супервентно, т.е. зависимо и определено базовыми физическими свойствами. Вот что получается, когда мы анализируем каузальное влияние ментального на ментальное с учетом Тезиса о супервентности:

Предположим, ментальное событие  $M$  является причиной другого ментального события  $M^*$ . Согласно ТС, однако, ментальное событие  $M^*$  в качестве своей базы имеет физическое событие  $P^*$ , ведь ментальное не существует автономно. Теперь применим ПКИ, точнее вариант ПКИ. В исходном виде принцип не совсем подходит, ведь он относится к случаям каузальной связи. Но его можно обобщить до любых генерирующих (определяющих) связей, в частности отношения супервентности. Тогда он будет звучать так:

*(ПКСИ) Если событие  $e$  определено событием  $s$  — каузально или отношением супервентности — тогда событие  $e$  не может быть определено другим, отличающимся от  $s$  событием, кроме случаев сверхдетерминации.*

Применив этот принцип, получим: если событие  $M^*$  имеет достаточное основание в событии  $s$ , это и есть его единственное основание (кроме случаев сверхдетерминации). Но которое из двух перечисленных оснований,  $M$  и  $P^*$ , является достаточным?  $M$ , вероятно, детерминирует  $M^*$  горизонтально, с течением времени, а  $P^*$  — вертикально, как база для



супервентного свойства. Известно, что, случись  $P^*$ ,  $M^*$  произошло бы с необходимостью. Тогда достаточной базой является  $P^*$ . То, что мы наблюдаем в этом случае, сам Ким называет максимой Эдвардса (в честь американского философа и теолога Джонатана Эдвардса): «Существует конфликт между вертикальной детерминацией и горизонтальной каузальностью. Фактически вертикальная детерминация исключает горизонтальную каузальность»<sup>30</sup>. Тогда остаётся лишь один вариант влияния ментального на ментальное —  $M$  может стать причиной  $M^*$ , став предварительно причиной базового события  $P^*$ . Единственный способ каузально влиять на супервентное свойство — только через каузальное влияние на его базу. Но это шах, ведущий к мату. Ведь это случай каузального влияния ментального на физическое. А в отношении возможности подобного влияния был уже вынесен отрицательный вердикт в связи с Аргументом исключения. Круг замкнулся, и теперь мы в полной мере наблюдаем трудность философской проблемы ментальной каузальности.

Ментальные состояния не могут казуально влиять не только на физические состояния, но и на другие ментальные состояния. Кажется, общепринятые взгляды на ментальное и его действенность непоследовательны. И попытки совместить веру в ментальную каузальность с важными метафизическими и логическими принципами приводят нас к патовой ситуации. Дж. Ким так обобщает эту ситуацию: «Следует заключить тогда, что психофизическая проблема для нас, желающих быть физикалистами, разделяется на две проблемы: проблему ментальной каузальности и проблему сознания, и они вместе представляют самый большой вызов для физикализма. Если физикализму суждено уцелеть в качестве нашего мировоззрения, он должен показать, каково наше место в физическом мире, и это означает, что он должен показать наш статус как сознательных существ, обладающих возможностью влиять на окружающую

---

<sup>30</sup> Kim J. Physicalism, or Something Near Enough. Princeton: Princeton University Press, 2008. P.35.

среду благодаря сознанию и ментальности. Аргументы, представленные выше, уже показали, что физикализму не удастся сохраниться без изменений...»<sup>31</sup>.

За последние несколько десятилетий вопросу о соотношении ментального и физического (проблема «сознания») и проблеме ментальной каузальности были посвящены многие сотни работ. Значительная их часть написана философами. Но растёт также количество исследований, предпринятых учёными-когнитивистами, нейрофизиологами, психологами, врачами, физиками. Философия, кажется, теряет эксклюзивные права в этой области. Но среди исследователей остается что-то вроде общего консенсуса: физикалистская позиция продолжает преобладать. Это позиция, согласно которой всё в окружающем нас мире в конечном итоге состоит из физических объектов и зависит от физических свойств. С появлением новых книг растёт количество аргументов, и авторы с критикой набрасываются на тексты своих коллег, так что у любопытствующей публики создаётся полное ощущение горячего спора. Но если в действительности спор и имеет место, этот спор, скорее, о деталях. Оригинальность авторов обычно выражается в способе аргументации, но не в явной новизне позиции. Одним из редких исключений являются тексты нашего коллеги, российского философа В.В. Васильева. Он предлагает провести ревизию физикализма.

### III. Решение проблемы ментальной каузальности в локальном интеракционизме

#### 1. Ключевые положения локального интеракционизма

По нескольким статьям и трем монографиям<sup>32</sup>, опубликованным в последнее время, можно проследить оригинальную философскую позицию,

---

<sup>31</sup> Kim J. Physicalism, or Something Near Enough. Princeton: Princeton University Press, 2008. P.31.

<sup>32</sup> Васильев В.В. Трудная проблема сознания. М.: Прогресс-Традиция, 2009; Васильев В.В. Сознание и вещи: очерк феноменалистической онтологии. М.: УРСС, 2014; Васильев В.В. В защиту классического компатибилизма: Эссе о свободе воли. М: ЛЕНАНД, 2017.

которую защищает российский философ Васильев. Автор называет её «*локальным интеракционизмом*». В соответствии с этой позицией утверждается специфический, нередуцируемый характер ментального. С точки зрения автора, квалиа, качественные характеристики воспринимаемого во внутреннем опыте, не могут быть сведены к физическому, к состояниям мозга. При этом, по его представлениям, квалиа не просто сопутствуют физическим процессам. Они не эпифеноменальны, не являются «номологическими бездельниками», а реально влияют на поведение. Без них мозг не мог бы работать так, как он работает, поэтому существование квалиа эволюционно оправданно. Но как тогда быть с принципом каузальной замкнутости физического? Достоинством и уникальной особенностью локального интеракционизма является попытка совместить тезис о ментальной каузальности и некоторый вариант ПКЗ. Положения локального интеракционизма можно резюмировать в следующих тезисах:

*Онтологический статус сознания:* Ментальные состояния существуют как субъективные переживания и нередуцируемы к физическим.

*Ментальная каузальность:* Ментальные состояния каузально влияют на поведение.

*Связь с состояниями мозга:* Они не могут быть полностью обусловлены только процессами в мозге.

*Каузальная замкнутость:* Принцип каузальной замкнутости не работает в некоторых локальных объектах, но сохраняется глобально в мире.

Рассмотрению этой позиции целиком следует посвятить отдельное исследование. Я же попытаюсь в этой главе разобрать один из интереснейших аргументов, который изобретает автор «Трудной проблемы сознания» — Аргумент каузальных траекторий (АКТ). Это рассуждение было разработано Васильевым в 2004г. при размышлении над книгой

«Сознающий ум» Д. Чалмерса<sup>33</sup>. На этом аргументе основывается важная часть теории локального интеракционизма и решение проблемы ментальной каузальности.

## 2. Аргумент каузальных траекторий за ментальную каузальность и против локальной супервентности

Аргумент сформулирован в виде трех посылок и вывода. Первая посылка — это утверждение о том, что (1) одинаковые события могут иметь разные причины. Вот шар попал в лузу, он мог там оказаться самыми различными способами, в результате самых различных соударений и движения по множеству траекторий. Как, впрочем, любая вещь могла оказаться там, где она есть, множеством различных способов. Посылка выглядит бесспорной. Вторая посылка — это констатация того, что (2) память отражает события, происходившие с субъектом в прошлом, т.е. отражает каузальную историю субъекта. Конечно, память может представлять каузальную историю ошибочным образом, но это, скорее, исключение, чем норма<sup>34</sup>. Столь же несомненно, по мнению российского философа, что воспоминания представляют собой приватные качественные состояния и образуют основу убеждений, желаний и других осознанных состояний субъекта. Третья посылка указывает на корреляцию поведения и желаний. Эта посылка не отрицается даже эпифеноменалистами. Эпифеноменалист, правда, считает, что корреляция не имеет отношения к причинно-следственной цепи. А вот представленный аргумент заставляет нас склониться к обратному.

Объединяя посылки, В. Васильев получает следующую картину: «Если реально, что одно и то же событие или комплекс событий могли бы порождаться разными причинами и если мой мозг в данный момент времени может рассматриваться как громадный комплекс физических событий

---

<sup>33</sup> Васильев В.В. Трудная проблема сознания. М.: Прогресс-Традиция, 2009; Васильев В.В. Сознание и вещи: очерк феноменалистической онтологии. М.: УРСС, 2014. С. 216.

<sup>34</sup> Васильев В.В. Трудная проблема сознания. М.: Прогресс-Традиция, 2009; Васильев В.В. Сознание и вещи: очерк феноменалистической онтологии. М.: УРСС, 2014. С. 217.

(некоторые составляющие которого могли бы иметь и иные причины), то очевидно, что он мог оказаться в данном состоянии в результате самых разных предшествующих событий. Иными словами, мой мозг... мог иметь иную каузальную историю, нежели та, что действительно была у меня и составляла мою жизнь. Но если бы у меня была иная история, то, как следует из второй посылки, я обладал бы сейчас другими воспоминаниями и эти другие воспоминания фундировали бы иные убеждения и желания. Поскольку же мои убеждения и желания... скоррелированы с моим поведением (третья посылка), то, обладая другими убеждениями и желаниями, я вел бы себя иначе, чем веду себя в данный момент...»<sup>35</sup>.

Итак, из аргумента следует, что, находясь в том же самом физическом состоянии, субъект мог бы вести себя различным образом. И эти различия обуславливаются только его качественными ментальными содержаниями, различными убеждениями и желаниями. И никакой анализ наличного физического состояния не способен дать точного прогноза действий субъекта. Ментальные состояния действенны, а значит, эпифеноменализм ложен. Но опровержение эпифеноменализма, с моей точки зрения, не самое критичное из следствий Аргумента каузальных траекторий. Гораздо более значимым мне кажется опровержение того, что является общим местом современных теорий сознания — тезиса о супервентности ментальных содержаний на состояниях мозга (ТС).

Понятие супервентности обозначает зависимость одних свойств или фактов от других. Если свойства типа *A* супервентны на свойствах типа *B*, то два объекта не могут различаться в свойствах *A* без различий в свойствах *B*. В современный философский дискурс это понятие ввёл уже упомянутый Д. Дэвидсон в знаменитой статье «Ментальные события» (1970): «Хотя согласно позиции, которую я занимаю, отрицается существование психофизических законов, она совместима с мнением, что ментальные

---

<sup>35</sup> Васильев В.В. Трудная проблема сознания. М.: Прогресс-Традиция, 2009. С. 217-218.

характеристики в некотором смысле зависят, или супервентны на физических характеристиках. Здесь супервентность означает невозможность двух событий, не отличающихся в физическом смысле, но отличающихся в ментальном отношении, невозможность изменения объекта в ментальном отношении без изменений в физическом отношении»<sup>36</sup>. Несмотря на то, что Дэвидсон защищает свой вариант решения проблемы сознания, тезис о супервентности ментального на физическом является распространенным. И его принятие могло бы служить критерием минимального физикализма. Его разделяют большинство философов: дуалисты свойств, эмерджентисты, функционалисты, сторонники теории тождества, элиминативисты, ведь ТС — это не решение вопроса о сознании и ментальной каузальности, а современная его постановка: как объяснить ментальное (сознание), если оно супервентно на физическом? Тем важнее оказывается аргумент, который подвергает этот тезис сомнению, отрицает супервентность. Впрочем, здесь необходимо сделать некоторые уточнения.

Возможны как минимум три формулировки тезиса о супервентности. (1) Ментальное супервентно на физическом в том случае, если верно, что если нечто  $X$  имеет ментальное свойство  $M$ , то имеется физическое свойство  $P$  такое, что  $X$  обладает  $P$ , и необходимо, что всякий объект, обладающий  $P$ , наделен  $M$ . (2) Ментальное супервентно на физическом в том смысле, что объекты, совершенно одинаковые в своих физических свойствах, не могут отличаться относительно ментальных свойств. (3) Ментальное супервентно на физическом в том смысле, что миры, одинаковые во всех физических отношениях, одинаковы и во всех ментальных отношениях; т.е. миры, одинаковые физически, совершенно одинаковы в целом<sup>37</sup>. Сила этих утверждений различна. Наиболее сильным является утверждение (1), т.к. оно фиксирует супервентность даже одного изолированного свойства. А

---

<sup>36</sup> Davidson D. *Mental Events // The Essential Davidson*. Oxford, 2006. P. 111.

<sup>37</sup> Варианты тезиса о супервентности приведены по книге Дж. Ким. Kim J. *Philosophy of Mind*. Cambridge, Massachusetts: MIT Press, 2006. P. 9-10.

наиболее слабым — (3), так как в нем утверждается супервентность только миров в целом. Если условия, описанные в утверждении (1), выполняются, то будут выполняться и условия (2). А если выполняются условия по утверждению (2), автоматически будут выполнены и условия по утверждению (3). Определение (3) фиксирует отношение глобальной супервентности, а определения (1) и (2) — локальной. Для примера возьмём свойство «быть банком». Это свойство не имеет локальной супервентности, но имеет глобальную. Так, здание может быть одним и тем же, но является ли это здание банком или, скажем, музеем, зависит от того, какой статус оно имеет и как используется обществом. Аналогично, свойство «быть оригинальной картиной Ван Гога» также не имеет локальной супервентности. Пусть 1000 картин не отличаются друг от друга ни на один атом, тем не менее одна из них может быть подлинной картиной Ван Гога, а другие — точными копиями. Это зависит не только от самой картины, но и от внешних по отношению к ней факторов. Итак, как соотносятся 3 варианта тезиса с супервентности с выводом Аргумента каузальных траекторий?

Очевидно, реальная возможность существования двух идентичных людей с разными ментальными состояниями, допускаемая в качестве результата аргумента, противоречит варианту (1) ТС. Но сам Васильев сообщает, что его выводы не противоречат «глобальной супервентности»: «последняя формулировка выражает так называемую "глобальную супервентность", и мы увидим, что предложенные мною аргументы не отрицают её. Вторая формулировка тоже не опровергается ими, если она допускает глобальную интерпретацию. Ошибочно лишь первое понимание»<sup>38</sup>. Что же получается? Перед нами — неожиданная и удивительно красивая картина мира. Ментальные состояния отличаются от физических. При этом они не эпифеноменальны, а действенны. Они влияют на поведение. Но это влияние не разрушает законы природы, в том числе

---

<sup>38</sup> Васильев. В. В. Трудная проблема сознания. М.: Прогресс-Традиция, 2009. С. 243.

фундаментальный закон сохранения энергии. Ментальное является посредником, условием влияния нелокальных физических факторов, а локально — причиной поведения. Ментальная каузальность защищена, тщетность усилий ученых найти корреляты ментальным состояниям объяснена, ПКЗ сохранён, и... в качестве бонуса получен ключ для решения двух философских проблем: «трудной проблемы сознания» и проблемы тождества личности. Из подобной картины следует, что сознательные состояния существуют потому, что необходимы для генерации человеческого поведения (без них поведение разумного человек было бы просто невозможно). И что тождество личности не связано только лишь с тождеством тела, а с тождеством каузальной истории, зафиксированной в ментальных состояниях. Что ж, если это построение, помимо своих бесспорных эстетических качеств, сможет выдержать критический анализ, то это крупный успех, и даже прорыв в философии. Итак, учитывая значимость выводов, вернёмся к аргументу ещё раз и попытаемся для начала проверить, действительно ли он опровергает тезис о локальной супервентности, сохраняя при этом глобальный вариант. Именно в этом случае он совместим с ПКЗ, вписывается в современную научную картину мира и согласуется с базовыми естественными установками.

#### *А. Локальная или глобальная супервентность?*

Вполне вероятно, что в исходном виде аргумент действительно направлен исключительно против локальной супервентности ментального на физическом. Но есть вероятность, что если аргумент верен, его вариацию можно использовать также и против глобальной супервентности. Тогда рассуждения строились бы следующим образом. (1) У одинаковых событий могут быть разные причины. Таким образом, если текущее состояние мира представляет собой гигантский комплекс событий, эти же события могли произойти в ходе других причин. (2) Память людей обычно отражает события, которые происходили с ними в прошлом. Но если бы у них было другое прошлое, их память имела бы другое содержание. (В) Следовательно,



возможно, что существовал бы мир, полностью идентичный нашему, но отличающийся от нашего содержанием ментальных состояний населяющих его людей. Это следствие противоречит тезису о глобальной супервентности: тезису о том, что миры, одинаковые физически, совершенно одинаковы в целом. Выходит, что аргумент каузальных траекторий направлен и против глобальной супервентности?

Автор предполагает подобного рода замечания и заведомо не согласен с выводом. Скорее всего, он критиковал бы первую посылку в нашем изложении. Хотя одно сингулярное событие (используя терминологию автора) может иметь множество предшествующих событий в качестве причины, весь комплекс событий, составляющий состояния мира в момент  $t$ , мог бы иметь только одно предшествующее состояние — то, которое было в действительности. Т.е. у текущего совокупного события, *События А*, может быть только один антецедент, *Событие В* (*ТСС — тезис о совокупном событии*). Но каковы основания считать именно так?

У сторонника локального интеракционизма есть две тактики защиты этого тезиса. Одна основывается на анализе принципов здравого смысла. По мнению автора, эти положения не могут быть оспорены, т.к. отражают устройство наших познавательных способностей. Одним из базовых положений В.В. Васильев считает убеждение «в тождественности будущего прошлому»<sup>39</sup>. На нём построены другие: вера в существование причин у каждого события, убеждение в том, что «при точном повторении в будущем тех событий, которые имели место в прошлом, за ними последует то, что следовало за ними тогда, когда мы раньше ощущали их»<sup>40</sup>, принцип каузальной замкнутости физического. По мнению автора, аксиома тождества прошлого и будущего обуславливает также веру в возможность существования только одной причины, только одного События А,

---

<sup>39</sup> Васильев. В.В. Трудная проблема сознания. М.: Прогресс-Традиция, 2009. С. 206.

<sup>40</sup> Васильев. В.В. Трудная проблема сознания. М.: Прогресс-Традиция, 2009. С. 208.

предшествующего Событию В. Мне кажется, что следование здесь далеко не очевидно. Но в данном случае я не берусь оценивать умозаключения Васильева, т.к. метод философствования, который применяет здесь автор, частично выходит за рамки аналитического дискурса и представляет собой вариант аналитической феноменологии, довольно редко использующийся современными философами. Впрочем, к выводам автора проложен и другой путь, который проще проследить и проверить. Он представлен в другой книге В. Васильева, «Сознание и вещи»<sup>41</sup>.

Философ сравнивает два описания события: отскок мяча от поверхности и отскок мяча от поверхности на такое-то расстояние. Оба события могут быть следствиями множества причин. Например, причиной могли быть удар по мячу другим мячом, или теннисной ракеткой, или неустойчивое исходное положения мяча на плоскости. Но второе описание является более конкретным и в этом смысле сужает горизонт возможных причин (предшествующих состояний). Кажется, что конкретизация описания событий уменьшает возможные антецедентные события. Отскок мяча на двести метров мог иметь гораздо меньше причин, чем просто отскок мяча. А состояние мозга в момент  $t$  является еще более «конкретным событием», чем отскок мяча на двести метров, и потому может иметь еще меньшее количество причин. Но явно не одну. Только предельно конкретное событие имеет исключительно одну причину. Таким событием и является Событие, или совокупное содержание всего чувственного опыта в момент  $t$ , — считает автор.

Кажется, эти рассуждения не являются полностью безупречными. Они строятся на представлении о снижении количества возможных причин при конкретизации описания. Но разве они убеждают нас в том, что совокупное событие<sup>42</sup> может иметь только одну причину? Возьмем сингулярное событие

---

<sup>41</sup> Васильев В.В. Сознание и вещи: очерк феноменалистической онтологии. М.: УРСС, 2014.

<sup>42</sup> В. Васильев использует термин примерно в том же смысле, «что Витгенштейн понимал под миром». Васильев. В.В. Трудная проблема сознания. М.: Прогресс-Традиция. 2009. С. 241.

и создадим для него исчерпывающее конкретное описание. Это описание будет бесконечно длинным, пусть так. Но будет ли у этого конкретного события возможность в качестве antecedента иметь больше одной причины? Кажется, что да, по крайней мере, в соответствии с посылкой 1 исходного аргумента. Вне зависимости от того, сколь точным, конкретным будет описание одного сингулярного события, к нему может вести больше одной каузальной траектории. А теперь подумаем о Событии. Разве Событие, совокупное событие, текущее состояние мира, по определению не является совокупностью всех конкретных сингулярных событий? Если так, то его предшественниками могут быть варианты antecedентов всех конкретных сингулярных событий. И таких предшествующих состояний должно быть больше, чем одно. Я даже думаю, что чем больше сингулярных конкретных событий входит в совокупность, тем больше вариаций ей может предшествовать.

К примеру, пусть комбинация шаров на бильярдном столе такова, что игрок может забить шар в лузу только с отскоком от правого или левого борта. Если шар уже попал в лузу, и мы не видели, как он был забит, мы должны допустить, что он туда попал одним из двух вариантов. Теперь представим, что футбольное поле или даже пустыня Сахара заставлена бильярдными столами с аналогичным расположением шаров, и у каждого стола находится по одному игроку. Вот все шары оказались в лузах. Какое количество комбинаций могло предшествовать попаданию всех шаров в лузы? Очевидно, что это число больше, чем число комбинаций, предшествующих попаданию одного шара. Мне кажется, таким же образом дело должно обстоять и с совокупным событием. А это значит, что если Аргумент каузальных траекторий верен, он, вероятно, опровергает не только локальную супервентность, но и глобальную, т.е. все три варианта формулировки тезиса.

В таком случае ментальное становится полностью автономным от физического, парящим свободно над физическим миром. А это — нарушение ПКЗ, одного из принципов здравого смысла. Теория, которая ведет к таким следствиям, вряд ли может выстоять. Впрочем, не стоит торопиться. Даже если нам удалось опровергнуть второй путь защиты ТСС, надежным мог быть первый. А для анализа его у нас недостаточно инструментов. Думаю, в таком случае оправданно вернуться к исходному аргументу и продумать его посылки. На мой взгляд, достаточно будет ограничиться первыми двумя. Ведь для получения интересующего нас вывода (о ложности ТС) достаточно только их. Мы сократим аргумент и приведем его в несколько адаптированный вид.

*В. Разъяснение двусмысленности в аргументе*

(1) Реально возможно, чтобы мой мозг пришел к состоянию С в результате различных событий (А или В и т.д.).

(2) Реально возможно, чтобы память отражала события, происходившие с субъектом в прошлом.

(В) Реально возможно, чтобы мой мозг находился в некотором состоянии С, но моя память при этом могла бы отражать одно из различных предшествующих событий (А или В и т.д.).

В новой формулировке аргумента сделаны некоторые правки. Первая посылка конкретизирована в отношении возможных каузальных состояний мозга, ведь именно об этом в ней речь. Введено уточнение: каждая из посылок представлена не как факт, а как реально возможный факт. Это кажется уместным дополнением, так как только в этом случае посылки становятся верными. Мозг может прийти к некоторому состоянию С разными путями, хотя и не обязательно к нему придет. А память реально может отражать прошлые события, хотя и неверно, что она отражает все прошлые события. Вывод сокращенного аргумента опровергает ТС (1). И насколько я

понимаю, с такой формулировкой должен быть согласен автор Аргумента каузальных траекторий.

Однако в таком виде в аргументе проявляется некоторая двусмысленность. Реально возможно, чтобы мозг пришел к состоянию С в результате разных событий, реально возможно, чтобы память отражала произошедшие с субъектом события. Но в этих двух посылках — можно возразить — речь идёт совсем о разных событиях! В одном случае — о событиях, происходящих в нейронных сетях, а в другом — о биографических событиях. Память не отражает каузальных цепей нейронных событий, даже память нейрофизиолога-мнемониста, это реально невозможно. Тогда, значит, связь между посылками отсутствует? И из посылок вывода о ложности ТС вовсе не следует?

Конечно, дело обстоит не так. Двойственность из аргумента устранить возможно. Но тогда, мне представляется, рассуждение должно иметь более сложный характер. Назовем это рассуждение АКТ 1:

(1) Реально возможно, чтобы мой мозг пришёл к состоянию С в результате различных нейронных событий (А, В и т.д.).

(2) Другие нейронные события могли соответствовать другим биографическим событиям.

(3) Реально возможно, чтобы память отражала биографические события.

(В) Реально возможно, чтобы мой мозг был в одном и том же состоянии С, но моя память отражала различные биографические события (А, В и т.д.).

Посылки представляются истинными. И, следовательно, истинным может быть вывод. Правда, есть существенное условие. Для истинности вывода нужно не только, чтобы реально возможными были каждая из описанных в

посылке ситуаций, но и их конъюнкция: все события должны быть возможными одновременно. Ведь реально возможно начать игру в шахматы и с хода пешкой, и с хода конём, но нельзя начать игру сразу с двух ходов, конём и пешкой. Мне кажется, что совпадение ситуаций, описанных в посылках аргумента, все-таки невозможно. Невозможна ситуация, когда есть два двойника, и их мозги находятся в идентичных состояниях, но пришли к этим состояниям в результате разных нейронных событий, а также разных автобиографических событий, и эти двойники имеют разные воспоминания.

### 3. Возражения на Аргумент каузальных траекторий

#### *А. Аргумент-двойник*

Доводом в пользу невозможности этой ситуации служит аналогичный аргумент, Аргумент-двойник, который, хотя и имеет, по моему мнению, ключевые черты АКТ 1, приводит к очевидно ложному выводу. Если аргумент двойник не работает, но принципиально не отличается от АКТ 1, то и АКТ 1 не работает. Я предлагаю рассмотреть рассуждение, которое я назову АКТ 2:

(1) Реально возможно, чтобы жесткий диск моего компьютера пришёл в текущее состояние различными путями.

(2) Файловая система хранит информацию (компьютерная память) о последовательности изменений файлов на диске.

(В). Реально возможно, чтобы жесткий диск моего компьютера был в текущем состоянии, но файловая система хранила другую информацию о последовательности записи файлов на диск

Это рассуждение похоже на АКТ 1, и доводы, приводимые в пользу истинности посылок АКТ 1, полностью применимы и в новом аргументе. Диск компьютера — такая же физическая система, как мозг человека. Следовательно, текущему состоянию диска могли предшествовать различные состояния. Файловая система в качестве метаданных хранит информацию о

времени создания, доступа и модификации каждого файла. Эта информация обычно представляется правильным образом и отражает индивидуальную каузальную траекторию жесткого диска. Отклонения, конечно, возможны, но мы говорим о норме. И это — общее представление, разделяемое как любителями, так и профессиональными программистами. И тем не менее, вряд ли кто-либо станет утверждать отсутствие локальной супервентности данных файловой системы на физическом состоянии жесткого диска и допустит возможность ситуации, описанной в выводе АКТ 2. Но почему? Почему вывод АКТ 1 представляется правдоподобным в случае с памятью человека, а аналогичный вывод в случае с памятью компьютера в АКТ 2 — неправдоподобным? Может, нам стоит поэкспериментировать с жесткими дисками и экспериментально проверить работу принципа супервентности в компьютерных системах?

И все-таки в случае с механическими системами в локальной супервентности информации на носителе вряд ли можно сомневаться. Здесь с нами согласится и Васильев. Об этом можно судить из его анализа чего-то подобного, робота-зомби с механической памятью. «Такой робот мог прийти к своему нынешнему состоянию... разными путями. Но, независимо от того, как он пришёл к нему, он, будучи чисто механической конструкцией, должен в этом состоянии одинаково реагировать на своё окружение. Итак, для поведения нашего робота-зомби значима не его индивидуальная история, а исключительно его наличное состояние»<sup>43</sup>. Наличное состояние робота может отражать прошлые события. Но, по мнению российского философа, механическая память робота не чувствительна к его фактической индивидуальной истории. На чём основана такая уверенность?

Возьмём убеждение, что я родился в 1976г., и запись компьютера о том, что первый файл на его диск записан в 1976г. В чём принципиальная разница? Оба «воспоминания» содержат хронологическую метку, оба могут

---

<sup>43</sup> Васильев В.В. Трудная проблема сознания. М.: Прогресс-Традиция, 2009. С. 221.

быть верными или ошибочными, оба в качестве основания имеют некоторый правдоподобный источник (часы реального времени и свидетельство о рождении), и оба влияют на поведение. Оба «воспоминания» могут как-то отражать другие события, и оба могут исчезнуть или измениться, если оказать непосредственное воздействие на физический носитель, мозг или жёсткий диск. Повреждения или значимые изменения в физическом носителе в обоих случаях приводят к изменениям или исчезновению воспоминаний.

Может, различие в том, что механизм компьютерной памяти относительно хорошо изучен, а механизм работы человеческой памяти всё ещё отчасти остаётся загадочным для учёных? Но отсутствие понимания всех механизмов работы человеческой памяти не является доводом в пользу наличия у нее нефизических свойств и отсутствия локальной супервентности. Или в том, что память человека может содержать конкретные визуальные образы прошлого, а память компьютера — только последовательность битов, которые могут быть проинтерпретированы самым различным образом? Например, запись о времени создания файла какой-то другой программой действительно может быть проинтерпретирована иначе: как математическая формула, данные об изображении, команда распечатать файл и т.п. С этим возражением можно работать (можно попытаться показать, что воспоминания человека могут также быть интерпретированы по-разному, как и любая сжатая информация; тактики атаки и защиты подобных утверждений присутствуют, например, в дискуссии Сёрла и Деннета о «Китайской комнате»). Но более простой способ атаковать АКТ 1 — заменить цифровой жесткий диск в Аргументе-двойнике на аналоговый носитель, например, аналоговую кассету или диафильм. Эти носители сохраняют информацию в максимально конкретном виде, и альтернативная интерпретация данных оказывается затруднена. Мне кажется, что параллельный аргумент и его несостоятельность должны показать то, что сам по себе АКТ 1 недостаточен



для опровержения локальной супервентности. Для опровержения этого стороннику АКТ 1 нужно показать, что аргументы неэквивалентны.

И в рассуждениях Васильева есть указание на ещё одно, возможно, ключевое отличие механической памяти от памяти человека. Видимо, именно оно должно делать АКТ 1 правдоподобным, а его механическую вариацию, АКТ 2, — нет. В представлении автора память человека — это приватные квалитативные состояния. И поэтому «индивидуальное прошлое не растворяется в нейтральных нейронных кодах, а продолжает существовать на уровне ментальных образов»<sup>44</sup>. Это утверждение нужно использовать как ещё одну посылку в аргументе. Она фигурирует там проходящим образом, как нечто само собой разумеющееся, но крайне важна и потому требует разъяснения и оценки.

Квалиа — термин, обозначающий качественный или феноменальный аспект ментальных процессов: «мучительность» боли, «красноту» красного, «горечь» обиды, «сладость» наслаждения. Его используют для того, чтобы провести различие между содержательным аспектом ментального процесса (контентом) и субъективным переживанием этого процесса, «каково это: испытывать» процесс на себе. Когда автор говорит о том, что человеческая память состоит из приватных квалитативных состояний, он имеет в виду наличие этого качественного аспекта. Вот я вспоминаю прошлое лето, выходные: дача, на траве стоят плетеный столик и кресло, книги и чашка чая на столе, почти ощущаю запах стриженной травы, слегка греет солнце, чувствую удовольствие с оттенком сожаления. Воспоминания как бы оживляют пережитые ощущения, мои личные, те, что никто другой, кроме меня, не переживал и о которых другие могут узнать только косвенным образом. Этот пример должен объяснить утверждение Васильева. Но если хорошо подумать, все ли содержимое памяти представлено таким же образом? Вот некто знает иностранные языки, т.е. помнит тысячи слов,

---

<sup>44</sup> Васильев В.В. Трудная проблема сознания. М.: Прогресс-Традиция, 2009. С. 221.

выражений и грамматические правил. Разве такое содержание памяти окрашено феноменально? Каково это: «помнить три иностранных языка»? И отличается ли это состояние от состояния «помнить четыре или пять иностранных языков»? Сейчас, когда я печатаю этот текст, я, очевидно, помню собственную биографию, но каково это: «помнить свою биографию»? Отличается ли это состояние от состояния «помнить биографию Стива Джобса или Бенжамина Франклина»? А было бы мое состояние другим, если бы вместо того, чтобы сейчас помнить, что я родился в Москве, я, например, помнил, что родился в Бейруте или Бостоне? Память об этих моментах, кажется, влияет на моё поведение. Но чтобы иметь кажимость влияния, они не должны быть как-то качественно окрашены.

Теперь попробуем расставить всё по местам. У нас есть АКТ 1 и АКТ 2. Единственным различием между ними является допущение, что память человека состоит из качественных частных состояний. Но из наших примеров очевидно, что по крайней мере часть содержания памяти не имеет «качественного вкуса». Должны ли мы, вопреки интуиции, провести различие в каузальной действительности «качественно окрашенных» и «качественно пустых» воспоминаний? Или углубиться в интроспективные исследования, чтобы все-таки найти «вкус» всех без исключения записей памяти? Мне кажется, однако, что обе альтернативы бесперспективны. И вернее считать, что память не имеет «качественных качеств». Тем более что такую точку зрения могут занимать не только элиминативисты (противники существования квалиа), но даже такие квалиафилы, как Д. Чалмерс.

*В. Память не является качественным состоянием*

В своей основной монографии «Сознающий ум» (1996) Д. Чалмерс, один из самых известных современных аналитических философов, утверждает, что ментальное исчерпывается психологическим и феноменальным (качественным). Некоторые ментальные свойства сводятся только к психологическим, некоторые — только к феноменальным, а

некоторые имеют двойственный характер. Боль, к примеру, имеет как психологические, так и феноменальные аспекты. Психологической составляющей является состояние, связанное обычно с повреждением организма и ведущее к особому поведению, реакции уклонения. Как мы видим, это описание функциональное — оно фиксирует входящие параметры и соответствующее этим параметрам поведение-реакцию. Феноменальным аспектом боли является её мучительное переживание. Концептуально это несвязанные аспекты: теоретически организм мог бы быть поврежден и демонстрировать характерное для боли поведение, но при этом не испытывать болезненных переживаний, и наоборот, мог бы быть не поврежден и не демонстрировать боли в поведении, но интенсивно её переживать. Но фактически в мире и в жизни организма они переплетены. А вот память Чалмерс относит целиком к психологическим процессам: «Согласно нашему анализу в параграфе 3, обучение и память — функциональные свойства, характеризующиеся каузальными ролями...»<sup>45</sup>. И причина в том, что процесс запоминания и хранения в памяти можно легко представить бессознательным, лишенным феноменальных свойств. Она проходит своеобразный тест: если ментальное понятие М представимо без какого-либо связанного с ним феноменального качества, значит, оно психологическое. В таком случае почему автор аргумента каузальных траекторий считает, что память состоит из частных квалитативных состояний?

Очевидно, автор, когда говорит о памяти, говорит не о процессе запоминания и хранения информации о прошлом, а о ее воспроизведении в текущем опыте субъекта. В текущем опыте часть воспоминаний действительно представляется феноменально окрашенной. Но тогда точнее будет сказать, что не память обладает квалитативностью, а текущее состояние, объектом которого являются события из прошлого, из памяти.

---

<sup>45</sup> Chalmers D. The Conscious Mindconscious Mind. In Search of a Fundamental Theory. Oxford, New York: Oxford University Press, 1997 (1996). P. 24.

Содержание памяти может быть воспроизведено в сознании с феноменальными свойствами, но не имеет их в процессе хранения. Качественные аспекты проявляются в наличном состоянии субъекта, а не в его памяти.

А что если представить, будто эти ментальные содержания все-таки живут внутри нас в виде ментальных образов, постоянно текущих феноменальных состояний? Непонятно, как это представить, и непонятно, чьими бы эти состояния были. Если сейчас, когда я пишу эти строки, продолжает существовать феноменальное воспоминание о моем воскресном дне летом, то не совсем понятно, где и для кого оно продолжает существовать. Ведь онтология феноменальных состояний субъективна: если феноменальные состояния существуют, они существуют представленными для субъекта. Уверен, что в нейронных сетях закодированы воспоминания, но существование неосознаваемых никем феноменальных состояний мне представляется маловероятным. Тем более это трудно представить в связи с прерывностью сознательного опыта. Даже если прошлые феноменальные состояния продолжают существовать где-то на периферии моего сознания, то они должны улечиваться во время и после наркоза или даже сна. Но было бы странно предполагать, что только ментальные события после сна имеют каузальную действенность, а все, что происходило до сна, «растворилось в нейронных кодах». Чтобы защитить аргумент, недостаточно показать возможность «проигрывать» в настоящем воспоминания в «театре сознания». Нужны доводы в пользу непрерывного параллельного существования феноменальных состояний прошлого и непротиворечивый способ помыслить это. А этих доводов, судя по всему, мы пока не находим, и неосознанные сознательные состояния помыслить невозможно. Таким образом, различие в квалиа представляется мне нерелевантным. И мы возвращаемся к исходной ситуации: если бы работал АКТ 1, работал бы и АКТ 2.

*С. Решение парадокса сближающихся в ментальных состояниях двойников*

Впрочем, может быть, и тезис о супервентности ментального на физическом способен привести к трудностям подобного характера? Кажется, такое сомнение озвучивает Васильев, когда предлагает нам представить двойников со сближающимися физическими состояниями. Вот в разных мирах существуют два человека с разными физическими состояниями, но их состояния постепенно сближаются. По мере этого сближения должны постепенно стираться и их ментальные различия. Если это напоминает амнезию, стирание тех аспектов прошлого, которые отличали их друг от друга, двойникам «придётся забыть очень многое или даже почти всё, если двойников очень много, а ведь это тоже реально возможно»<sup>46</sup>. Чем больше двойников с альтернативными траекториями, тем больше воспоминаний должны быть стёрты. Чем больше двойников, тем меньше каждый из них может помнить. И человек, у которого бесконечное множество двойников, не должен помнить ничего. А значит, «сознанием в полной мере обладают только те люди, у которых нет двойников. Чем больше у вас двойников, тем больше вы и они должны быть похожи на зомби, хотя эти двойники никак с вами не связаны»<sup>47</sup>. Такое следствие действительно похоже на абсурд. Оно явно не согласуется со здравым смыслом и естественными установками. Неужели мы вынуждены принимать его, если отстаиваем тезис о супервентности и допускаем множественность каузальных траекторий?

Для того чтобы ответить на этот вопрос, нужно представить ситуацию с двойниками с бóльшими подробностями. Итак, есть два двойника, два почти неотличимых по физическому строению человека, живущих в разных мирах. Пусть одни из них, Дэн, родился в Бостоне, а другой, Фил, в Бейруте. И у них разные занятия: Дэн бизнесмен, а Фил — музыкант. Один проводит всё время в офисе, другой — играет джаз на фортепиано в каком-то маленьком ресторанчике. В какой-то момент параллельных, но не

---

<sup>46</sup> Васильев. В.В. Трудная проблема сознания. М.: Прогресс-Традиция, 2009. С. 219.

<sup>47</sup> Там же, С. 219.

пересекающихся жизней мозги двойников приходят в одинаковые состояния, одинаковые вплоть до каждого нейрона, атома, элементарной частицы. И в этот момент их спрашивают в каждом из миров: «Эй, ты кто? и чем ты здесь занимаешься?». Согласно ТС их ответ должен быть идентичным. Предположим, они оба отвечают: «Чего тебе? Я музыкант, я играю на фортепиано». Но как такое помыслить? Как бизнесмен мог дать такой ответ? Неужели он забыл все те часы, которые провёл в офисе? Он действительно должен был забыть многое, очень многое. Но очень многое он должен был и «помнить». К примеру, он должен помнить мелодию и гармонии джазовых стандартов «Каравана» и «Мисти», неподражаемые виртуозные пассажи легендарного Арта Тэйтума, должен тонко чувствовать свинговый ритм, синкопы, уметь импровизировать. Разве похоже все это на поведение и состояния зомби? Разве его нельзя назвать полноценным сознательным существом? Допустим, кроме Дэна и Фила были и другие двойники, и все они пришли к тому же состоянию, и все теперь говорят, что джаз — их призвание. Разве от этого сознание Дэна или Фила станет менее богатым? Вряд ли нам нужно так думать. Парадокс работает только в том случае, если допустить, что они только забывают, но не приобретают новые воспоминания. Только тогда это вело бы к сокращению возможного содержания памяти.

Действительно ли утверждение о том, что двойники по мере схождения их физических состояний должны забывать отличающие аспекты прошлого, ведёт в итоге к сокращению возможного содержания памяти? Мне кажется, это могло бы быть истинным только в случае конечного числа возможных воспоминаний. В самом деле, если от конечного числа возможных воспоминаний отнимать все большее и большее их количество, будет оставаться меньше возможностей. Но если количество бесконечно, то от вычитания все большего и большего числа различий мощность исходного множества не уменьшится. Хорошо, тогда, может быть, бесконечное число

двойников будет представлять опасность для содержания памяти? Мне кажется, что и здесь не возникает проблемы. Ведь если из бесконечного множества натуральных чисел удалить бесконечное множество нечётных чисел, останется бесконечное множество чётных. И это множество по мощности будет таким же, как исходное множество.

Но, по моему мнению, абсурдной ситуации (уподобления двойников зомби) не возникнет, даже если количество возможных воспоминаний конечно. Дело в том, что двойники не только «забывают» отличающие их аспекты прошлого, но и приобретают новые воспоминания. Так, двойник-бизнесмен в нашей ситуации не только забыл часы, проведенные им в офисе, но, судя по всему, и приобрёл множество дополнительных воспоминаний, связанных с игрой на фортепиано. Содержание его памяти не уменьшилось, а только изменилось. И если физически состояние его мозга сближалось с состоянием музыканта постепенно, стоит предположить, что постепенно менялось содержание его памяти. Постепенно он забывал часы, проведенные в офисе, и убеждался, что вместо офиса проводил время в ресторане и что смотрел он не в экран компьютера, а в ноты. Хотя все это представить очень сложно, имея представление о сложности устройства мозга, не менее сложно представить вероятность того, чтобы два мозга оказались в абсолютно тождественном состоянии.

Из рассуждений, приведенных выше, следует, что возможно непротиворечиво помыслить двойников со сближающимися ментальными траекториями. А это значит, что Тезис о супервентности не ведет к противоречивым последствиям. Но локальный инетракционизм не даёт желаемой перспективы, и тогда разве не возвращаемся мы к исходной дилемме? После критики локального интеракционизма разве не вынуждены мы отказываться от ментальной каузальности вовсе? Или придётся соглашаться с тождеством ментального и физического? Обе жертвы крайне нежелательны, а это значит, следует искать другое решение.

#### IV. Решение проблемы ментальной каузальности в биологическом натурализме Дж. Сёрла

##### 1. Ключевые положения биологического натурализма Дж. Сёрла.

Альтернативный выход представляет позиция американского философа Дж. Сёрла. Его теория, биологический натурализм, также предлагает вариант решения проблемы ментальной каузальности и сформулирована им в ряде монографий<sup>48</sup>. Биологический натурализм похож на локальный интеракционизм тем, что в нем также отрицается тождественность ментального и физического. Сёрл тоже настаивает на нередуцируемой природе ментального<sup>49</sup>. И также утверждает каузальную эффективность ментальных состояний. Но проблему ментальной каузальности американский философ пытается решить по-другому, не ставя под сомнение супервентность ментального на физическом. В этом — ключевое отличие его позиции от локального интеракционизма. Позицию Дж. Сёрла в отношении сознания и ментальной каузальности можно резюмировать в следующих тезисах:

*Онтологический статус сознания:* Ментальные состояния существуют как высокоуровневые характеристики, но нередуцируемы к низкоуровневым.

*Ментальная каузальность:* Ментальные состояния каузально влияют на наше поведение.

*Связь с состояниями мозга:* Ментальные состояния полностью каузально обусловлены процессами в мозге.

---

<sup>48</sup> Сёрл Дж. Открывая сознание заново. Пер. с англ. Грязов А. М.: Идея-Пресс, 2002; Searle J. Intentionality: An Essay in the Philosophy of Mind. Cambridge: Cambridge University Press, (1983) 1999. Searle J. Mind: A Brief Introduction. Oxford: Oxford University Press, 2004.

<sup>49</sup> В одной из своих монографий Дж. Сёрл пишет: «Я настаиваю на том, ...что мы можем принять очевидные факты физики... без того, чтобы в то же самое время отрицать очевидные факты нашего внутреннего опыта — например, что все мы обладаем сознанием и что наши сознательные состояния имеют достаточно специфические нередуцируемые феноменологические свойства». Searle J. The Rediscovery of Mind. Cambridge, 1994 (1992). P. 28.



*Каузальная замкнутость*: Принцип каузальной замкнутости работает, но принцип каузальной исключительности не работает.

Первые два тезиса, как мы видели, являются общими для локального интеракционизма и биологического натурализма. Нередуцируемость ментального для Сёрла обусловлена исключительными характеристиками ментальных феноменов. Всего Сёрл перечисляет четыре особенных свойства ментальных состояний: (а) субъективность, (б) квалитативный характер, (в) интенциональность и (г) единство опыта<sup>50</sup>. Под субъективностью ментального имеется в виду именно онтологическая, не эпистемическая субъективность. Сёрл пишет: «ментальные состояния обладают нередуцируемой субъективной онтологией»<sup>51</sup>. Для Сёрла ментальные феномены существуют только как часть внутреннего опыта человека (или животного).

Когда Сёрл говорит о квалитативном характере, он имеет в виду качественный характер внутренних состояний. По его мнению, ментальное необходимо имеет этот уникальный для каждого субъекта характер, «то, что значит быть» в конкретный момент этим самым субъектом. К примеру, ощущение вкуса пива, запаха свежее испеченного хлеба, звуков третьей симфонии Бетховена — все это внутренние, непосредственно данные субъекту и недоступные для других переживания. Под интенциональностью Сёрл подразумевает направленность состояний на объекты: радость чему-то, сожаление о чем-то, ожидание чего-то, надежда на что-то. Почти все ментальные состояния имеют такое свойство. И, наконец, единство опыта — это объединение в опыте зрительных образов, тактильных ощущений, звуков,

---

<sup>50</sup> Searle J. *The Rediscovery of Mind*. Cambridge, 1994 (1992). P. 117.

<sup>51</sup> Searle J. *The Rediscovery of Mind*. Cambridge, 1994 (1992). P. 284.

запахов и проч. Сёрл подчёркивает здесь холистическое свойство сознательного опыта<sup>52</sup>.

Итак, каузальная эффективность ментального тоже утверждается обоими философами. Только российский философ считает это утверждение выводом из очевидных посылок, а Сёрл — очевидностью. По мнению Сёрла, ментальная каузальность — это эмпирический факт, с которого и нужно начинать. «Я думаю, что мы ощущаем необходимую [каузальную — *Д.В.*] связь почти всю нашу сознательную жизнь...»<sup>53</sup>. Вот, — говорит он на очередной лекции, — я хочу поднять руку, и проклятая штука поднимается, поднимается из раза в раз. Только чудак способен сомневаться в том, что причиной движения руки является желание.

Сёрл удивительным образом сочетает тезисы о ментальной каузальности и каузальной замкнутости физического мира. В третьем из выделенных выше тезисов он утверждает, что ментальные состояния с их специфическими характеристиками вызваны процессами в мозге и только ими. Но это еще ничего не говорит о том, чем именно являются ментальные состояния. Может даже показаться, что Сёрл считает, будто процессы в мозге являются причиной каких-то других процессов, происходящих не в мозге, а где-то ещё, в духовной субстанции, например. Но это не так. Его позиция по этому вопросу зафиксирована в первом тезисе. Ментальные состояния реализованы в самом мозге как высокоуровневые характеристики системы.

В качестве аналогии высокоуровневой характеристики Сёрл приводит такое свойство, как текучесть. Одна молекула  $H_2O$  не обладает свойством текучести, но множество молекул — обладают. При этом текучесть не является каким-то мистическим или необъяснимым феноменом. Она

---

<sup>52</sup> Резюмирующее описание теории Сёрла можно найти в его статье «Биологический натурализм»: Searle J. *Biological Naturalism* // Max Velmans and Susan Schneider (eds.). *The Blackwell Companion to Consciousness*. Oxford: Blackwell Publishing Ltd, 2007. P. 25.

<sup>53</sup> Searle J. *Mind: A Brief Introduction*. Oxford: Oxford University Press, 2004. P. 202.

складывается из свойств скоплений молекул. Причиной текучести, согласно Сёрлу, являются свойства скопления молекул, при этом текучесть «реализована» в скоплениях молекул. Аналогичным образом, по его мнению, обстоит дело с сознанием. Причиной ментальных состояний являются процессы в мозге, при этом сознание реализовано в мозге.

Минимальным расхождением между локальным интеракционизмом и биологическим натурализмом представляются основания второго тезиса. Васильев считает, что ментальная каузальность требует доказательства, а Сёрл, кажется, видит возможность обосновать её простой демонстрацией. Является ли ментальная каузальность эмпирическим фактом или гипотезой, требующей доказательств?

Думаю, позиция сторонника локального интеракционизма здесь более верна. Действие желания как причины мы не можем наблюдать. Ну и что, что после того, как мы ощущаем желание поднять руку, «чертова штука» поднимается из раза в раз? Ментальная каузальность — не факт, не очевидность, а гипотеза, которую нужно доказывать. Фактом является не наличие причинно-следственной связи между желанием и действием, а временное следование. Вот я хочу поднять руку, и вот она поднимается. Если бы фактом было наличие причинно-следственной связи между желанием и действием, фактом была бы и причинно-следственная связь между днём и ночью. Вот день, вот ночь. Они всегда следуют друг за другом. Но разве день является причиной ночи? Причиной наступления дня и ночи является нечто иное — движение планеты вокруг своей оси. Даже если последовательность есть, причина может находиться в другом месте. Так может обстоять дело и с ментальной каузальностью. Последовательность является эмпирической данностью, но необходимую связь, каузальную эффективность нужно доказывать.

Теперь разберёмся в пунктах, в которых сосредоточены основные различия. Сёрл утверждает, что ментальные состояния полностью

обусловлены процессами в мозге. Т.е. что ментальные состояния детерминированы состояниями мозга и только ими. Но какого рода эта связь? Сёрл считает, что ментальные состояния детерминированы мозгом, потому, что они в *качестве причины* имеют физические процессы в мозге и реализованы в мозге в качестве высокоуровневой характеристики. «Если обобщить [мою позицию — Д.В.] в этот момент, я утверждаю, что два феномена могут быть как связаны каузальной связью, так и реализованы один другим, если они находятся на разных уровнях описания»<sup>54</sup>. Это высказывание оригинально и даже удивительно потому, что сочетает в себе две, казалось бы, противоположные идеи. Сёрл как бы смешивает здесь понятие причины и субстрата. Несколько необычно утверждение, что нечто может быть реализовано в чем-то, и это что-то одновременно может быть его причиной. Но не только соотношение последних двух тезисов должно вызывать удивление. Как согласуются утверждения, что сознание есть высокоуровневая характеристика мозга и что при этом оно не редуцируется к состояниям мозга? Ведь высокоуровневые характеристики должны редуцироваться к характеристикам низкого уровня, а Сёрл утверждает нередуцируемость ментального.

Концепция Сёрла нарушает различие, которое мы провели в начале исследования: между расшифровывающей функцией подчинительной союза «потому» и каузальной. Кажется, он демонстрирует, что союз «потому» может указывать сразу на оба отношения: отношение концептуальной и каузальной связи. Примеры самого Сёрла лучше прояснят его позицию.

## 2. Иллюстрации к каузальным связям в биологическом натурализме

Рассмотрим переход воды в газообразное состояние в процессе кипения и проанализируем его с позиции Сёрла<sup>55</sup>. Если отбросить связь с агентом (пока оставим ментальную каузальность в стороне и рассмотрим

<sup>54</sup> Searle J. Intentionality. An Essay in the Philosophy of Mind. Cambridge: Cambridge University Press, (1983) 1999. P. 266.

<sup>55</sup> Этот пример Сёрл рассматривает в книге «Открывая сознание заново». Searle J. The Rediscovery of Mind. Cambridge, 1994 (1992). P. 87.

только очевидно физические процессы), кипение можно объяснять так: вода закипает потому, что разогревается зажженной конфоркой до соответствующей температуры. Т.е. (1) разогрев воды на огне приводит к переходу её в газообразное состояние. Кажется, это вполне законное каузальное объяснение. Но без труда найдется и другое описание, которое претендует на роль каузального. Кипение происходит потому, что кинетическая энергия, передающаяся в результате окисления углеводов (горения топлива) молекулам  $H_2O$ , заставляет их двигаться так быстро, что между ними разрываются (полярные и водородные) связи. Т.е. (2) повышение кинетической энергии до определенного уровня приводит к разрыву связей между молекулами воды. Таким образом, у нас два совершенно законных каузальных объяснения перехода в газообразное состояние. Первое — на макроуровне (макрособытие  $A \rightarrow$  макрособытие  $B$ ), а второе — на микроуровне (микрособытие  $A \rightarrow$  микрособытие  $B$ ). Оба описания позволяют проследить связь между антецедентом и консеквентом. Эти описания отражают последовательность событий во времени, поэтому вслед за Сёрлом назовём их описаниями «слева направо»<sup>56</sup>. Есть ли между этими описаниями конкуренция? По мнению философа, в данном случае проблему конкуренции, даже если она возникает, решить не сложно. Достаточно провести соотношение между событиями микро- и макроуровней. Т.е. идентифицировать разогрев воды с повышением уровня кинетической энергии молекул воды, а газообразное состояние воды — с состоянием, когда молекулы воды не объединены полярными и водородными связями между собой. Как только идентификация осуществлена, проблема исчезает: у нас не две различные причины, не два конкурирующих за каузальную эффективность явления, а одно, но с двумя описаниями на двух разных уровнях.

Наверное, тогда можно добавить и другие каузальные описания, тоже без конкуренции: микро-макро и макро-микро слева-направо: (3) повышение

---

<sup>56</sup> Searle J. The Rediscovery of Mind. Cambridge, 1994 (1992). P. 87.

кинетической энергии молекул воды до определенного уровня является причиной кипения (4) разогрев воды до определённого уровня приводит к разрушению полярных и водородных связей. Таким образом, есть уже четыре способа каузального описания слева-направо: микро-микро, микро-макро, макро-макро и макро-микро<sup>57</sup>. Эти способы возникают в силу того, что одно и то же явление можно описывать на разном уровне. И, по всей видимости, конкуренция за каузальную эффективность между этими описаниями снимается через идентификацию явлений нижнего и верхнего уровней.

Но с позиции Сёрла в описанном явлении существует ещё один тип каузальных связей — «снизу-вверх». Этот тип связи наиболее экстравагантен. Судя по его рассуждениям<sup>58</sup>, Сёрл утверждает наличие каузальности между одномоментными, синхронными событиями на разных уровнях. Т.е. повышение кинетической энергии молекул воды он также считает причиной повышения температуры воды. И, соответственно, (5) разрыв связей между молекулами воды является причиной перехода воды в газообразное состояние. Таким образом, по мнению Сёрла, правомерно еще и пятое описание причины кипения воды: снизу-вверх, микро-макро. Конкуренция в этом случае, по его мнению, также не является проблемой. Ведь речь идёт об одном и том же явлении.

---

<sup>57</sup> Ibid., P. 87.

<sup>58</sup> Эти рассуждения о ментальной каузальности Дж. Сёрла приведены в Эпilogue к монографии «Интенциональность. Эссе по философии сознания». Вместо нашего примера с кипением воды Сёрл в этом тексте анализирует работу двигателя внутреннего сгорания, тоже на двух уровнях. Searle J. Intentionality. An Essay in the Philosophy of Mind. Cambridge: Cambridge University Press, (1983) 1999. P. 264-272.

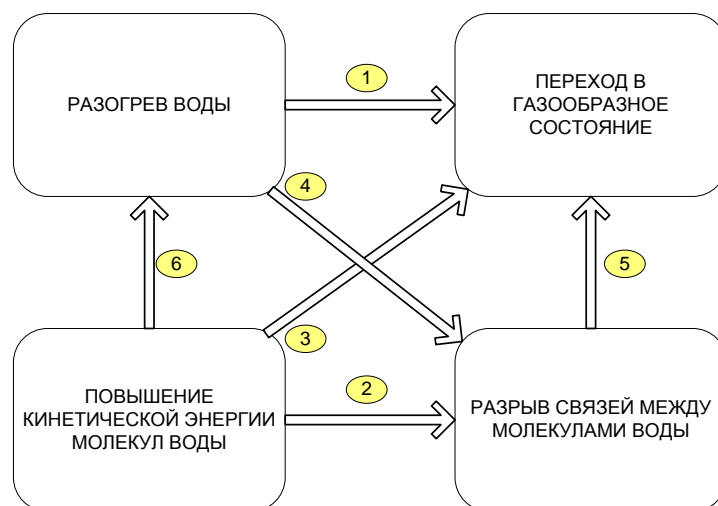


Рис. 1. Схема каузальных связей процесса кипения воды

Этот пример является важным инструментом для иллюстрации подхода к феномену сознания и решения проблемы ментальной каузальности Сёрлом. Он, наверное, должен служить доводом в пользу существования множественности достаточных причин для простых физических процессов. Теперь остается провести следующую аналогию между соотношением высокоуровневых и низкоуровневых характеристик физических процессов — с одной стороны, и соотношением ментальных и нейронных процессов — с другой. Сёрл её и предлагает.

Философ утверждает, что отношение между мозгом и сознанием, физическими и ментальными состояниями, является не чем-то удивительным, а вполне тривиальным. И что мозг является причиной ментальных состояний, также как причиной этих ментальных состояний являются предшествующие ментальные состояния. Мозг каузально влияет на сознание, и сознание — на последующие состояния сознания. Таким образом, Сёрл вписывает ментальное в каузальные цепочки природы и представляет конфликт ментальной и физической причин иллюзией.

Рука поднимается вверх. Причиной этого действия, по его мнению, являются как желание поднять руку агентом, так и возбуждение такой-то сети нейронов в его мозге. Но это — не две конкурирующие причины.

Желание поднять руку и возбуждение сети нейронов — это описания событий на разных уровнях: желание — на макроуровне, а возбуждение нейронов — на микроуровне, подобно тому, как разогрев воды является описанием процесса на макроуровне, а повышение кинетической энергии молекул воды — на микроуровне.

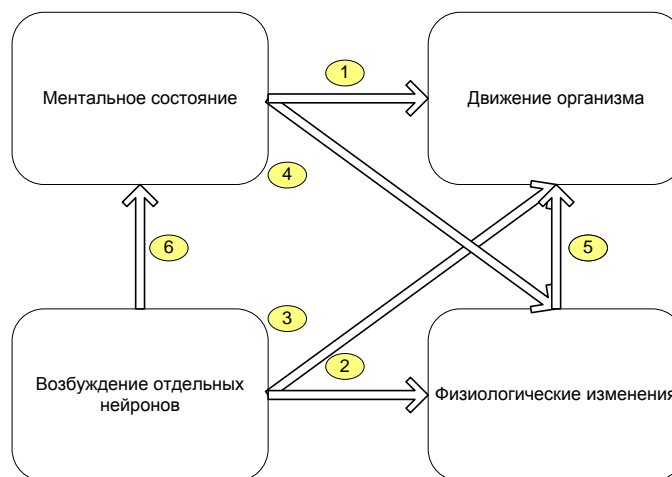


Рис. 2. Схема каузальных связей движения организма<sup>59</sup>

Рассуждения Сёрла вполне понятны. Полезным кажется его указание на возможность параллельно описывать явления на макро- и микроуровнях. Но (1) действительно ли когерентно такое описание причин физических явлений? И (2) действительно ли проходит аналогия между высокоуровневыми и низкоуровневыми характеристиками физических процессов, с одной стороны, и нейрофизиологическими процессами и состояниями сознания — с другой, в рамках теории сознания Сёрла?

### 3. Внутренние противоречия теории Дж. Сёрла

Само описание причинных связей в примере Сёрла вызывает у меня сомнение, в особенности 5-ое направление каузального действия — синхронного. Причиной газообразного состояния воды, по мнению Сёрла, является свободное движение молекул. Но ведь газообразное состояние и есть состояние, в котором молекулы движутся свободно. Причиной жидкого

<sup>59</sup> Подобный рисунок Сёрл приводит в том же эпилоге. Searle J. Intentionality. An Essay in the Philosophy of Mind. Cambridge: Cambridge University Press, (1983) 1999. P. 270.



состояния воды философ называет ограниченное движение молекул в условиях полярных и водородных связей. Но именно ограниченное движение молекул в условиях полярных и водородных связей и является жидким состоянием. Между двумя парами этих состояний имеет место редуцивно-эмерджентное отношение явлений разного уровня или логическая связь, но не каузальная. Т.е. газообразное состояние воды редуцируется к свободному и быстрому движению её молекул и является эмерджентным по отношению к свободному и быстрому движению молекул. Проблема для Сёрла возникает в связи с симметрией описываемых отношений. Философ представляет здесь каузальные отношения как симметричные, но в таком случае понятие причинно-следственной цепи становится, на мой взгляд, некогерентным<sup>60</sup>. Причинная связь асимметрична, причины определяют действия, а не наоборот.

Предположим, что в момент  $t_1$  происходят низкоуровневые события. И в этот же момент происходят высокоуровневые события. Изменения высокоуровневых событий не происходят без изменений низкоуровневых событий. Существует очевидная регулярность. Если допустить, что низкоуровневые события являются причинами, а высокоуровневые события — следствиями, то по той же самой логике высокоуровневые события являются причинами, а низкоуровневые события — следствиями. Следствие является причиной, а причина — следствием. Порочный круг можно разорвать, но только допустив асимметрию в отношении разных уровней свойств. Скажем, утверждая, что каузально эффективными всегда являются низкоуровневые свойства. В пользу этого хода можно привести доводы. К примеру, оправдано свойства целого выводить из свойств его частей. Физика считается самой фундаментальной наукой именно потому, что изучает

---

<sup>60</sup> К подобным выводам приходит в своей диссертации и А.В. Кузнецов. Он пишет: «Понятие эмерджентной причинности [Серла — Д.В.] противоречит определению причины и действия, которые должны быть разведены по времени. События, происходящие одновременно, не могут быть по отношению друг к другу причиной и действием согласно дефиниции». Кузнецов А.В. Проблема ментальной каузальности в аналитической философии сознания. Диссертация на соискание степени кандидата философских наук: 09.00.03 — История философии. М., 2016. С. 109.

элементарные частицы и законы их взаимодействия. Но именно для Сёрла этот ход закрыт. Ведь сам философ неоднократно утверждает возможность каузального действия не только «снизу вверх», но и «сверху вниз». См., например, Рис 1., связь 4. Понятие казуальной зависимости, мне кажется, предполагает асимметричность, а Сёрл её нарушает. Поэтому приведенный способ каузально описывать процесс кипения мне кажется некогерентным.

#### 4. Недостатки проведенных аналогий и иллюстраций

Думаю, что и аналогии Сёрла работают не совсем удачно — с ними возникает проблема именно в рамках биологического натурализма. Аналогия между ментальным и высокоуровневыми характеристиками физических процессов плохо согласуется с другими положениями биологического натурализма, да и не схватывает специфических свойств ментального по теории Сёрла (субъективность, квалитативность, интенциональность, единство). Высокоуровневые характеристики физических процессов (например, кипение) действительно реализованы в низкоуровневых микроструктурах. Но в данном случае они редуцируемы к характеристикам микроструктур, совпадают в пространстве и выводимы из движения микроструктур. Наблюдая, скажем, поведение молекул воды, можно было бы предположить свойства газообразного, жидкого и твердого состояний. И переход от микро- к макроуровням можно проследить эмпирически.

Проведем исследование растекающейся жидкости под микроскопом. Сначала на максимальном увеличении наблюдаем движение одной молекулы воды. Если мы поменяем кратность увеличения, то увидим несколько молекул, их скопление и общее движение. Если еще уменьшить кратность, то, наконец, станет видно движущуюся каплю воды, т.е. каплю, обладающую свойством текучести. Хотя это — увлекательное исследование, оно не шокирует. Изменяя степень увеличения, мы, тем не менее, не изменяли перспективу. Обнаруживая, как скопления движущихся молекул формируют каплю, мы не обнаруживаем в какой-то момент, что не

наблюдатель рассматривает каплю, а капля — радужку и зрачок глаза. Однако именно такой радикальный скачок получился бы, если бы, по логике Сёрла, изменяя параметры оптики наблюдения нейрофизиологических процессов, мы вдруг увидели бы онтологически субъективные ментальные состояния агента.

Мне кажется, что аналогия работала бы только в случае, если бы Сёрл отождествлял ментальные процессы с физическими. Ведь высокоуровневые характеристики разогрева и кипения воды можно отождествить с низкоуровневыми процессами увеличения кинетической энергии молекул воды. Но в теории сознания Сёрла нельзя ожидать такого отождествления. Ведь среди его ключевых тезисов — утверждение о нередуцируемости ментального и его уникальных специфических черт, в первую очередь — субъективности. Сознание, согласно его мнению, онтологически субъективно, т.е. существует как приватные субъективные состояния. Но это принципиально отличает сознание от физического. И, следовательно, аналогия в теории Сёрла не проходит. Из всего этого можно сделать вывод о том, что решение проблемы ментальной каузальности в биологическом натурализме не оптимально. Другой альтернативой является теория коллеги Сёрла и его наиболее знаменитого оппонента, философа Д. Деннета.

## V. Решение проблемы ментальной каузальности в телеофункционализме Д. Деннета

### 1. Ключевые положения телеофункционализма Деннета

Д. Деннет является одним из самых важных современных аналитических философов. Он — автор множества монографий и сотен статей. Главный вклад Деннета в философию сознания заключается в том, что он заставил коллег отвлечься от исключительно «кабинетных изысканий», показал, что исследования в области когнитивистики, нейрофизиологии и биологии

имеют отношение к обсуждению философских проблем<sup>61</sup>. Разработанный Деннетом подход можно назвать телеофункционалистским (ТФ). «Я — своего рода "телеофункционалист", наверное, один из первых телеофункционалистов»<sup>62</sup>, — пишет он в монографии «Объясненное сознание». Теория сознания Деннета оснащена множеством эмпирических знаний, но в ней можно выделить философскую основу. Метафизические положения телеофункционализма сводятся к следующему:

*Онтологический статус сознания:* Ментальные состояния существуют как высокоуровневые характеристики физических свойств.

*Ментальная каузальность:* Ментальные состояния каузально влияют на наше поведение.

*Связь с состояниями мозга:* Ментальные состояния полностью реализованы процессами в мозге.

*Каузальная замкнутость:* Принцип каузальной замкнутости работает в физическом мире, но некоторые события могут быть объяснены только высокоуровневыми закономерностями.

Деннет согласен с Сёрлом в том, что ментальные состояния представляют собой высокоуровневые характеристики физических свойств. Однако в отличие от Сёрла он не считает, что эти свойства — биологического характера. Деннет убежден, что любой механизм, обладающий функциональными качествами мозга, будет обладать ментальными состояниями<sup>63</sup>. Как и Сёрл, Деннет уверен, что ментальные высокоуровневые характеристики каузально эффективны. Однако в отличие от Сёрла он не считает, что мозг является причиной появления сознания. По его мнению, ментальные свойства реализованы в физическом субстрате как

---

<sup>61</sup> Этот призыв Деннет сделал в статье «От кабинетной философии — в поле». Dennett D. *Brainchildren: Essays on Designing Minds*. Cambridge, Massachusetts: The MIT Press, (1981), 1998. P. 289.

<sup>62</sup> Dennett D. *Consciousness Explained*. Boston, Massachusetts: Little Brown & Co, 1991. P. 459.

<sup>63</sup> Dennett D. *Consciousness Explained*. Boston, Massachusetts: Little Brown & Co, 1991. P. 280.

высокоуровневые — в низкоуровневых. Деннет также, судя по всему, разделяет принцип каузальной замкнутости физического мира, но допускает наличие высокоуровневых причин. Скорее всего, он не разделяет принцип исключения, хотя в открытом виде этого утверждения я у него не нашел.

Телеофункционализм Деннета является современным развитием функционалистской программы объяснения сознания. Функционализм — это позиция, согласно которой сущность ментальных состояний выражается в способе их функционирования, в той роли, которую они играют в других процессах. Определить наличие ментального содержания можно по результатам обработки системой входящий сигналов. Например, боль с функционалистской позиции — это состояние, ведущее к реакции избегания; симпатия — наоборот, к притяжению. Убеждение в том, что «сегодня будет дождь», — это что-то вроде предрасположенности взять зонтик или плащ после просмотра прогноза погоды в случае пребывания в состоянии «желания остаться сухим». Если обобщить, то по этой логике *убеждения* представляют собой состояния, определённые данными органов восприятия, а также другими убеждениями, и при взаимодействии с желаниями формирующие поведение или действия. А *желания* — это состояния, находящиеся в связи с другими желаниями и намерениями, при взаимодействии с убеждениями формирующие поведение или действия.

Функционализм предполагает возможность множественной реализации<sup>64</sup> ментальных состояний. То есть, по мнению функционалистов, субстрат и внутренняя механика системы не определяют ментальных содержаний. Биологический носитель — только один из вариантов. Вполне возможно ментальные содержания реализовать и на механических носителях, в частности на кремниевых чипах. Такой подход характеризует все виды

---

<sup>64</sup> Это понятие ввел в философию сознания Х. Патнэм. В серии статей, направленных против теории тождества, он показал, что ментальные состояния могут быть реализованы различными физическими состояниями, поэтому ментальное и физическое не может быть в строгом смысле тождественно друг другу. Art, Mind, and Religion. Capitan W.H. and Merrill (eds.). University of Pittsburgh Press. P. 37-48.

функционалистских теорий. Взгляды Деннета, тем не менее, имеют некоторую специфику.

### Своеобразие функционализма Деннета

Позицию Деннета можно для контраста сравнить с другим видом функционализма — функционализмом машины Тьюринга (ФМТ)<sup>65</sup>. Эта теория несколько более ранняя и, возможно, менее совершенная. Для иллюстрации различий используем два ключевых вопроса: (1) что представляют собой ментальные содержания (content)? (2) что общего имеется между двумя системами, когда они имеют одно общее убеждение? Посмотрим, как на них ответят сторонники ФМТ и ТФ.

Ответ на первый вопрос, скорее всего, будет идентичным для обоих вариантов функционализма. Он служит идентификатором монизма. «Каждое ментальное состояние — это какое-то физическое состояние», — считает большинство функционалистов. А вот по поводу второго вопроса позиция может быть различной. Сторонник ФМТ скажет, что системы, имеющие общее убеждение, могут быть описаны как выполняющие один и тот же алгоритм и находящиеся в одном и том же месте выполнения этого алгоритма. Например, как если бы синхронно запущенная презентация PowerPoint на компьютере IBM и на компьютере Macintosh была остановлена на одном и том же слайде. Сторонник ТФ ответит более аккуратно. Он будет утверждать, что единственная общность между двумя системами, разделяющими одно убеждение, будет заключаться в том, что поведение обеих систем можно будет надежно предсказать, допуская это убеждение. Т.е. (1) телеофункционалист не требует ни общности в физической реализации, ни общности «программы» (мыслящие существа могут в силу образования, воспитания и наследственных данных значительно отличаться «алгоритмами»). Могут быть запущены разные презентации на разных

---

<sup>65</sup> В «Объясненном сознании» Д. Деннет пишет: «Тогда являюсь ли я функционалистом? И да, и нет. Я не функционалист Машины Тьюринга». Dennett D. *Consciousness Explained*. Boston, Massachusetts: Little Brown & Co, 1991. P. 460.

машинах, только бы эти презентации вели к одним и тем же практическим выводам и действиям. В этом смысле телеофункционализм Деннета более либерален, чем ФМТ. Но здесь нужно сразу сделать оговорку. Предъявляя меньшие требования для идентификации ментальных содержаний в отношении алгоритмов, автор теории вводит дополнительное требование: надежное и осмысленное соединение системы с внешним миром.

Деннет не случайно называет свою позицию телеофункционалистской. Вероятнее всего, это указывает на (2) связь понятий функции и направленности в его теории ментальных содержаний. Для наличия соответствующих интенциональных состояний недостаточно иметь аналогичную реакцию на входящую информацию. Реакция должна соотноситься с функцией системы в целом, её интересами, приоритетами, целями. В случае с живыми организмами это — биологические, адаптационные функции. Реакция должна быть не просто соответствующей, она должна иметь соответствующий смысл.

Недостаточно, чтобы программа приводила к похожим действиям, она должна осмысленным образом состыковывать интенциональную систему с внешним миром. По мнению Деннета, семантическое содержание приобретает только в целенаправленном взаимодействии с внешним миром. Логика выживания глобально определяет интенциональные содержания человека. Робот тоже может являться системой для выживания, и соответственно, с позиции американского философа, иметь интенциональные содержания. Но у него тогда должны быть цели, интересы, приоритеты. Не все механизмы так состыкованы с внешним миром. Лишенный собственного интереса термостат только очень условно может соответствовать критериям ментальных содержаний. Для человека сохранение температуры тела является принципиальным условием выживания, а для термостата — просто рабочим назначением. К тому же чувствительность термостата не позволяет ему точно дифференцировать, настраивать ответные «действия». Термостат

можно использовать как для регулирования потока хладагента, поступающего в радиатор, так и для запуска ядерной ракеты. Биологические сенсоры и механизмы управления гораздо избирательней в репертуаре действий, они почти идеально настроены под соответствующие воздействия внешнего мира и обеспечивают адекватную реакцию на стимул.

Но, если у мыслящих агентов разные программы, как возможны надёжные предсказания? Почему одни мыслящие существа способны «понимать» состояния и предсказывать поведение других? Ключом к разрешению этого вопроса является некоторый стандарт, общая норма, определяющая поведение всех существ. Это стандарт рациональности. И существованию этого стандарта мы обязаны общему эволюционному процессу. Несмотря на то, что каждый вид и даже каждая особь могут иметь свою собственную программу рациональности, они все написаны по одному лекалу. У всех программ один общий программист — естественный отбор.

Одна и та же рациональность стоит за строением птичьего крыла и... проектом атомной подводной лодки, созданной человеком. Рациональность, которая повсюду встречается в природе и которая в течение столетий объяснялась креационистской моделью как свидетельство рациональности создателя, объясняется теорией естественного отбора. Из ее основного принципа следует, что если имели место борьба видов и естественный отбор, то популяции, которые живут сегодня, являются наиболее приспособленными к условиям существования, или обладают *оптимальной конструкцией*. Они выработали наиболее эффективные модели реакций на воздействия среды и способны к стабильному воспроизводству. Общей целью для всех живых существ является сохранение вида, а следовательно, сохранение жизни особи, когда она может произвести потомство. И поведение, которое ведет к достижению этой цели — наиболее оптимальное или рациональное. В контексте теории эволюции можно еще раз уточнить определения «убеждение» и «желание». Деннет пишет, что «убеждения» и



«желания» в конечном итоге демонстрируют необходимость выживания и отражают оптимальное устройство органической системы. На этом и основаны механизмы предсказания поведения агентов. «Если эволюция проделала свою работу, ...наши предсказания должны быть достаточно надежными, чтобы ими пользоваться»<sup>66</sup>.

Любопытно также отметить, что то, что отличает теорию контента Деннета от некоторых функционалистских теорий контента, сближает его позицию с концепцией интенциональности — такой, какой она появилась в философском обращении в конце 19в. Интенциональность (от латинского *intendo* — указывать на что-то, целить во что-то) — это отличительная черта психического, как было подмечено австрийским философом Ф. Brentano, указывает на способность ментальных актов, с одной стороны, содержать нечто (обладать контентом), а с другой — быть направленными на что-то. Убеждение, к примеру, в том, что Земля круглая, с одной стороны, *содержит в себе* некоторую информацию и отношение к ней, но также, с другой стороны, указывает на некоторый объект за пределами ментального. Как пишет Brentano в книге «Психология с эмпирической точки зрения» (1874): «Каждый психический феномен характеризуется посредством того, что схоласты средневековья называли интенциональным (а также, пожалуй, ментальным) *существованием в нем некоторого предмета*, и что мы, хотя и не вполне избегая двусмысленности выражений, называли бы отношением к содержанию, *направленностью на объект* (под которым не следует понимать некоторую реальность), или имманентной предметностью. Каждый [психический феномен] содержит в себе объект, хотя и не каждый одинаковым образом. В представлении нечто представляется, в суждении нечто признается или отвергается, в любви — любитесь, в ненависти — ненавидится, в желании — желается и т.д.»<sup>67</sup>. Деннет сохраняет эту

---

<sup>66</sup> Dennett D. Brainstorms: Philosophical Essays on Mind and Psychology. Cambridge, Massachusetts: The MIT Press, 1998. P. 9.

<sup>67</sup> Brentano Ф. Психология с эмпирической точки зрения. Т.2, раздел 7. Представления и суждения как два различных класса психических феноменов // Вопросы философии, 1995, №2.

двусмысленность термина: в его концепции интенциональность указывает и на содержание, и на направленность. Это — и функциональное репрезентативное состояние, и направленность в мир, на объект.

Деннет, как и Васильев, как и Сёрл, предлагает свое решение проблемы ментальной каузальности. Однако в данном случае решение Деннета не представляет собой строгий аргумент или эксплицитную теорию. Когда философа спрашивают о ментальной каузальности, он всегда ссылается на вымышленную им ситуацию, которая называется «История двух черных ящиков»<sup>68</sup>. Деннет считает, что описание этой ситуации само по себе проливает свет на проблему ментальной каузальности и демонстрирует ее решение.

## 2. «История двух черных ящиков» как иллюстрация каузальной эффективности высокоуровневых свойств

Представим, что некогда были найдены два соединенных между собой кабелем чёрных ящика. На первом расположены две кнопки:  $\alpha$  и  $\beta$ , а на втором — три световых индикатора: красный, зеленый и оранжевый. Когда стали исследовать ящики, обнаружили, что при нажатии кнопки  $\alpha$  на первом ящике, на втором всегда загорается красный индикатор, а при нажатии кнопки  $\beta$  — зеленый. Оранжевый индикатор, казалось, вообще не используется. После многократных повторений ученые убедились:

$\alpha$  является причиной красного

$\beta$  является причиной зеленого

Сигнал передавался через провод. Было очевидно, что, когда нажимали кнопку  $\alpha$ , в проводе происходило что-то одно, а когда кнопку  $\beta$  — что-то другое. Возможно, различалось количество импульсов тока, а может, разным было напряжение. Для того чтобы выяснить, в чём состояло отличие, ученые подключили к кабелю «прослушивающее» устройство. И вскоре обнаружили, что, когда нажималась любая из кнопок, по кабелю передавался поток

<sup>68</sup> Dennett D. Darwin's Dangerous Idea: Evolution and the meaning of Life. New York, 1995. P. 412.

импульсов с интервалами, как бы нулей и единиц, всего 10 000 битов. Но каждый раз последовательность была разной. Что в таком случае определяло, «какого цвета» этот поток?

Чтобы разгадать головоломку, ученые вскрыли ящик с цветовыми индикаторами. Там они обнаружили компьютер с огромным процессором, памятью и программой, занимавшей почти весь жесткий диск. Когда на процессор поступала последовательность из 10 000 битов от другого ящика, процессор осуществлял несколько миллиардов операций, в результате которых он выдавал 0 или 1. Если на процессор приходила та же самая комбинация, он выдавал тот же самый результат, 0 или 1, зеленый или красный. Ничего сверхъестественного не происходило. Кроме одного любопытного факта. Компьютер во втором ящике перед выдачей одного и того же ответа на один и тот же запрос проходил через разные физические состояния. Это объяснялось тем, что программа записывала каждый входящий запрос, и состояние памяти компьютера никогда не оставалось прежним. Наблюдая за работой компьютера, ученые решили, что любая последовательность из 10 000 битов будет определяться им как красная или зеленая. Гипотеза, однако, провалилась, как только они стали подменять последовательности битов, исходящие из первого ящика, новыми, собственными. Вдруг заработал оранжевый индикатор, как если бы второй ящик обнаружил мошенничество.

На любую новую последовательность, введенную учеными самостоятельно, загорался оранжевый. Нет, нельзя было сказать, что компьютер как-то отличал последовательности, «сделанные вручную», от последовательностей, исходящих из первого ящика. Когда вручную дублировались последовательности из первого ящика, второй ящик работал нормально, т.е. выдавал зеленый или красный. Вот только если изменялся хотя бы один бит, загорался оранжевый индикатор. Выходило, что все возможные последовательности делятся на зеленые, красные и... оранжевые.

Только оранжевых было значительно больше, на порядки больше. Это делало закономерность с красным и зеленым индикаторами ещё более интригующей. Что все-таки определяло, каким будет «цвет» последовательности символов? В каждом конкретном случае ничего удивительного не было. Ученые могли проследить каузальную цепочку, начиная от ввода любой последовательности в компьютер до вывода им результата. В этом смысле аппарат был полностью детерминирован физическими законами. Однако то, что было удивительно, *так это отсутствие возможности предсказать*, исследуя одну лишь последовательность символов, какой результат выдаст компьютер. Ученые знали лишь, что строка будет оранжевой с почти стопроцентной вероятностью, кроме случаев, когда она была выдана первым ящиком. В последнем случае с вероятностью миллиард к одному она была либо красной, либо зелёной. Но никто *не мог заранее определить один из двух цветов*, не пропустив эту строку через программу в ящике Б. Ученые решили, что разгадку даст только вскрытие первого ящика.

Они открыли этот ящик и обнаружили в нём еще один компьютер с гигантской памятью и программой. При нажатии кнопки, программа обращалась к встроенным часам, переводила время в бинарное выражение и с помощью этого кода выбирала, к какой части памяти обратиться, чтобы подготовить последовательность для второго ящика. Стало понятно, что именно использование внутренних часов гарантировало, что по кабелю всегда передавалась новая комбинация. При этой псевдо-произвольности, всегда, когда нажималась кнопка  $\alpha$ , загоралась красная лампочка, а  $\beta$  — зеленая. Впрочем, ученые, наконец, обнаружили аномалию: примерно раз в миллиард случаев происходила «ошибка», и после нажатия кнопки  $\alpha$  загорался зеленый цвет. Эта незначительная погрешность сделала аппарат ещё более загадочным для ученых.

Но однажды в лабораторию зашли два программиста, Эл и Бо, которые создали этот аппарат, и всё объяснили. Когда-то они оба работали над экспертными системами: создавали базу знаний, в которой должны были содержаться все истинные утверждения о мире и выводимые из них утверждения. Они писали эти программы на разных языках программирования. Оба преуспели в этом деле так, что скоро их программы могли генерировать бесконечное множество истинных утверждений: «Дом больше семечка», «Корова — самка домашнего быка», «Снег бел». Но, как выяснилось, экспертные системы не могли дать полезного нового знания человечеству. И программисты забросили свои проекты, превратив разработки в философскую игрушку.

Они выбрали способ транслировать (переводить) утверждения между машинами и соединили программы так, что при нажатии кнопки  $\alpha$  первая машина отправляла истинное утверждение, а при нажатии  $\beta$  — ложное. Вторая машина сопоставляла утверждение на входе с собственными «убеждениями» и, если получала истинное утверждение, подтверждала его красным сигналом, а если ложное — откликалась зелёным. Когда данные на входе невозможно было обработать, т.е. когда они выражали бессмыслицу либо содержали некорректные выражения, машина выдавала оранжевый цвет. Конечно, когда ученые вносили коррективы в передаваемые сигналы без понимания того, что они собой должны представлять, это было самой распространённой реакцией. Так раскрылось, что удивительное свойство красных последовательностей было свойством «быть истинным утверждением», а зелёных — «быть ложным утверждением». Неожиданно, поняв устройство аппарата, ученые сами смогли создавать сколько угодно зелёных и красных последовательностей. Они даже открыли последовательности, которые сначала были красными, а потом становились зелёными, как, например, сообщение: «это сообщение передавалось менее 100 раз». Ведь на 101 раз это сообщение должно было становиться зелёным.

### 3. Интерпретация мысленного эксперимента

Мысленный эксперимент «Два черных ящика» хорош своим почти детективным сюжетом, и вряд ли есть читатель, который хотя бы чуть-чуть не был заинтригован любопытным аппаратом и его свойствами. Однако нас интересует польза, которую философские исследования ментального могут получить от подобной иллюстрации. Не является ли эта громоздкая мысленная конструкция чем-то вроде «машины Руба Голдберга», чрезвычайно сложным и запутанным устройством для осуществления простейшей функции? Быть может, задачу, которую решает мысленный эксперимент, можно решить более лаконичным образом? Да и в чем состоит эта задача? Эта история, насколько мне известно, публиковалась в двух книгах Деннета: «Опасной идее Дарвина» (1990) и последней монографии «Помпы для интуиции» (2013)<sup>69</sup>. И в обоих случаях Деннет представлял очень краткое разъяснение этого мысленного эксперимента.

Для его философского стиля характерны скорее яркие образы и сравнения, нежели четкие дефиниции и формально структурированные аргументы. Поэтому я вынужден представить собственную интерпретацию. С моей точки зрения, мысленный эксперимент Деннета иллюстрирует существование высокоуровневых свойств, а также их каузальную эффективность.

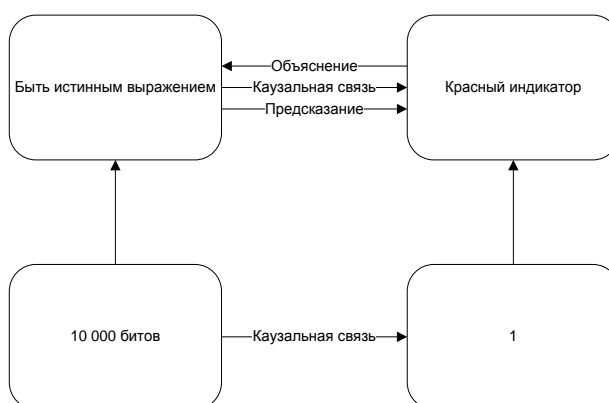


Рис. 3. Два черных ящика Д. Деннета

<sup>69</sup> Также эта история была опубликована в журнале *Activitas Nervosa Superior*, 2009. #52. Vol. 2. P. 81-84.

На базовом уровне в описанной системе из двух ящиков-компьютеров имеется последовательность символов, нулей и единиц. А на высоком уровне — свойство «быть истинным высказыванием», «быть ложным высказыванием» или «быть бессмысленным выражением». Поскольку каждый раз в ящик Б передаётся новое закодированное сообщение, из одной только последовательности символов *невозможно предсказать*, какой индикатор загорится. Также с помощью одной только ссылки на последовательность входящих символов *невозможно объяснить* результат, цвет индикатора. На вопрос, почему в случае отправки некоторых последовательностей загорается красный индикатор, а не какой-то другой, есть только один удовлетворительный ответ: потому что эти последовательности в закодированном виде выражают истинные утверждения о мире. По замыслу Деннета, единственным способом объяснить и предсказать, почему и какой индикатор загорится, является ссылка на свойство высокого уровня, в данном случае — семантическое. Результатом мысленного эксперимента я считаю следующие два тезиса:

- (1) Механическая система, состоящая из двух соединенных ящиков и спроектированная для выполнения синтаксических операций, проявляет семантические свойства. То есть помимо низкоуровневых, базовых свойств, она обладает высокоуровневыми свойствами.
- (2) Свойствами, которые определяют регулярность и каузальную связь между нажатием кнопок на одном ящике и вспышками индикаторов на другом, являются эти семантические, высокоуровневые свойства последовательностей.

Синтаксические свойства, т.е. последовательности битов, не содержат какого-то общего паттерна и потому не могут обосновывать каузальную закономерность. Вот, к примеру, человек идет по песчаному пляжу и оставляет за собой след. Этот след зависит от формы подошвы и веса человека, но никак не от цвета и материала подошвы. Вымышленная

ситуация Деннета построена так, чтобы семантическое содержание представлялось как каузально релевантное свойство, как форма или вес по отношению к следу, а не как цвет или материал.

Можно ли было каузальную действенность семантических свойств в механической системе продемонстрировать как-то иначе? Аппарат Деннета напоминает светофор, управляемый вручную. Две кнопки на одном устройстве зажигают соответствующие лампочки на другом. Такая модель светофора демонстрировала бы (а) *каузальную регулярность*. Но этого недостаточно для философских целей. Для объяснения каузальной регулярности и предсказания состояний такого «светофора» достаточно описаний физического уровня. Чтобы появилась (б) *возможность описывать работу* «светофора» на *интенциональном уровне*, система должна демонстрировать автономность, хотя бы в какой-то степени управляться изнутри и реагировать избирательно на входящие сигналы. Тогда можно будет описывать процесс так: отправленные первым компьютером данные второй компьютер «считает» соответствующими такому-то цвету. Кажется, необходимую автономность система начинает демонстрировать тогда, когда впервые загорается оранжевый индикатор. Но только возможности приписывать интенциональные состояния тоже недостаточно. Необходимо (в) *заблокировать возможность предсказаний на основе только лишь синтаксического анализа*. Для этого нужно создать ключ для генерации бесконечного числа последовательностей, сохранив при этом регулярность в работе системы. Ключ должен находиться вне системы, чтобы его нельзя было обнаружить и расшифровать. Таким ключом в аппарате Деннета является описание внешнего мира. «Именно тот факт, что каждый ящик несет в себе многообразную семантическую связь с одним объектом (описанием мира — Д.В.), выраженную, однако, в разной "терминологии" и по-разному аксиоматизированную, является основой регулярности»<sup>70</sup>.

---

<sup>70</sup> Dennett D. Intuition Pumps and Other Tools for Thinking. N.Y.: W.W. Norton & Company, 2013. P. 195.



Возможно, ключом мог быть также какой-то внешний текст, например, текст «Войны и мира» на двух языках. Но тогда синтаксические последовательности, содержащие этот текст в закодированном виде, были бы конечными. А значит, со временем, даже не зная ключа (текста Толстого), ученые смогли бы предсказывать, какая последовательность будет «красной», а какая — «зелёной». Мне кажется, установка действительно соответствует минимальным требованиям и не чересчур усложнена. Но достигает ли Деннет с её помощью поставленных целей? Удаётся ли ему показать, что семантические высокоуровневые свойства способны обладать каузальной эффективностью в механических системах?

#### 4. Критика телеофункционализма и возможные ответы

##### *А. Невыполнимость условий мысленного эксперимента*

На первый взгляд, описанный выше аппарат представим и возможен. Но уверенность в этом рассеивается, стоит лишь обдумать механизм его функционирования более тщательно. Видимо, программисты фактически все-таки не могли реализовать аппарат так, чтобы свойства «быть истинным» или «быть ложным высказыванием» регулярно вызывали соответствующую цветовую индикацию. Во-первых, истинность части высказываний должна со временем быть опровергнута, во-вторых, что-то должно было измениться с момента создания базы и, в третьих, то, что было истинным для машины А, могло не быть истинным для машины Б. К примеру, высказывание «Изотоп антигелия-4 был зарегистрирован на ионном коллайдере RHIC» могло быть ложным до 2011, истинным после 2011г. и даже бессмысленным до появления слов «изотоп», «антигелий» и «ионный коллайдер». А выражение «Мою программу написал Эл» истинно для первой машины и ложно для второй. Таким образом, в работе аппарата отсутствовали бы регулярности, которые описаны в мысленном эксперименте. И, при нажатии  $\alpha$ , на ящике Б иногда загорался бы красный индикатор, а иногда — зеленый и даже оранжевый.

Парировать возражение можно небольшим усовершенствованием эксперимента. Допустим, что программисты предварительно договорились, какие истины включить в экспертную программу. Тогда их базы данных должны соответствовать друг другу, а регулярность работы аппарата — соблюдаться. Но в таком случае неясно, что является каузально эффективным свойством. «Быть истинными» или «быть ложным выражением» в качестве таких свойств уже не работают. Т.к. в случае с предварительно согласованной базой высказываний программисты могли бы допустить совместные ошибки. Также в ходе исследования ящичков ученые могли случайно набрать последовательность символов, которая была бы истиной, но при этом не вошла бы в согласованный массив знаний Эла и Бо. В качестве кандидата на каузальную эффективность можно принять свойство «быть истинным выражением для... машин А и Б»<sup>71</sup>. Но оправдано ли это?

*В. Машина не обладает семантическими свойствами*

Далеко не очевидно, что машины могут иметь убеждения. Для многих даже очевидно обратное. По крайней мере, машины, как они описаны в мысленном эксперименте Деннета, точно не представляются мыслящими существами. Поэтому эти машины вряд ли могут считать что-то истинным или ложным. Но высказывание может быть истинным или ложным для самих программистов. Тогда правильной всего свойством верхнего уровня считать только свойство «быть истинным или ложным для Эла и Бо». Иначе говоря, в аппарате Деннета нельзя найти высокоуровневых интенциональных свойств, для их наличия нужны программисты. Таким образом, мысленный эксперимент возвращает нас к исходным пунктам, к необъяснимой и противоречивой ментальной каузальности.

Так допустимо возражать, если предполагать возможность каузальной связи между событиями, отстоящими во времени и пространстве.

---

<sup>71</sup> Не факт, что Деннет согласился бы с такой поправкой. Возможно, она разрушает его построение. Ведь семантические характеристики происходят именно из связи с внешним миром. Они связаны с тем, что в последовательностях битов существует репрезентация мира, а не каких-то внутренних состояний.

Ментальные состояния агентов, безусловно, имели отношение к работе этого аппарата тогда, когда программировалась база знаний. Но в момент нажатия кнопок, они уже не имеют значения. Для верности можно представить, что два ящика были раскопаны следующими поколениями на месте, где раньше находился Лос-Анджелес, например, в 3105 году. К тому времени программисты давно уже умерли. Вряд ли уместно говорить, что каузальным действием обладают ментальные состояния уже несуществующих людей.

На это можно ответить: пусть между ментальными состояниями программистов и состоянием компьютеров в момент исследования нет контакта, но генеалогическая связь прослеживается. Ментальные состояния программистов повлияли на исходные состояния компьютеров, а эти исходные состояния каузально связаны с каждым последующим их состоянием, и так — до самого последнего, до момента, когда ученые начали исследовать ящики А и Б в лаборатории. Помогает ли такое рассуждение преодолеть временной зазор? Допустим, но тогда можно дальше усовершенствовать мысленный эксперимент. Например, можно представить, что массив истинных данных был получен путем сбора информации о внешнем мире непосредственно машинами-роботами. Тогда ментальные состояния не будут иметь прямого отношения к последовательностям битов. Впрочем, скептик и здесь, наверное, не будет удовлетворён. Он может спросить: а реализуем ли такой сценарий? Возможно ли создать самообучающуюся энциклопедию? И даже если возможно, что будет обозначать информация, собранная машиной? Разве машина вообще может быть чувствительна к семантическим содержаниям? Она способна распознавать и совершать операции только с формами, содержание которых может быть интерпретировано сознательными мыслящими агентами. Но сама машина не способна интерпретировать эти сигналы. Только для мыслящих существ последовательность битов обозначает истинное высказывание или

что-то ещё. Только мыслящие существа наделяют материальные объекты смыслом. Машина — только инструмент, а никак не мыслящий агент.

Это возражение достаточно популярно. Одним из самых влиятельных сторонников такой позиции является уже известный нам и Деннету Джон Сёрл. Поэтому неудивительно, что в текстах Деннета можно найти развернутый подходящий ответ. Это — ещё одна вымышленная история, или помпа для интуиции. Новая история Деннета должна подорвать уверенность в том, что только сознательные разумные существа наделяют знаки значением.

«Итак, предположим, вы решили по каким-либо причинам продлить свою жизнь и узнать, что произойдет в двадцать пятом веке»<sup>72</sup>, — начинает свою историю автор. Чтобы дождаться этого момента, придётся длительное время провести в состоянии глубокого сна, комы. Когда вы проснетесь, ваших детей и внуков уже не будет. Так что нужно заранее позаботиться о сохранности тела, например, создав специальную защитную капсулу. Создание такой капсулы для хранения — не единственная забота. Нужно обеспечить её снабжение и безопасность. Выбирать придётся между двумя вариантами. Первый — сразу найти идеальное место, куда будут подведены источники питания и энергии. Этот вариант плох тем, что обстоятельства могут измениться и идеальное место может перестать быть таковым. Второй вариант — создать движущуюся конструкцию, гигантского робота, и установить капсулу в него. Вторая, более сложная идея, требует, чтобы робот был способен «выбирать» то или иное действие исходя из ваших интересов. У него должны быть органы восприятия, способность управлять движением, даже планировать. В конечном итоге сложно предсказать, какие неприятности могут произойти в следующие столетия. К тому же нельзя исключать возможности создания других таких же роботов, которые будут бороться за ограниченные ресурсы, — продолжает Д. Деннет. Вполне

---

<sup>72</sup> Dennett D. The Intentional Stance. Cambridge, Massachusetts: The MIT Press, 1996 (1987). P. 295.

возможно, роботу придется состязаться или сотрудничать с ними. В таком случае его нужно запрограммировать так, чтобы, обладая основной целью (обеспечить ваше выживание), он мог выстраивать и свои, промежуточные. Вполне возможно, эти цели даже в какой-то момент станут преобладать, даже вступят в противоречие с основной целью. Его может толкнуть на самоубийство другой робот, стремящийся подчинить все окружающее своей миссии. Или заставить помогать и защищать другой организм. В соответствии с описанием Деннета, очевидно, что программа и поведение такого робота должны быть очень сложными. Такой робот даже, наверное, прошел бы Тест Тьюринга: по поведению его сложно будет отличить от обычного человека. Ему можно будет приписать подлинные интенциональные состояния. Он будет вести себя так, будто мыслит, имеет убеждения и желания... хотя все это будет запрограммировано вами заранее. Разве он не будет мыслящей машиной?

Если все это не убедило вас, значит, вне зависимости от деталей, вы не способны допустить интенциональность у любых механических объектов. Но если это так, вы... не способны допустить существование интенциональности даже у людей. История в руках Деннета вдруг обретает неожиданный оборот. До этого момента она могла показаться несурзой и уж слишком надуманной фантастикой. Но Деннет объявляет, что она — всего лишь немного замаскированная история эволюции человека, вариация на тему, представленную в знаменитой книге этолога и эволюционного биолога Ричарда Докинза «Эгоистичный ген»<sup>73</sup>.

Две стратегии выживания в биологической истории обозначают стратегии выживания двух царств: растений и животных. А сам робот — это мы. Как ни парадоксально, люди — это машины для выживания генов, продукт их творчества. Это их эгоистичные планы мы выполняем, их цели, а не наши собственные, являются нашим *raison d'être*. В своей основе гены,

---

<sup>73</sup> Dennett D. The Intentional Stance. Cambridge, Massachusetts: The MIT Press, 1996 (1987). P. 298.

конечно, весьма просты. Но они обладают бессознательной целенаправленностью: «желанием» сохранять собственную «жизнь» и преумножать число своих копий. И им удается реализовать свои планы. Жизнь одного гена намного дольше человеческой. Она почти бесконечна. Так кто кем в результате управляет, мы генами или гены нами? Кто стоит у руля? Правда, управление генов организмом происходит не в реальном времени, не как управление, скажем, марионеткой. Генам приходится предвидеть все заранее. Поскольку точные условия окружающей среды в будущем для них неизвестны, они строят самые общие предположения о рисках и возможностях. Направляя синтез белка, они программируют функциональность организма задолго до того, как «механизм выживания» встретится с реальными опасностями. Когда их работа проделана, организму приходится справляться самостоятельно. Но тем не менее, именно гены остаются в центре эволюционного процесса и являются авторами «наших» замыслов.

Деннет считает, что наличие исходно семантической составляющей в нашем сознании — это иллюзия. «[Мозг — Д.В.] не может быть создан, чтобы решать невозможную задачу, но он может быть создан так, чтобы решать её приближенно, подражая работе невозможного объекта (семантического процессора)»<sup>74</sup>. Мозг решает эту задачу, используя соответствия между структурными (синтаксическими) закономерностями в своем внутреннем устройстве и окружающей среде. Правда, что мы оперируем интенциональным содержанием, точнее, как бы оперируем. Но даже люди не обладают первичной, внутренней интенциональностью. Значения состояния и действия людей наделяются постепенно, в ходе эволюционного процесса, во взаимодействии с миром. И этот процесс не имеет абсолютного завершения. Семантические свойства в любом случае производны. И четкой грани между только лишь «синтаксическими

---

<sup>74</sup> Dennett D. The Intentional Stance. Cambridge, Massachusetts: The MIT Press, 1996 (1987). P. 61.

операциями» и «семантическими операциями» не может быть. Допущение чёткой разделительной линии между ними — это тупиковый путь, ошибочный путь эссенциализма, на который философия всё время норовит свернуть, может быть, еще со времён Сократа. Абсурдность философского поиска сущностных свойств и проведения повсюду границ Деннет иллюстрирует с помощью псевдоаргумента, аргумента о невозможности существования млекопитающих<sup>75</sup>:

1. Каждое млекопитающее имеет мать-млекопитающее.
2. Если млекопитающие существуют, их к настоящему времени было ограниченное количество.
3. Но у самого первого млекопитающего должна была быть мать-млекопитающее, и т.д. Это приводит к противоречию с пунктом 2.

Из этого аргумента будто бы следует, что млекопитающих не существует. Но это же абсурд. Несмотря на формальную корректность аргумента, следствие ложно. Ошибочны посылки, скрытое в них предположение наличия чёткой грани между млекопитающими и их предками. И полностью автономные семантические двигатели — это такая же химера, как «первое млекопитающее». Млекопитающие постепенно эволюционируют из рептилий. И между двумя видами стоит множество транзитных «как бы млекопитающих» и их матерей (из отряда терапсидов). Аналогичным образом эволюционируют и семантические двигатели. Они не появляются внезапно. Более того, идеального семантического двигателя не существует. Поэтому вряд ли стоит удивляться, что воображаемый аппарат из двух ящиков не может абсолютно точно схватить свойство «быть истинным выражением». Он работает на «почти истинных выражениях»... в большинстве случаев. Как, впрочем, и мозг человека.

---

<sup>75</sup> Dennett D. Intuition Pumps and Other Tools for Thinking. N.Y.: W.W. Norton & Company, 2013. P. 178.

*С. Высокоуровневые свойства в принципе не обладают каузальной эффективностью*

Впрочем, оппонент Деннета мог бы усомниться и в другом тезисе автора: в том, что в мысленном эксперименте с «Двумя черными ящиками» каузально эффективным свойством является какое-либо высокоуровневое свойство. Оппонент мог бы попытаться доказать, что в реальности каузальная сила есть только у другого, низкоуровневого физического свойства, например, у свойства «быть частью такого-то блока данных». При реализации базы вероятней всего, что истинные выражения располагались в каком-то одном блоке, скорее всего, даже пространственном, а ложные — в другом. Тогда каузальным действием обладают не семантические характеристики, а, например, пространственные.

Пространственное расположение действительно могло быть общностью, отличающей одни наборы битов от других. Но данные на жестком диске могли располагаться по-разному. Даже если пространственная общность существовала, она не была обязательной. При смене «места» хранения последовательности она всё равно вызывала бы ту же ответную реакцию второй машины. Т.е. пространственная общность являлась бы общностью случайной, в отличие от общности семантической. Это была бы случайная корреляция, а не каузальная зависимость.

Аналогичным образом, в качестве претендента на каузальную роль не подходит и характеристика «принадлежать к блоку данных, связанных с  $\alpha$ », т.к. вполне возможно, чтобы последовательности были связаны иначе, но регулярность бы соблюдалась. [Под этой регулярностью я подразумеваю отправку на второе устройство все время различных последовательностей битов без синтаксических и иных общностей и способность второго ящика отсортировать их должным образом]. Для примера представим такое устройство, в котором металлический стержень соединяет между собой взрывчатое вещество и открытый огонь. Пока теплопроводность металлического стержня низкая, взрыва не происходит, т.к. недостаточно



тепла передаётся от огня. Но стоит повысить теплопроводность стержня (например, заменив его на стержень с другим химическим составом), как произойдет взрыв.

Существует, однако, еще один верный способ произвести взрыв такого устройства. По Закону Видемана-Франца, теплопроводность всегда при нормальных условиях коррелирует с электропроводностью (т.к. оба этих свойства обуславливаются движением свободных электронов). Соответственно, повышая электропроводность стержня в устройстве (заменяя химический состав стержня), мы также создаём условия для взрыва. При этом кажется, что электропроводность является только сопутствующим свойством, а не казуально эффективным<sup>76</sup>.

Описанное выше устройство демонстрирует, что только одно из двух коррелирующих между собой свойств может быть казуально эффективным. При этом в случае с данным устройством корреляция высокоуровневых свойств систематическая, регулярная. А в случае с пространственным расположением данных, корреляция явно случайная, нерегулярная, несистематическая. Последовательности битов могут располагаться как угодно, при этом истинные выражения, «выражения, истинные для машин Эла и Бо», будут вызывать один и тот же эффект. Контрфактический анализ помогает определить, какое из свойств антецедента является казуально эффективным, а какое — нет. И именно семантическое свойство выдерживает контрфактическую проверку, а не низкоуровневое физическое.

#### *D. Функциональные свойства не обладают казуальной эффективностью*

Автор мысленного эксперимента с двумя ящиками предполагает, что семантические свойства казуально эффективны. При этом мы ранее установили, что эти свойства Деннет считает функциональными. Но могут ли функциональные качества быть казуально эффективными? Оппонентов в

---

<sup>76</sup> Подобный пример приводит Блок в статье «Может ли разум изменить мир?». Block N. Can the Mind Change the World? // Boolos G. (ed.). Meaning and Method. Essays in Honor of Hilary Putnam. Cambridge: Cambridge University Press, 1990. P. 146.

этом случае тоже достаточно, как и в случае с утверждением о принципиальной способности компьютеров иметь интенциональные состояния.

Возьмём для рассмотрения несколько свойств, определённых функционально, и протестируем интуиции в отношении их каузальной эффективности. Для начала возьмем свойство «провокационность». Это — функциональное свойство, т.к. оно определено в терминах своего эффекта. Считается, что мулета тореадора провоцирующим образом действует на быка, она возбуждает в нём агрессию. Таким образом, мулета тореадора обладает свойством «быть провокационной». Но она также обладает свойством «быть красного цвета». По общим представлениям, быка провоцирует красный цвет. Красный цвет каузально эффективен в отношении агрессивного поведения быка. Можно ли то же самое сказать в отношении «провокационности» мулеты? Провоцирует ли быка «провокационность» мулеты? «Провокационность» мулеты — это второпорядковое свойство. Как и другие функционально определенные свойства, оно зависит от наличия первопорядкового свойства. И именно свойство «быть красного цвета» находится в каузальных отношениях с событиями на входе и выходе. Видимо, в «провокационности» мулеты нет дополнительной каузальной работы. Бык не достаточно умен, чтобы осознать, что мулета провокационна. Ему достаточно красного цвета. Провокационность мулеты может спровоцировать членов Всемирного общества защиты животных, но не быка.

А вот — другой пример, «снотворность». Это свойство тоже функциональное и тоже второпорядковое. Оно определено в терминах эффекта, результирующего состояния — сна. Но оно предполагает существование первопорядкового химического свойства, вызывающего сон. «Быть снотворным» — значит обладать химическим свойством, каузально эффективным в отношении сна. Одни снотворные эффективны в силу

наличия в них барбитуратов, другие — бензодиазепинов, и т.д., но вряд ли какие-то — в силу свойства «снотворности». Только в случае, если пациент сам убеждает себя в «снотворности» лекарства, т.е. в случае эффекта «плацебо», лекарство может приводить к сну в силу самой «снотворности». Даже если химический состав медикамента каузально неэффективен в отношении сна, убеждение пациента в действенности препарата может быть «снотворным». Это, однако, особый случай. В целом, вероятно, что свойство «быть снотворным» как функциональное свойство, как правило, каузально бездейственно.

Описанные выше примеры функциональных свойств и соответствующие рассуждения приводит философ из Нью-Йоркского университета, Нед Блок, в статье «Может ли разум изменить мир?»<sup>77</sup>. По его мнению, такие примеры показательны. Они иллюстрируют тот факт, что функциональные свойства не могут быть каузально эффективными. Исключениями являются те особые случаи, когда они каузально эффективны в силу убеждений в их действенности разумного существа. В стандартных же случаях, т.е. когда функциональные свойства определены посредством их результирующего эффекта, и этот эффект не возникает в результате осознания разумного существа, они каузально бессильны<sup>78</sup>. Блок считает, что интуиции в отношении этих примеров верны.

Сомневаться в каузальной эффективности функциональных свойств приходится и в силу других оснований. Блок убежден, что для каузальной связи нужны контингентное, эмпирическое и номологическое отношения между событием и его следствием. А таких связей здесь нет. Медикамент обладает свойством «снотворности» потому, что обладает *другим* свойством, которое в случае употребления этого медикамента номологически

---

<sup>77</sup> Block N. Can the Mind Change the World? // Boolos G. (ed.). Meaning and method. Essays in honor of Hilary Putnam. Cambridge: Cambridge University Press, 1990.

<sup>78</sup> Block N. Can the Mind Change the World? // Boolos G. (ed.). Meaning and method. Essays in honor of Hilary Putnam. Cambridge: Cambridge University Press, 1990. P. 156.

гарантирует сон (или повышает его вероятность). Т.е. номологическое отношение имеет место не между функциональным свойством и эффектом, а между другим, химическим или физическим свойством и эффектом. «Свойство снотворности» же находится в логической, концептуальной связи со своим эффектом. Таким образом, объяснение со ссылкой на функциональное свойство носит аналитический характер. Между функциональным свойством и действием есть логическая связь. Чтобы между ними присутствовала еще и номологическая связь, должны быть специальные основания, как, например, в следующем случае. Представьте, что любимое свойство моей тёти — это «снотворность», а дяди — «быть сном». Между этими свойствами существует логическая связь. Но вдобавок между этими свойствами есть ещё и номологическая связь: мой дядя отслеживает любые изменения в любимом свойстве моей тёти так, чтобы в каждый момент его любимое свойство логически следовало из её любимого свойства. Вот в этом случае имеет место двойная связь. Но её нет в стандартной ситуации с функциональными свойствами.

Однако в рассуждениях Блока настораживает уже то, что в качестве примеров он использует заведомо сомнительные случаи и свойства. Известно, что красный цвет безразличен быку, т.к. у него дихроматическое зрение (впрочем, Блок это и сам не отрицает, и тем не менее — использует этот пример). Бык одинаковым образом реагировал бы на красную, синюю, или зеленую ткань. Тогда что, если не красный цвет, возбуждает быка? «Провокационность» мулеты? Кажется, здесь Блок сам оказывается в ловушке собственной иллюстрации. Даже если мы согласны с его антифункционалистскими настроениями, этот пример не выполняет своей функции. Со снотворным качеством пример получается тоже не очень удачным. Блок как бы нарочно напоминает о псевдообъяснениях. Вспомните мольеровского «Мнимого больного»: опиум действует снотворно, так как в нём имеется «снотворная сила». Мольер использует это объяснение с

сарказмом, высмеивая неосведомленность объясняющего. Возможно, на эти ассоциации частично опирается и Н. Блок. Но это — психологический приём, а не убедительное рассуждение. Вряд ли сторонник функционализма сам стал бы использовать «снотворность» в качестве иллюстрации каузально эффективных функциональных свойств.

Мне кажется, впрочем, что это соображение не имеет конечной убедительности. Вспомним, как Э. Энском и Г. Райл (а кроме них — также А. Мелден и др.) критиковали действенность ментальных состояний с помощью Аргумента логической связи. Тогда речь шла о ментальных состояниях безотносительно их функционалистской интерпретации. И тем не менее, рассуждения Блока напоминают этот аргумент. Тогда можно выстроить защиту наподобие той, что выстраивалась в ответ на АЛС. Можно попытаться разъединить строгую концептуальную связь, например, показав, что наличие того или иного функционального свойства не обязательно приводит к ожидаемому эффекту. Функциональные свойства взаимодействуют с другими функциональными свойствами, и результат этих взаимодействий могут давать различный эффект. Если бы существовала жёсткая логическая сцепка, таких вариаций не было бы.

Возьмём свойство «быть будильником». Это — функциональное свойство, т.к. оно определяется в терминах функции, роли, которую играет это устройство, а не материала или его внутреннего механизма. Будильник — это устройство (электронное, механическое или даже органическое!), в заданный момент подающее сигнал к пробуждению. Допустим, некто пытается дать такое каузальное объяснение: он проснулся в шесть утра от звонка будильника. По логике Блока, раз будильник будит, то логически следует, что он должен разбудить. Но разве из того, что будильник будит, следует то, что звонок раздастся в шесть утра? Звонок мог раздаться в любое время, причем даже не в то время, на которое был поставлен: часы могли отстать или заторопиться, да и будильник мог вообще сломаться или быть

незаведённым. К тому же тот, кого будят, может и не проснуться, несмотря на то, что функциональное свойство будильника — будить. «Быть будильником» и «будить», безусловно, как-то логически состыкованы, но без номологической связи одно все равно не будет причиной другого.

Ранее было дано функциональное определение боли: состояние, вызванное повреждением органов, сопровождающееся «убеждением, что что-то повреждено», «желанием это прекратить» и ведущее к реакции избегания. Но разве обязательно, чтобы испытывающий боль демонстрировал «реакцию избегания»? Если человек, к примеру, под пытками стремится убедить своих палачей, что он не способен чувствовать боль, не стонет, не выражает страданий в мимике, разве он не испытывает боль? Или если перед быком тореадор размахивает провоцирующей мулетой, но бык так и остаётся бездвижным (это ведь вполне возможно) — что, мулета перестает быть провоцирующей или перестает быть красной? Нет, просто бык мог устать. Ментальные состояния, определенные функционально, всегда взаимодействуют с другими ментальными состояниями (убеждениями и желаниями) и не могут образовывать жёстких логических связей с последующим поведением. Таким образом, как это уже было сформулировано в контексте рассмотрения функционализма, *убеждения*, определенные функционально, представляют собой состояния, определённые данными органов восприятия, а также другими убеждениями, и при взаимодействии с желаниями формирующие поведение или действия. А *желания* — это состояния, находящиеся в связи с другими желаниями и намерениями, при взаимодействии с убеждениями формирующие поведение или действия.

Впрочем, сам Деннет, как я уже говорил, не вступает в открытую полемику с Блоком и его сторонниками. Его мысленные эксперименты лишь намечают направление, в котором можно решать проблему ментальной каузальности, но не представляют собой явного решения. Пожалуй, здесь я

могу предложить собственный вариант решения проблемы ментальной каузальности. Это решение будет состоять из трех шагов. На первом шаге я применю тактику, подобную той, что применялась выше с Аргументом каузальных траекторий. Я представлю аргумент, аналогичный Аргументу исключения, но ведущий к очевидно ложным следствиям. Вторым шагом я приведу иллюстрацию каузальной эффективности высокоуровневых свойств и сопровожу ее анализом. На третьем шаге я покажу, что именно может не работать в механике Аргумента исключения Дж. Кима. Это трехчастное решение проблемы ментальной каузальности работает в рамках телеофункционалистской теории сознания.

## VI. Предлагаемое решение проблемы ментальной каузальности

### 1. Аргумент обобщения Н. Блока против Аргумента исключения

Зачастую лучшая тактика защиты — это нападение. И в данном случае такая тактика уместна. Можно показать, что Аргумент исключения, направленный против ментальной каузальности и нередуцируемых свойств, противоречит важным научным утверждениям. Он подвергает сомнению установки научного знания, в частности убеждение в каузальной эффективности свойств и объектов, входящих в теории специальных наук. Под специальными науками имеются в виду все науки, отличающиеся от фундаментальной физики. Чтобы показать это, необходимо найти сходный по логической структуре и опирающийся на те же ключевые посыпки Аргумент-двойник. Но он должен быть направлен не против ментальной каузальности, а против чего-то, что оспорить значительно труднее. Подобный метод аргументации был уже мной использован выше для критики Аргумента каузальных траекторий. В данном случае он тоже применим. В качестве Аргумента-двойника можно использовать уже разработанный Аргумент обобщения, предложенный Н. Блоком<sup>79</sup>. Аргумент называется так

---

<sup>79</sup> Block N. Do Causal Powers Drain Away? // Philosophy and Phenomenological Research. 2003. LXVII. July. P. 471-497.

потому, что обобщает, экстраполирует выводы, сделанные Кимом относительно эффективности ментальных свойств, на эффективность свойств, входящих в аппарат специальных наук. Суть рассуждения — в следующем:

Допустим, что все свойства макрообъектов определены свойствами их частей, вплоть до элементарных частиц. Тогда свойства макрообъектов супервентны на свойствах микрообъектов. При этом свойства микрообъектов являются достаточной причиной для всех событий с макрообъектами. Тогда у свойств макрообъектов нет каузальной роли и, следовательно, согласно принципу исключения, эти свойства эпифеноменальны (т.е. каузально неэффективны). Они проигрывают в конкуренции за каузальную эффективность. Эта логика применима и ко второпорядковым свойствам. Первостепенными свойствами являются свойства индивидов. Второстепенные свойства — свойства свойств первого порядка. Все второстепенные свойства супервентны на первостепенных свойствах. Если Аргумент исключения верен, то и второстепенные свойства не играют каузальной роли. Но такими являются все свойства объектов специальных наук. Следовательно, химические, биологические и физиологические свойства тоже не являются каузально действенными. Все объекты и их свойства за пределами физики являются эпифеноменальными, а законы специальных наук не схватывают каузальных регулярностей. Иными словами, «существует только физика и коллекционирование марок». Такой вывод, кажется, следует из Аргумента обобщения.

А что, если элементарные частицы физики делимы до бесконечности и фундаментального, базового уровня вообще не существует? На каком уровне тогда искать каузальные закономерности? А ведь это — реальная возможность, обсуждаемая современной физикой. Тогда проблема представляется еще более драматичной. Каузально эффективными перестают быть все известные нам объекты, и даже открытые в будущем объекты будут



каузально бездейственными. Теряется вообще надежда открыть каузальные законы. Поиск таких законов становится подобным заполнению водой бочки с пустым дном. С открытием каждого нового, более низкого уровня каузальность перетекает в него, и так до бесконечности. Выявление причин превращается в погоню за неуловимой целью. Кажется, перед нами новая дилемма: принимать Аргумент исключения и вместе с ним — Аргумент обобщения, требующий редукции всех специальных наук (да и значительной части физики), или отрицать Аргумент исключения.

В пользу второй альтернативы, отрицания АИ, кажется, достаточно доводов. С практической точки зрения специальные науки вполне оправдывают себя. Они работают с регулярностями, на которые можно положиться. К тому же доводом в пользу отрицания АИ является и ментальная каузальность. Если АИ поражает одновременно все супервентные свойства, ментальные свойства оказываются «в хорошей компании». А проблема ментальной каузальности становится частным случаем проблемы всех высокоуровневых наук. Это значит, что проблемы вообще может не быть. Если следствие, к которому ведет аргумент, абсурдно, нужно проверить посылки или форму аргумента. Не наш ли это случай? Быть может, здесь заложена какая-то ошибка, и Аргумент обобщения сводит к абсурду Аргумент исключения?

Такого казуса можно попытаться избежать. Сам автор Аргумента исключения Дж. Ким считает, что существует путь между Сциллой и Харибдой. Можно сохранить АИ против нередуцируемых ментальных свойств и не попасть в ловушку Аргумента обобщения. Для этого нужно показать, что утечка каузальной эффективности от высокопорядковых к низкопорядковым<sup>80</sup> свойствам не ведет к утечке каузальной эффективности от макро- к микрообъектам. Свойства объектов, — рассуждает он, —

---

<sup>80</sup> Здесь я вслед за Кимом нарочно использую термины «высокопорядковое» и «низкопорядковое» свойство, чтобы зарезервировать термин «уровень» исключительно за иерархией макро-микро.

действительно образуют особую иерархию. Над свойствами нижнего порядка надстраиваются свойства верхнего: над физическими — химические, над химическими — биологические и т.д. Ментальные свойства находятся где-то в верхней части иерархии. Они надстроены над биологическими, т.к. все ментальные свойства присущи только системам с биологическими свойствами (по крайней мере, до создания ИИ) и не наоборот. Но вся эта иерархия свойств не является иерархией от малого к большому. Она построена не движением от микро- к макрообъектам, а с помощью отношения реализации: физические свойства реализуют химические свойства, химические — биологические. По шкале микро-макро движения нет. Биологическими, ментальными и физическими свойствами обладает один и тот же объект. Таким образом, регресс от высокопорядковых свойств к низкопорядковым происходит внутри одного и того же класса объектов. К примеру, свойством «весить 80 кг» может обладать объект, обладающий свойством «быть Homo sapiens» и свойством «желать горячего чая». Разные свойства, свойства разного порядка, относятся к одному объекту. Таким образом, из Аргумента обобщения не следует бесконечный регресс в микро-макро иерархии. Макрообъекты обладают своими особыми каузальными свойствами, отличными от свойств составляющих их частей. Следовательно, атомы и составляющие их частицы не имеют абсолютной монополии на каузальную эффективность. И возможность деления эти частиц до бесконечности не представляет угрозы для каузальности в целом. Впрочем, так Ким блокирует только один вариант утечки каузальности: от макро- к микрообъектам. Но утечка может проходить и в другом направлении: от высокопорядковых свойств к первопорядковым.

Пусть составные объекты обладают каузальной силой, отличной от каузальной силы их частей, но возможно, только самые «простые» их свойства каузально эффективны? Утечка каузальности может произойти не вниз, с уровня макро- на уровень микрообъектов, а как бы «в сторону», на

одном уровне. Если только «простые», фундаментальные, физические свойства объектов каузально эффективны, свойства, которые изучают специальные науки, оказываются эпифеноменальными. Тогда законы этих наук не являются каузальными законами. Впрочем, остается ключевой вопрос: какие свойства должны попасть в категорию физических? Только масса, спин, электрический заряд? А, например, свойство «быть жидким» или «быть газообразным» должно попасть в категорию физических? Кажется, с одними только базовыми свойствами не объяснить растворяющее действие жидкостей или процесс кипения. Для того чтобы физический мир был каузально замкнутым, чтобы имеющиеся в нем феномены можно было полностью объяснять, требуется значительно расширять набор физических свойств. И в него, по-видимому, войдут свойства как элементарных частиц, так и атомов, молекул, клеток, сложных живых организмов, планет, звезд, т.е. свойства, приписываемые объектам в химии, биологии, физиологии, геологии т.п. Среди них будут и функциональные свойства. При этом такие свойства не будут каузально неэффективными, а только «каузально гетерогенными». Т.е. их действенность будет определяться их реализацией в других свойствах.

Разъяснения, представленные выше, отчасти спасают АИ от абсурдных следствий. В частности, они удостоверяют, что АИ не ставит под сомнение каузальные свойства макрообъектов. Но могут остаться сомнения в отношении следствий для специальных наук. Действительно ли из АИ не следует, что свойства объектов, приписываемые им в специальных науках, эпифеноменальны? Действительно ли сеть, образуемая АИ, захватывает только нередуцируемое ментальное содержание? Защита Дж. Кима построена на представлении о свойствах с микро-базой. Именно опора на это представление позволяет философу включить в физический каузально эффективный мир высокопорядковые свойства. Свойства с микробазой, по его мнению, — это любые свойства, у которых в качестве основания имеются

микросвойства. Все специальные науки оперируют именно такими свойствами. Вот свойства молекул воды, например, имеют микробазу, т.к. основаны на свойствах атомов водорода и кислорода. Свойства воды редуцируемы к свойствам  $H_2O$ , ведь вода, считает Ким, и молекулы  $H_2O$  редуцируемы друг к другу. И именно поэтому свойства воды не конкурируют со свойствами водорода и кислорода за казуальную эффективность. Ведь нечто не может конкурировать с самим собой. Если каузально эффективна основа, каузально эффективны и производные свойства. Так, психологические свойства каузально эффективны в силу своей реализации и тождественности со свойствами биологическими, биологические — с химическими и т.д. Видимо, это Ким и имеет в виду, когда говорит о каузальной гетерогенности свойств высокого порядка. С ментальными свойствами (по крайней мере, квалитативными), однако, ситуация другая. В отличие от свойств, приписываемых объектам в специальных науках, ментальные свойства не имеют микробазы и поэтому выступают конкурентами физических в борьбе за каузальную значимость. И, соответственно (по АИ), в этой борьбе они проигрывают.

Впрочем, в такой позиции есть еще брешь. Не только ментальные квалитативные свойства сопротивляются редукции. Вызывает сомнение возможность редукции высокоуровневых свойств в специальных науках к свойствам фундаментальной физики. И камнем преткновения здесь является множественная реализация. Некоторые высокоуровневые свойства могут иметь различную реализацию на более низком уровне. В высокоуровневых науках они могут рассматриваться как один естественный вид, а в низкоуровневых — как множество. Вот даже вода. Её тождественность с молекулами  $H_2O$  — уже слишком грубое обобщение. Фактически проба воды дает не только мономерные молекулы  $H_2O$ , но и полимерные, такие как  $(H_2O)_2$  и  $(H_2O)_3$ . Все разновидности молекул находятся в статистическом балансе, но постоянно образуют короткоживущие связи. Т.е. даже вода

может иметь множественную реализацию. Более того, как свидетельствуют некоторые химики, даже сама процедура идентификации воды состоит не в анализе составляющих её элементов, а тестировании её функциональных свойств. Определение естественного вида в химии может проходить не с помощью анализа составляющих, а посредством анализа реакций. Значит, многие важнейшие свойства, изучаемые специальными науками — свойства функциональные.

Примеров естественных видов и высокоуровневых свойств с множественной реализацией достаточно. Так, даже свойство «быть твердым телом» имеет как минимум две реализации: кристаллическая решетка (как у алмаза) и аморфная субстанция (как стекло). В кристаллах атомы расположены регулярно, с периодичностью. А в аморфных телах они колеблются вокруг хаотически расположенных точек. Но? тем не менее, и аморфные, и кристаллические твердые тела представляют собой один естественный вид, одно свойство. Они подчиняются единообразным законам, которые изучает подраздел физики — механика твердого тела.

Все эти соображения показывают, что естественные виды, фигурирующие в законах специальных наук, могут быть по-разному реализованы. А следовательно, они не редуцируются к одному свойству как к микробазе. А если они не редуцируются, значит, они конкурируют за каузальную эффективность с базовыми первопорядковыми физическими свойствами и в соответствии с принципом каузальной замкнутости физического мира должны проигрывать в этой конкуренции и быть эпифеноменальными. Таким образом, Аргумент исключения исключает из числа каузально эффективных свойств множество свойств, составляющих теоретическую базу специальных наук, подрывает статус их законов и, тем самым, сам сводится к абсурду. Ибо такое следствие вносит слишком много разрушений и, соответственно, неприемлемо. Все эти рассуждения приводят нас к мысли, что, если ментальная каузальность и выглядит на каком-то

основании проблематичной, она представляется не более проблематичной, чем каузальность второпорядковых свойств и свойств, изучаемых специальными науками. Т.е. ментальная каузальность находится «в хорошей компании». Это — одно из направлений критики Аргумента исключения.

Итак, можно критиковать Аргумент исключения с помощью Аргумента-двойника, такого как Аргумент обобщения, и его следствий. Другое направление критики — с помощью контрпримеров. Можно найти или придумать ситуации, в которых причинами чего-либо будут выступать не фундаментальные физические, а высокоуровневые свойства. Для критики аргумента Кима здесь я хочу предложить собственный пример, который должен продемонстрировать каузальную работу «исторических» свойств. Исторические свойства объектов не являются базовыми физическими свойствами. Если они имеют каузальное влияние, то оно может быть и у ментальных событий. Замысел предлагаемой мной истории сходен с замыслом автора «Истории про два черных ящика», но есть и важное отличие. Новый пример позволяет продемонстрировать, как одинаковые по фундаментальным физическим свойствам события благодаря свойствам высокого уровня могут привести к разным действиям. Это мне представляется преимуществом.

## 2. Контрпример к Аргументу исключения. Библиотека первых изданий

Представим, что в Марракеше живет старый библиотекарь, который посвятил свою жизнь собиранию древних философских книг. Он собирает первые издания философских трактатов. В его Библиотеке древних философских трактатов хранятся почти все первые издания манускриптов, папирусов, древних сочинений. Библиотекарь знает историю каждой философской книги в мире и может датировать абсолютно любую работу. Поэтому в его библиотеке нет ни одного произведения, которое было бы не первым изданием. К этому вопросу он относится особенно щепетильно. Для полноты коллекции ему не хватает только одного произведения. И вот в его

День рождения на его столе оказываются две совершенно одинаковые книги. Это и есть недостающее произведение — подарок от его друга, коллекционера из Рабата. Книги абсолютно идентичны. Это — полные физические дубликаты. Увидев подарок, старый библиотекарь улыбается, внимательно их рассматривает, а потом одну книгу ставит на библиотечную полку, а другую требует отослать своему знакомому обратно в Рабат. Как объяснить такое странное поведение — ведь библиотекарь собирает *все* первые издания философских книг?

Дело в том, что библиотекарь знает каузальную историю абсолютно каждой книги. И знает, что по заказу его друга была сделана абсолютно точная физическая копия первого издания этой книги. Таким образом, одна книга является первым изданием, а другая — нет. Именно поэтому только один из томов подходит для его коллекции.

В вымышленной ситуации одна книга — в библиотеке древних книг в Марракеше, а другая отправлена в Рабат. При этом книги физически абсолютно идентичны. Как объяснить разную траекторию этих книг? Из описания ситуации разница между книгами понятна: она в том, когда и как были изданы эти экземпляры. Только одна книга оказалась в библиотеке потому, что только она была первым изданием. Это объяснение вполне удовлетворительно. Оно позволяет предсказать развитие аналогичных событий и в будущем. Если вдруг в следующий раз старому библиотекарю придут еще две книги, мы уже не так удивимся его поведению. И при наличии абсолютно полных знаний сможем в другой раз сами предсказать, какая книга окажется на полке, а какая — на почте. Но какого рода это объяснение? Мне представляется, что это объяснение — каузальное. Оно выявляет причину различной траектории книг.

О сущности каузальных связей у философов нет согласия. Существует множество теорий относительно того, что лежит в основе суждений о каузальности. В этой работе, пожалуй, нет возможности обсуждать детали

этих разногласий. Однако существуют общие места, точки согласия, самые общие критерии каузальности, которые обычно принимаются. На них мы и будем опираться. Этими общими критериями каузальной связи являются регулярность, контрфактическая зависимость, временное и пространственное соседство. Итак, проверим соответствие нашего описания этим критериям. Во-первых, мы видим присутствие регулярности. Событие «поступление первого издания философской книги» регулярно сочетается с событием «пополнение библиотеки». В этой регулярности нет и не предвидится исключений. Кто-то может возразить: но библиотекарь, в конце концов, может ошибиться или поменять свои критерии. Тогда регулярность нарушится! Это означает, что регулярность — простое совпадение. Что ж, тогда можно усовершенствовать мысленный эксперимент: вместо человека-библиотекаря поставить машину — робота, который будет обладать всей информацией и отслеживать каузальную траекторию всех книг. Такой робот будет механически отсортировать древние книги от подделок и сохранит требуемую регулярность. Таким образом в мысленном эксперименте будет выполнено первое требование.

Во-вторых, между событием «поступление первого издания философской книги» и событием «пополнение библиотеки» существует контрфактическая зависимость. Если произойдет событие «поступление первого издания философской книги», то произойдет событие «пополнение библиотеки», и если событие «поступление первого издания философской книги» не произойдет, то не произойдет и «пополнение библиотеки». Таким образом, выполняется и второе, важнейшее требование к каузальной связи. К этому можно добавить и другие традиционные критерии: предшествование во времени причины действию и их пространственную смежность. Эти критерии тоже выполнимы. Кажется, можно заключить, что связь между событиями каузальная. Теперь стоит уточнить, между какими именно свойствами имеется эта каузальная связь.



Книги являются физическими объектами. У них есть физические свойства. И они могли бы выступать носителями каузальных закономерностей. Но в предложенном мной случае они не имеют каузальной эффективности. Дело в том, что книги, по условиям эксперимента, являются физическими двойниками, то есть в физическом отношении ничем не отличаются друг от друга. И тем не менее, они оказываются в разных местах: одна — в Марракеше, а другая — в Рабате. Причиной их различных траекторий служит не какое-то их физическое свойство, а каузальная история, свойство «быть первым изданием философского трактата» или, что более точно, реляционное отношение между конкретным экземпляром книги, её местонахождением и воспоминаниями библиотекаря или состояниями блока памяти робота, отношение R. Это реляционное свойство в данной интерпретации каузально действенно. Следовательно, это — контрпример для аргумента Кима. Чтобы убедиться в этом, рассмотрим три наиболее вероятных возражения.

*Возражение 1. Можно было бы показать каузальную эффективность высокоуровневых свойств на более тривиальном примере*

Для того чтобы привести контрпример аргументу Кима, большинство тривиальных случаев не подходит. Так, во множестве стандартных ситуаций физические высокоуровневые свойства реализованы локально в самом объекте. Интерпретируя эти ситуации, можно настаивать на утечке каузальности к низкоуровневым физическим свойствам. В нашем же случае локальные физические свойства объекта недостаточны для объяснения устойчивой закономерности. Несмотря на то, что книги неразличимы по своим физическим параметрам, их появление приводит к различным действиям. Для объяснения этого необходимо соотношение каузальной траектории книги и памяти библиотекаря или робота. Более того, в описанной ситуации могут возникать любые физические события (могут поступать любые книги), при этом закономерность может быть каузально

объяснена только с обращением к высокоуровневому реляционному свойству R.

*Возражение 2. Аргумент спасает каузальную эффективность высокоуровневых свойств, но не спасает ментальную каузальность*

Предположим, что интерпретация описанной выше ситуации верна. Но разве она позволяет нам утверждать, что ментальные свойства каузально эффективны? Свойство «быть древним философским трактатом» — это просто реляционное отношение между экземпляром книги (конкретным физическим объектом), её местоположением и информацией в памяти библиотекаря или блоке памяти робота (также реализованной в физическом носителе). Это свойство R — просто отношение между физическими объектами. В свою очередь, ментальные состояния таковыми не являются. Ментальные состояния — это субъективные качественные состояния. Доказав каузальную эффективность свойств типа R, мы еще не доказали возможность каузальной эффективности ментальных состояний. Так мог бы рассуждать оппонент нашего аргумента.

Но мне эта критика не представляется основательной. Целью нашего эксперимента было показать, что каузально эффективными могут быть любые другие свойства помимо базовых физических свойств. Свойство R не является базовым физическим свойством, хотя оно глобально супервентно на этих свойствах. Поэтому наша цель достигнута. Возможно, что и ментальные свойства являются подобными высокоуровневыми свойствами и что они супервентны, хотя и не редуцируемы к базовым свойствам физического субстрата.

*Возражение 3. Аргумент не спасает каузальную эффективность высокоуровневых свойств*

Состояния библиотекаря (или робота) различны в случаях предъявления ему различных по истории, но идентичных физически книг. Эти различные состояния и являются причиной разных действий, — мог бы

возразить оппонент. С этим возражением я, безусловно, согласен. Но вопрос именно в том, почему библиотекарь приходит к таким разным состояниям, ведь перед ним появляются физически идентичные книги. Для библиотекаря они, конечно, с самого начала не идентичны: он знает, какая из них старая, а какая — нет. Но именно это и показывает, что у книг имеются каузально эффективные высокоуровневые свойства. «Старая» и «новая» не есть свойства фундаментальной физики. При том, что только эти свойства дают единственно доступную возможность объяснить каузальную закономерность.

Оппонент мог бы сказать, что тогда можно упростить условие мысленного эксперимента. Например, представить, что библиотекарь всегда берёт книгу, которая справа, и отправляет обратно книгу, которая слева. В таком случае реализация одинаковых физических событий тоже вела бы к разным действиям. Но при этом вряд ли можно было бы говорить о каузальной эффективности высокоуровневых свойств.

На это можно, в свою очередь, возразить: книги могли бы оказаться по разные стороны от библиотекаря. И тогда последовали бы другие действия. Пример с таким алгоритмом не схватывает интересной «высокоуровневой» закономерности. В то время как наш пример с Библиотекой древних философских трактатов схватывает важную казуальную закономерность, которая построена на высокоуровневых свойствах книг. Какой бы физический объект ни оказался перед юниолиотекарем, с какой бы стороны он ни оказался, библиотекарь поступит с ним в соответствии с его каузальной историей.

### **3. Что не так с Аргументом исключения Дж. Кима.**

Аргумент обобщения доказывает, что если ментальная каузальность и проблематична, по такой же логике проблематична каузальность всех свойств, приписываемых объектам в специальных науках, т.е. всеми науками, кроме фундаментальной физики. Примеры с «Двумя черными ящиками» и «Библиотекой первых изданий» показывают, что высокоуровневые свойства

иногда являются единственным способом объяснить регулярности в мире. Это указывает на их каузальную эффективность. Но для решения проблемы остается выяснить, что же тогда «не так» с аргументом Дж. Кима. Чтобы это показать, удобно представить аргумент в формальном виде.

ПКЗ<sup>1</sup> *Если физическое событие имеет причину в момент  $t$ , оно имеет достаточную физическую причину в момент  $t$ .*

ПКИ: *Если событие  $e$  имеет достаточную причину  $c$  в момент  $t$ , никакое событие в  $t$ , отличающееся от  $c$  не может быть причиной  $e$  (кроме случаев сверхдетерминации).*

(ПОСС) *Не существует систематической сверхдетерминации физических событий.*

Структура рассуждения Дж. Кима такова:

1. Допустим, ментальное событие  $m^0$  является причиной физического события  $p$ ,  $m^0 \rightarrow p$ .
2. Согласно ПКЗ<sup>1</sup>, у события  $p$  есть также достаточная причина  $p^0$ ,  $p^0 \rightarrow p$ . (ПКЗ<sup>1</sup>)
3.  $m^0 \neq p^0$ .
4. Это — не случай сверхдетерминации. (ПОСС)
5. Должна быть одна достаточная причина. (ПКИ)
6. Этой причиной является физическая причина  $p^0$ . (ПКЗ<sup>1</sup>)

Следовательно, у ментального события нет каузальной действительности. Вывод: либо  $m=p$ , либо  $m$  — эпифеноменально.

Думаю, что в описанном выше рассуждении ошибка заключается в применении ПКЗ<sup>1</sup>. ПКЗ<sup>1</sup> допускает широкую и узкую трактовку физических свойств. Если использовать узкую трактовку, то получается, что причинами всех событий должны быть фундаментальные свойства физики. И эти свойства конкурируют за каузальную эффективность с ментальными. Но это

не обязательно должно быть так. Некоторые регулярности существуют именно из-за высокоуровневых свойств. Эти свойства подпадают под широкую трактовку физических свойств. Ментальное как высокоуровневая характеристика физического может быть частным случаем физических свойств в широком смысле. Тогда будет неправильно полагать, что, помимо высокоуровневых свойств, причиной являются еще какие-то свойства. То есть шаг 2 в рассуждении окажется ложным.

Одна из последних работ Дж. Кима называется «Физикализм или что-то около того». Это — наиболее актуальное и полное изложение его метафизической позиции. Как следует из названия, позиция автора — это «почти физикализм». Судя по аргументу, мне, однако, кажется, что позиция автора это не «почти физикализм», а даже «нечто большее, чем физикализм». Это — некий «пуританский физикализм». Такой физикализм разделять мы вовсе не обязаны, тем более в ущерб ментальной каузальности.

## VII. Итоги первой главы

В начале главы было выяснено, что наиболее правдоподобной теорией действия является стандартный каузализм. Ключевым проponentом стандартного каузализма во второй половине XX в. стал Д. Дэвидсон. Стандартный каузализм предполагает наличие ментальной каузальности, то есть возможности того, чтобы ментальные события были причинами физических событий. Однако ментальная каузальность оказывается проблематичной в свете аргументов, предложенных современными метафизиками. Основным препятствием служит Аргумент исключения, в наиболее остром виде сформулированный Дж. Кимом в 1992г. Аргумент построен на общепринятых посылах и, тем не менее, ведет к отрицанию возможности ментальной каузальности.

В ходе работы были рассмотрены несколько вариантов защиты ментальной каузальности. Они были выбраны таким образом, чтобы наиболее репрезентативно представить различные подходы. Первым было

рассмотрено решение проблемы ментальной каузальности в локальном интеракционизме. Был проанализирован ключевой аргумент этой позиции — Аргумент каузальных траекторий. Этот аргумент должен показать ложность тезиса о локальной супервентности и обосновать особую роль ментальных состояний в поведении физических объектов. Однако было показано, что этот аргумент не лишен недостатков и, вполне возможно, опровергается с помощью Аргумента-двойника.

Следующей была рассмотрена позиция американского философа Дж. Сёрла — биологический натурализм. По мнению этого философа, ментальная каузальность вовсе не является проблематичной. Однако, как было показано, его позиция внутренне противоречива. Сёрл смешивает понятия реализации, супервентности и каузальности. Это приводит к новым парадоксам. Из проведенного анализа следует, что решение проблемы ментальной каузальности в теории биологического натурализма не является оптимальным.

В качестве следующей альтернативы рассмотрена теория другого американского аналитика, Д. Деннета, — телеофункционализм. Деннет убежден, что проблема ментальной каузальности решена. Его позиция представлена в виде мысленного эксперимента — ситуации, которая должна иллюстрировать каузальную эффективность высокоуровневых свойств. Однако сама по себе эта иллюстрация недостаточна для того, чтобы служить решением проблемы ментальной каузальности. Отсутствие у Деннета разъяснений не позволяет сделать вывод о том, что именно в аргументе исключения он отрицает. Ввиду этого автор главы представляет собственное трехчастное решение.

Это решение основано на Аргументе обобщения (представленном в формулировке Н. Блока), авторском мысленном эксперименте Библиотека первых изданий и анализе структуры Аргумента исключения. Предложенный подход снимает проблему сочетания ментальной каузальности и физикализма и обосновывает ментальный кауализм. Ментальный кауализм

предполагает, а) что действие — это событие, определенное ментальной каузальностью; б) реализация свободы воли и моральная ответственность невозможны без действий агентов, и в) ментальная каузальность возможна в условиях физикализма. Теперь, когда одно препятствие на пути к свободе воле и моральной ответственности преодолено, можно перейти к следующему — детерминизму. Об этом — вторая глава диссертации.

## ГЛАВА 2. Современные теории свободы воли и детерминизм

### I. Проблема соотношения свободы, детерминизма и индетерминизма

В предыдущей главе было рассмотрено первое препятствие, которое обнаруживают современные аналитики для свободы воли — физикализм. Во второй главе будет рассмотрено второе препятствие — детерминизм. В самом начале главы будет проанализирована связь понятий свободы и ответственности в том виде, как она представляется современным аналитикам. В качестве иллюстрации будут использованы примеры из юридической практики. Автором будет показано, что под моральной ответственностью обычно понимается такое отношение между агентом и его действием (или последствиями этих действий), которое вызывает ответные реакции (reactive attitudes)<sup>81</sup> у других людей, такие как негодование, осуждение, похвала, благодарность и т.д. В исследовании будет продемонстрировано, что моральная ответственность именно в этом смысле служит индикатором наличия свободы воли и используется аналитическими философами, в частности Д. Перебумом, А. Миле, М. Талбертом.

Большинство современных философов считают, что моральная ответственность указывает на то, что свобода воли предоставляет агентам какой-то дополнительный контроль над действиями и тем самым дает основания для ответных реакций. В свою очередь, этот дополнительный контроль обеспечивается соблюдением двух условий: Принципа альтернативных возможностей и Принципа автономии (или «ответственности в конечном счете»). Следующим шагом в тексте главы будут освещены Аргумент последствий в формулировке ван Инвагена и Аргумент манипуляций в формулировке Д. Перебума. Эти аргументы будут

---

<sup>81</sup> В качестве перевода термина «reactive attitudes» я использую русскоязычное словосочетание «ответные реакции». Может показаться, что это тавтология. Однако, мне кажется, что здесь это выражение уместно, так как реакция может быть на любое событие, а ответная реакция может быть только на намеренное действие. Именно в качестве реакции на намеренное действие могут быть оправданы такие реакции, как негодование, осуждение, похвала.



приведены для того, чтобы выявить противоречие между детерминизмом и этими двумя условиями свободой воли.

В качестве решений этого противоречия я представлю две наиболее влиятельные современные инкомпатибилистские позиции: либертарианство Р. Кейна и твердый инкомпатибилизм Д. Перебума. Согласно позиции Кейна свобода и ответственность несовместимы только с детерминизмом. Согласно позиции Перебума свобода и ответственность несовместимы как с детерминизмом, так и с индетерминизмом. В качестве основания для второго утверждения будет представлен Аргумент исчезающего агента Д. Перебума. Этот аргумент является улучшенным вариантом Аргумента случайности Д. Юма.

В процессе исследования будет представлена критика оснований инкомпатибилистской позиции. Будут проанализированы примеры, предложенные Г. Франкфуртом в 1969г. против Принципа альтернативных возможностей, и критика Аргумента последствий Д. Деннетом, а также будет предложена критика Аргумента манипуляций Д. Перебума и собственный ответ на Аргумент исчезающего агента — Аргумент четырехмерного агента. Эти рассуждения станут основанием позиции суперкомпатибилизма — утверждения о том, что свобода и ответственность совместимы как с детерминизмом, так и с индетерминизмом. Суперкомпатибилизм — наиболее распространенная сегодня позиция среди теоретиков свободы воли. Это подтверждает К. Тимпе: «...современные компатибилисты — это сторонники свободы "в любом случае"; иными словами, они считают, что у нас есть свобода воли, если детерминизм истинен, ...но также и если детерминизм ложен»<sup>82</sup>. Таким образом, результатом второй главы будет обоснование этой наиболее распространенной позиции.

---

<sup>82</sup> Timpe K. Free Will. Sourcehood and its Alternatives. London, 2013. С. 175.

#### А. Ответственность как индикатор свободы

Условием свободы является отсутствие внешнего принуждения со стороны других агентов или непреодолимых обстоятельств. Действия, которые агент совершил под угрозой, под физическим давлением или вследствие манипуляции других агентов, не являются свободными. Интуитивно кажется, что понятие свободы предполагает отсутствие принципиальных ограничений, препятствий, сдерживающих факторов со стороны других агентов. Так, свобода слова предполагает отсутствие цензуры. Свобода движения — отсутствие барьеров для движения. Свобода мысли — отсутствие стереотипов, догм и манипулирования в интеллектуальной деятельности. Это — негативная часть определения. Чтобы понятие «ограничения» было осмысленным, нужно определить то, что могло бы быть ограничено. Видимо, свобода воли как минимум предполагает отсутствие принципиальных ограничений реализации желаний агента в поведении. Но все ли действия, реализующие желания самого агента, совершённые без принуждения или ограничения со стороны внешних агентов, свободные? Д. Юм считал, что это именно так. В его понимании свобода заключается в реализации желаний. В «Исследовании о человеческом познании» он писал: «...мы можем подразумевать под свободой только способность действовать или не действовать согласно решениям воли; если мы хотим оставаться в покое, мы можем это сделать, а если хотим двигаться, то можем сделать и это»<sup>83</sup>. Однако большинство современных аналитических философов не согласны с этим утверждением.

Современные теоретики свободы воли считают, что даже когда агент беспрепятственно реализовывает свои желания, свободы воли у него все равно может не быть. Это подтверждает Л. Экстрем: «Большая часть философов, однако, считают, что некто может действовать в соответствии с собственными желаниями (а также переживать ментальные состояния

---

<sup>83</sup> Hume D. An Enquiry Concerning Human Understanding. Oxford, 1995. P. 95.

желаний), при этом не совершая свободные действия»<sup>84</sup>. У этого утверждения несколько оснований. Во-первых, можно сопоставить возможности человека и животных. Животные иногда могут реализовывать свои желания в поведении. Но суждение о свободе воли у животных звучит противоречиво. Второе очевидное основание — различие свободных и несвободных действий у людей. Его можно выявить с помощью лингвистического анализа выражений (1) «он мог сделать это» и (2) «он мог бы сделать это, если бы пожелал», как это делает П. ван Инваген. Философ показывает неэквивалентность этих выражений, то есть он доказывает, что нельзя интерпретировать первое высказывание с помощью второго. Для демонстрации этого ван Инваген приводит контрпример. В «Эссе о свободе воле» он сопоставляет два утверждения:

А) «Смит мог съесть конфету красного цвета»

Б) «Смит мог съесть конфету красного цвета, если бы пожелал»<sup>85</sup>

С точки зрения ван Инвагена, различие между этими утверждениями обнаруживается, если допустить, что Смит патологически боится вида крови, а конфеты кроваво-красного цвета. Из совокупности этих обстоятельств следует, что Смит не смог бы съесть красную конфету, хотя смог бы, если бы пожелал. То есть при истинности Б, А оказывается ложным. Значит, нельзя обходиться классическим анализом свободы как возможности поступать сообразно желаниям. Тем более что научные открытия расширяют спектр известных физических причин желаний и действий. Следующий пример, приведенный М. МакКеной и Д. Перебумом в «Свободе воли: введение в современную дискуссию»<sup>86</sup>, является аналогичным случаем:

---

<sup>84</sup> Ekstrom L. Volition and the Will // A Companion to the Philosophy of Action. New Jersey: Wiley Blackwell, 2010. P. 101.

<sup>85</sup> Van Inwagen P. An Essay on Free Will. Oxford: Clarendon Press, 1983 (1986). P.116.

<sup>86</sup> McKenna M., Pereboom D. Free Will: A Contemporary Introduction. New York, 2016. P. 177.

*Выбор щенка.* Отец хочет подарить дочери щенка. В день рождения он приносит ей на выбор белого и чёрного. Отец решил, что подарит ей того, которого она сама предпочтет. Дочка выбирает белого. Является ли её действие свободным? Описание действия в этой вымышленной ситуации полностью соответствует требованиям рабочего определения свободы действия. У девочки есть желание, и оно беспрепятственно реализуется в действии. Но что, если девочка не могла поступить иначе? Допустим, что в вымышленной ситуации девочка боится собак черного цвета. Девочка не может сформировать желание взять второго щенка. Ее желания и, следовательно, действия ограничены. Действие соответствует желанию, но у нее не может сформироваться другого желания. Тогда это не свободное действие? А если девочка в силу своих психических особенности не могла устоять перед желанием взять первого увиденного ею щенка? Вот она увидела белого и взяла его на руки, но могла увидеть черного, и тогда бы он остался у нее. Выбор белого щенка соответствовал ее желаниям и не был никем навязан. Но был ли это свободный выбор?

Суждения о наличии свободы воли в этом вымышленном сюжете могут показаться спорными. Вдобавок история производит впечатление надуманности. Возможны ли вообще такие специфические психологические особенности? А если невозможны, какую пользу эти примеры несут для прояснения понятия свободы? Поэтому ситуация с выбором щенка только ставит вопрос. Но существует способ прояснить интуиции в отношении свободы и описанного случая. Большинство современных аналитических философов считают, что это можно сделать с помощью понятия «ответственности». Ответственность является наилучшим индикатором наличия свободы. Применение понятия ответственности фиксирует случаи, когда мы считаем, что свобода воли существует.

Когда современные аналитики говорят об ответственности, в частности о моральной ответственности, они «обычно имеют в виду отношение,

которое может быть установлено между людьми и их действиями или между людьми и последствиями их действий»<sup>87</sup>. Ответственность связывает людей и их поступки таким образом, что люди, совершившие поступки, становятся объектом особых ответных реакций со стороны других людей. Этими реакциями в том числе служат осуждение или похвала. Подобные реакции чаще всего предполагают наличие контроля действующего агента над собственным поведением. Вероятней всего, контроль совпадает с наличием свободы воли. Именно поэтому в современной аналитической философии большая часть теорий свободы воли также является и теориями ответственности. «Моральная ответственность предполагает, что у людей есть некоторый контроль над их действиями. Это связано с тем, что действия, за которые агент несет моральную ответственность, должны быть связаны с ним таким образом, чтобы оправдывать в отношении агента моральные оценки и моральные реакции, такие как осуждение и поощрение. Но как мы должны характеризовать контроль, необходимый для моральной ответственности? Многие философы считают, что такого рода контроль предполагает свободу воли»<sup>88</sup>.

О наличии свободы можно судить по наличию ответственности. Поэтому, чтобы определить, в каких случаях наступает ответственность, то есть в каких случаях считается, что человек обладает свободой воли, удобней всего сначала обратиться к судебной практике. В этом отношении судебные прецеденты представляют хороший иллюстративный практический материал. Они показывают, в каких случаях принято считать человека ответственным, а в каких — нет. Анализ применения уголовного права, по мнению другого исследователя свободы воли и нейрофилософа, Патрисии

---

<sup>87</sup> Talbert, M. *Moral Responsibility*. Cambridge: Polity, 2016. P. 5. Российские философы часто дают другое определение понятию ответственности, связанное с моральными ценностями. В частности, в книге «Этика» ответственность определяется как «понятие, которое отражает процесс перехода от моральных ценностей и принципов к возникающей на их основе деятельности». Этика. Под. Ред. А.А. Гусейнова. М, 2016. С. 157. Однако для целей настоящего исследования я буду использовать то определение, которое чаще используется в современной аналитической философии.

<sup>88</sup> Talbert, M. *Moral Responsibility*. Cambridge: Polity, 2016. P.15.

Черчленд, очень полезна для философского исследования, т.к. выводит дискуссию за рамки «простого академического волейбола»<sup>89</sup> в реальный мир.

#### В. Действия без принуждения не обязательно являются свободными

Из мировой юридической практики видно, что суды учитывают внешние ограничения и обстоятельства преступлений для определения вины. Этими обстоятельствами служат не только принуждение, но и обстоятельства формирования личности. Так, биография, воспитание, окружающие обстоятельства влияют на решение и могут смягчать или увеличивать вину. Например, насилие, унижения, психологические травмы, перенесенные в детстве, формируют характер. Характер, сложившийся в таких обстоятельствах, предопределяет антисоциальное поведение. Поэтому ответственность в данном случае может быть смягчена.

Суды учитывают и внутренние факторы, ограничивающие контроль и, соответственно, ответственность. Такими факторами являются физиологические и психологические пороки. Практика учета подобных обстоятельств существует еще со времен Аристотеля<sup>90</sup>. В американском праве показательна с этой точки зрения защита с помощью ссылки на «невменяемость» подсудимого. Кодификация этой защиты была связана с «правилами МакНотена», введенными в 1840-х. Эти правила устанавливали освобождение от ответственности в случае интеллектуальных недостатков, «дефектов разума» преступника. Правила были введены в связи с убийством Дэниелом МакНотеном Эдварда Драммонда, которого МакНотен принял за британского премьер-министра Роберта Пила. В результате суда МакНотен был освобожден от ответственности. Его признали невменяемым, так как он находился в состоянии заблуждения, предполагая, что его преследуют, и не мог препятствовать своим побуждениям совершить убийство, чтобы это преследование прекратить.

---

<sup>89</sup> Churchland P. Touching a Nerve: Our Brains, Our Selves. New York, 2013. P. 194.

<sup>90</sup> По статье R. Sapolsky. The Frontal Cortex and the Criminal Justice System // Philosophical Transactions, 2004. P. 1787-1796.

Позже, в первой трети XX в. суд округа Колумбии дополнил интеллектуальные критерии невменяемости критериями отсутствия достаточного волевого контроля. От ответственности стали освобождаться люди, действующие по «непреодолимому импульсу»<sup>91</sup>. Согласно этому дополнению подсудимый стал признаваться невиновным, если вследствие психического расстройства он не мог справиться с собственным желанием, даже если при этом обвиняемый понимал суть действия и его противоправность, то есть не обладал «дефектами разума». Психическая ограниченность стала определяться не только дефицитом рациональности, но и отсутствием волевого контроля.

Такие принципы невиновности присутствуют в стандарте Дэрхема 1954г. и в Примерном Уголовном Кодексе 1962г. Наиболее шумным процессом с применением защиты подобного рода является дело Д. Хинкли — покушение на президента Р. Рейгана. В ходе спланированного преступления Хинкли произвел шесть выстрелов из пистолета: он ранил президента и трех его сопровождающих. При этом суд признал Хинкли невиновным. В отличие от МакНотена, обвиняемый не испытывал заблуждений в отношении Рейгана и его окружения и не считал себя в опасности. Но адвокатам удалось показать, что подсудимый не мог противостоять стремлению совершить преступление. Будучи одержимым страстью к актрисе Джоди Фостер, Хинкли считал преступление единственным способом привлечь ее внимание.

Впрочем, этот конкретный случай вызывает противоречивые оценки (был ли Хинкли действительно неспособен совладать с собой? не было ли это просто удачной уловкой адвоката?). Вполне возможно, он демонстрирует даже некоторый перекосяк в снисходительности правовой системы. Реакция большинства граждан США, наблюдавших за процессом, была негативной.

---

<sup>91</sup> По УК РФ также юридическим критерием невменяемости являются два ограничения, интеллектуальное и волевое, причём достаточным для признания лица невменяемым является наличие хотя бы одного. Уголовное право России. Части Общая и Особенная / Под ред. А. И. Рарога. М., 2004. С. 118—119.

Поэтому он стал причиной отката судебной практики в сторону более жестких критериев невменяемости<sup>92</sup>. Но вероятно, что откат этот временный. И суды в качестве обстоятельств дела будут принимать доказательства расстройства воли подсудимых, особенно в связи с тем, что в последние несколько десятилетий такие свидетельства становятся более доступными и объективными.

Это следствие прогресса в техническом и концептуальном плане: во-первых, появления технологий неинвазивного сканирования и мониторинга работы мозга (ЭЭГ и ПЭТ), а во-вторых, возможности ассоциировать множество ментальных состояний с нейрофизиологическими. Прогресс предоставил возможность «подглядывать» в субъективный мир агента, давать оценку его когнитивным состояниям и способностям. Стало возможным более объективно судить, мог ли обвиняемый в данном случае сдержаться, поступить иначе. Появился способ изучать непосредственно ключевые обстоятельства преступления — психологические особенности и состояния мозга обвиняемого.

Исследования показали, какие именно части мозга отвечают за контроль желаний у человека. Оказалось, что контролирующие и сдерживающие функции сосредоточены преимущественно в префронтальной коре (ПФК) головного мозга. Это — место расположения своего рода суперэго, центр цензуры и управления импульсами. В ПФК происходит выбор «трудных» сценариев поведения: в случаях, когда нужно пересилить автоматический навык или отсрочить получение вознаграждения, выполнить действие более трудоемким, но более эффективным в долгосрочной перспективе способом. В частности, это подтверждается следующими экспериментами. Испытуемому представляется для запоминания список из 16 предметов. Они могут запомнить их двумя способами: самый простой — это

---

<sup>92</sup> В 1984г. американский конгресс устранил «дефекты волевого контроля» из критериев невменяемости на уровне федеральных судов в рамках Акта по реформе защиты по невменяемости.



попытаться запомнить предметы в представленной последовательности, как есть. Более сложный, требующий усилий, но более эффективный способ — сгруппировать предметы в семантические категории. Эксперимент показывает, что переход ко второму способу всегда требует активации ПФК.

Другой пример показывает роль ПФК в управлении эмоциями. Испытуемым предлагают для просмотра видеоролик с крайне неприятными сценами. В одном случае они должны просто просмотреть этот ролик, а в другом — подавлять естественные отрицательные реакции. Как показывает мониторинг, в первом случае ПФК почти не задействована, а во втором принимает активное участие.

Случаи снижения волевого контроля известны давно, хотя ассоциация этих расстройств с дефектами ПФК и лобных долей в целом до развития средств сканирования активности мозга была неподтвержденной. Самый известный случай подобных нарушений — случай, произошедший более 150 лет назад с американским железнодорожником Финеасом Гейджем. При взрывных работах по прокладке путей Гейдж получил тяжелую травму. Металлический прут диаметром 3 см вошел в череп Гейджа ниже левой глазницы и вышел на границе лобной и теменной костей. К значительному удивлению и радости товарищей, Гейдж быстро реабилитировался и уже через два месяца смог вести активный образ жизни. Но его личность сильно изменилась, так что знакомые даже перестали его узнавать. До происшествия Гейджа знали как человека ответственного, уравновешенного, трудолюбивого, справедливого в отношении к подчиненным. После травмы он стал импульсивным, неспособным сдерживать «животные склонности», лишённым чувства уважения к своим коллегам. Он изменился настолько, что, несмотря на физическую реабилитацию, не смог вернуться на постоянную работу. Вероятней всего, причиной такого радикально изменения характера было нарушение связи между левой лобной долей и остальными частями мозга.

Случай с Гейджем вряд ли просвещает нейрофизиологов. Спустя 150 лет осталось мало свидетельских показаний о ментальных изменениях Гейджа, да и точность сохранившихся сомнительна. Но прецедент сыграл важную роль в концептуальном осмыслении. Этот и подобные случаи показали, насколько нарушения мозга влияют на поведение, на способность контроля. С того времени доказательная база о связи ПФК и контроля эмоциональных импульсов принципиально увеличилась. Подтверждено, что здоровые дети, мозг которых еще находится в состоянии развития (вопреки расхожему мифу, мозг продолжает развиваться даже после подросткового возраста), демонстрируют ограниченные способности контролировать эмоции и импульсивное поведение и выносить моральные оценки. Доказано, что алкоголь угнетает контролирующую функцию ПФК.

Накоплено много прецедентов нарушений части ПФК в результате инсультов. Очень показательное поведение, которое демонстрируют такие пациенты при тестировании: они, с одной стороны, полностью понимают, как рационально следует решать ту или иную задачу, но при этом неспособны справиться с эмоцией, вынуждающей их вести себя нерационально. В ходе одного из нейрофизиологических тестов пациентам с нарушениями в части ПФК предлагают на выбор два варианта: один, который дает мгновенное, но небольшое вознаграждение, и другой — вознаграждение значительно большее, но с отсрочкой. В ходе тестирования пациенты обычно комментируют: «Я знаю, я знаю, как работает этот тест. Я должен выбрать этот более сложный путь потому, что вознаграждение в результате будет более значительным. Поэтому я именно это сейчас и выберу...», — и в тот же момент выбирают более «простой» путь<sup>93</sup>. Все это еще раз показывает, что функция волевого контроля может быть расстроена даже при сохранении когнитивных способностей. И люди с такими особенностями не могут нести ответственность за свои поступки. Это

---

<sup>93</sup> Sapolsky R. The Frontal Cortex and the Criminal Justice System // Philosophical Transactions. Nov., 2004. P. 4134.

доказывает, что действия без внешних ограничений, соответствующие желаниям агента, не всегда являются свободными.

### С. Два принципа свободы воли

Что же тогда за компонент нужно добавить к отсутствию принуждения, чтобы получить свободу, необходимую для ответственности? Выше была прослежена связь идеи свободы с отсутствием внешних и внутренних ограничений. Но это определение можно перевернуть с негативного на позитивное и таким образом сделать его более содержательным. Отсутствие ограничений — это не что иное, как наличие возможностей. Тогда получается, что свобода существует, когда есть возможности, когда доступны альтернативы. Вероятно, и случай с выбором двух щенков, и случаи с психологическими дефектами из юридической практики не подходили под определение свободных действий, т.к. не предполагали альтернатив для агентов. Следовательно, **(1) чтобы агент был свободен в действии и нес моральную ответственность, он должен иметь возможность поступить иначе**<sup>94</sup>. Это утверждение называется «Принципом альтернативных возможностей», и оно фигурирует во множестве современных работ. Согласно этому принципу будущее по крайней мере в некоторые моменты представляет собой разветвляющиеся пути или, используя поэтическое выражение Борхеса, сад расходящихся тропок. Делая выбор, агент актуализирует одну из открытых возможностей, вступает на одну из множества тропок. И этот выбор должен быть ему иногда доступен.

Важно, впрочем, не только доступ к выбору, но и связь агента с этим выбором. Выбор должен принадлежать агенту. Эту идею можно выразить в «Принципе автономии» (или «самоопределения»). Этот принцип представляет еще одно важное условие свободы: **(2) агент является свободным и, соответственно, морально ответственным, только если**

---

<sup>94</sup> Аналогичным образом принцип альтернативных возможностей формулирует Г. Франкфурт в статье «Альтернативные возможности и моральная ответственность». Frankfurt G. Alternate Possibilities and Moral Responsibility // Russel P. and Deery O. (eds.). The Philosophy of Free Will. Oxford, New York: Oxford University Press, 2013. P. 139.

**причиной его действий «в конечном счете» является он сам.** Агент должен обладать контролем над собственными действиями, сам устанавливать выбор, понимая суть альтернатив, не будучи принужден к нему внутренними или внешними обстоятельствами. Свободное действие должно иметь в качестве причины ментальное состояние, и это ментальное состояние должно принадлежать именно агенту. К примеру, свободный выбор во время голосования предполагает не только наличие нескольких кандидатов, но и то, что выбор исходит от самого агента. При этом агент должен сам понимать, что он голосует, за кого и почему. Так, выбор кандидата, сделанный на основании выпавших в лотерее результатов, не является свободным, как и выбор, основанный на брошенной монетке.

**Эти два принципа следует считать общепринятыми условиями свободы воли.** Они были обнаружены философами несколько веков назад, но продолжают сохранять актуальность и для аналитиков конца XX — начала XXI вв. Эти принципы «представляют собой интерпретацию двух различных аспектов свободы, которые должны быть отражены в любой достоверной теории — в частности, принципы самоопределения (или автономии) и доступности альтернативных возможностей. Любая концепция свободного агента должна предполагать возможности и автономию в каком-то смысле...»<sup>95</sup>, — утверждает Г. Уотсон в знаменитой статье «Свободное действие и свобода воли» (1987). С ним согласна и Л. Экстром: «Современным теоретикам свободы воли не мешало бы обратиться к весьма полезной статье Г. Уотсона 1987г. Она придаёт структуру дискуссии... и смысл широкому спектру публикаций о проблеме свободы воли»<sup>96</sup>.

Но, быть может, эти принципы логически связаны, и тогда возможно сократить число требований? Быть может, если агент автономен и сам определяет свое поведение, в то же самое время он имеет альтернативы?

---

<sup>95</sup> Watson G. Free Action and Free Will // Mind, 1987, Vol. 96, No. 382. P. 145.

<sup>96</sup> Ekstrom L. Free Will is Not a Mystery // Kane R. (ed.). The Oxford Handbook of Free Will. 2011, P. 366.

Иначе как можно будет определить наличие или отсутствие контроля? Вот агент поступил так, но по собственной инициативе мог поступить иначе. Значит, есть контроль, и агент — источник действия. А если в любом случае он поступил бы так, а не иначе, значит, от него ничего не зависело? Есть, однако, основания считать, что эти принципы независимы. Такими основаниями служат контрпримеры — ситуации, когда агент совершает самостоятельный выбор, а альтернатив поведения у него нет. Приведу одну такую ситуацию в пример. Джон находится дома и занимается собственным делом. Вдруг из соседней квартиры раздается крик и призыв о помощи. Джон слышит этот призыв, однако решает ничего не предпринимать. Ему кажется, что жизнь соседа не так важна, как то, чем он сейчас занимается. Однако предпринять он в любом случае ничего не может: преступники, напавшие на соседа, предварительно заперли Джона в его квартире снаружи и перерезали все телефонные линии. Таким образом, представляется, что у Джона действительно нет альтернатив, и тем не менее он сам принимает решение бездействовать.

Думаю, этот и подобные примеры могут показать, что удовлетворение второго принципа возможно без удовлетворения первого. Таким образом, показано, что эти принципы логически независимы. Следовательно, сократить критерии свободы воли до одного условия пока не получится.

### 1. Свобода воли и детерминизм

Итак, пока предположительно свобода воли заключается в отсутствии принуждения и требует удовлетворения перечисленных выше двух условий (или, по крайней мере, одного из них). Но могут ли эти условия в принципе быть удовлетворены? Очевидно, что существуют действия, совершаемые под принуждением, действия, детерминированные внешними и внутренними факторами (другими агентами или обстоятельствами), существуют ограничения воли, снимающие ответственность. Но при внимательном анализе, становится не совсем очевидно, что воля вообще может быть

свободной. Основным препятствием для свободы воли и реализации двух обозначенных принципов является детерминизм.

#### А. Каузальный детерминизм

Детерминизмом называют представление о том, что все события происходят по причине и что все причины с необходимостью определяют результат. Точнее, это каузальный детерминизм. Иначе тезис о каузальном детерминизме можно выразить так: факты прошлого в сочетании с законами природы определяют будущее, одно будущее<sup>97</sup>. Эти две формулировки взаимно дополняют друг друга, т.к. законы природы и являются по определению тем самым механизмом, который обеспечивает необходимость перехода от событий прошлого к событиям будущего. Есть и другие варианты детерминизма, например, теологический, генетический, логический детерминизм и др. Согласно теологическому детерминизму все события предопределены замыслом и волей монотеистического божества. Вопросы соотношения теологического детерминизма и свободы воли были заняты средневековые мыслители и схоласты. В генетическом детерминизме утверждается, что поведение всех живых существ определено их генетикой. А тезис логического детерминизма — в том, что все осмысленные предложения языка либо истинные, либо ложные. В том числе предложения о будущем, настоящем и прошлом. Любой из вариантов детерминизма может представлять проблему для свободы воли. Так, если все события предопределены волей божественного существа, человеку не остается пространства для свободы. Аналогичным образом, пространства для свободы нет, если все реакции организма определены генами. Возможность свободы проблематична также, если все предложения, в особенности о будущем, имеют истинное или ложное значение.

Однако эти варианты детерминизма не являются общепринятыми в отличие от каузального или номологического детерминизма. Каузальный

---

<sup>97</sup> По определению ван Инвагена, детерминизм — «это тезис о том, что в любой момент возможно только одно будущее». Van Inwagen P. An Essay on Free Will. Oxford: Clarendon Press, 1983 (1986). P. 3.

детерминизм (КД) до настоящего времени является распространенной метафизической позицией<sup>98</sup>. И он, по всей видимости, представляет основное препятствие для свободной воли. Это можно показать с помощью двух аргументов, каждый из которых оспаривает соответствующий аспект свободы. Из этих рассуждений следует, что свобода воли не сочетается с необходимостью и детерминизмом. И эта несовместимость представляет ядро современной проблемы свободы воли.

#### В. Два аргумента против свободы в условиях детерминизма

Первое рассуждение будем называть «Классическим аргументом», второе — «Аргументом последствий». Оба этих рассуждения играют существенное значение в современной дискуссии о свободе воли. Первый аргумент основывается на представлении о свободном действии как действии, источником которого является агент. Второй аргумент определяет свободное действие как действие, у которого есть альтернативы. Оба аргумента показывают невозможность свободы воли в детерминистическом мире, или несовместимость свободы воли с детерминизмом событий и необходимостью.

Свобода, в соответствии со вторым принципом, предполагает контроль или то, что агент сам является автономной причиной своих действий. Но разве такое возможно? Человек принимает решение выполнить действие. Это его произвольное действие, действие, которое вызвано его желаниями и намерениями. Это действие в качестве причины имеет его текущее ментальное состояние. Но это состояние, в свою очередь, имеет другую причину — ментальное или физическое состояние. Регресс продолжается в прошлое до его детства, младенчества и рождения. Но и здесь цепочка причин не прерывается, ведь каждое событие, в свою очередь, имеет другое событие в качестве причины. Все события в качестве причин имеют события прошлого, в том числе события до появления агента. События прошлого в

---

<sup>98</sup> Kane R. Introduction: The Contours of Contemporary Free-Will Debates. The Oxford Handbook of Free Will. Kane R. (ed.), 2011, P. 6.

сочетании с законами природы, согласно детерминизму, определяют все последующие события, в том числе действия агента. Следовательно, все действия в качестве причин имеют события за пределами существования агента. И никакие действия не имеют в качестве исходной причины агента. Соответственно, воля человека не является свободной? Причиной конкретного действия является агент и его состояние, но всегда есть и более ранняя причина, которая с необходимостью определяет исход и т.д. Какое тогда значение имеет сам агент? Где здесь пространство для свободы?

Не больше перспектив, кажется, и с другим определением. В соответствии с первым принципом, свободное действие — это действие, у которого есть альтернативы. Согласно тезису о детерминизме, все действия определяются событиями прошлого и законами природы. События прошлого в сочетании с законами природы с необходимостью определяют будущее, только одно будущее. Т.е. при данном прошлом дано только одно будущее. Тогда у действий и желаний агентов нет альтернатив. Соответственно, никакие действия и никакие желания не являются свободными. А если нет свободы, не может быть ни ответственности, ни моральных реакций. Другая дорога приводит к тому же парадоксу.

Возможно, что так. Но перед тем как убедиться в этом, следует как минимум проверить еще один ход рассуждений. Допустим, при заданном прошлом будущее определено. Но, может быть, для свободы достаточно будет наличия альтернатив в контрфактической ситуации? Если бы прошлое было другим, при тех же законах природы будущее тоже было бы другим, — разве это не альтернатива?. Вот в момент  $t$  агент хотел произвести действие  $X$  и сделал  $X$ . Но у действия была альтернатива — действие  $Y$ , т.к. если бы в момент  $t$  агент хотел сделать  $Y$ , он бы и сделал  $Y$ . Может, такой контрфактический анализ спасает свободу действия? Подобный способ защиты свободы в детерминистическом мире был предложен американским богословом Дж. Эдвардсом, а затем многократно использовался другими



философами, в том числе Муром. Он выражается в интерпретации ситуации «он мог сделать иначе» как ситуации «если бы он хотел сделать иначе, тогда бы он мог сделать иначе». И эта вторая ситуация не противоречит необходимости причин и законов природы.

Однако это, вероятней всего, не выход, так как подобная интерпретация ситуации «он мог сделать иначе» неверна. Другая ситуация действительно могла бы привести к другим действиям. Но другой ситуации быть, кажется, не могло. Сама ситуация в  $t$  была предопределена ситуацией в  $t-1$  и законами природы, и так — до появления данного агента и вообще до появления всех агентов. Наш агент просто не мог иметь другие желания и убеждения и, следовательно, в момент  $t$  не мог поступить иначе. Девочка, которой отец предложил выбрать щенка, выбрала белого. Если бы она захотела, она выбрала бы чёрного. Но в описанном случае она не могла захотеть выбрать чёрного, и потому, похоже, выбрать черного она не могла. Значение выражения «он мог сделать иначе», скорее всего, подразумевает не только, что «если бы он хотел сделать иначе, он мог бы сделать иначе», но и что «он мог захотеть иного»<sup>99</sup>. А это плохо совмещается с детерминистическим миром.

#### С. Каузальный детерминизм и квантовая физика

Впрочем, может, проблема решается? Может, она существует только в философской традиции и именно потому, что метафизические установки не прошли должной ревизии? В самом деле, если свобода воли, необходимое условие ответственности, не может сочетаться с детерминизмом, возможно, нет не свободы воли, а детерминизма? Есть некоторые основания так полагать. В то время как философские рассуждения, построенные на эмпирических данных специальных наук (анатомии, нейрофизиологии, психологии и т.п.), склоняют нас к убеждению в детерминизме, современная

---

<sup>99</sup> Подобный аргумент приводит Chisholm R. Human Freedom and The Self. Brown, 1964. P.6.

физика, самая фундаментальная из наук, кажется, больше тяготеет к индетерминизму.

Детерминистическое представление о мире стало преобладать в результате успехов ньютоновской механики. Особенно эти успехи были заметны в астрономии. Ньютоновская теория позволила впервые с высокой точностью предсказывать передвижение небесных тел. Успехи в предсказаниях событий, столь отдаленных в пространстве и времени, вселили уверенность, что вскоре можно будет предвидеть любые события и даже состояния всей вселенной в сколь угодно отдаленном будущем. Мир стал представляться прогнозируемым механизмом, а наука — его новым оракулом. Казалось, что ученым стоит только определить, измерить исходное положение вещей и выявить постоянные законы, управляющие изменениями. И тогда будущее станет открытой книгой. Будущее галактик и каждого атома можно будет просчитать, увидеть со стопроцентной определенностью. Если, конечно, позволят вычислительные мощности.

Тем не менее, надеждам не суждено было воплотиться. И препятствием стали отнюдь не вычислительные мощности. Физику XX в. ждал концептуальный переворот — открытие квантового мира. Эксперименты, направленные на изучение поведения элементарных частиц, подорвали оптимизм в отношении однозначных предсказаний и показали, что прогнозирование событий гораздо более проблематично. Согласно доминирующей, *Копенгагенской интерпретации*, эти проблемы оказались не эпистемологического, а онтологического характера. Ограничения измерений и предсказаний носят не условный, а абсолютный характер. То есть принципиально невозможно установить одновременно точную позицию и импульс (количество движения) элементарной частицы (*принцип неопределённости Гейзенберга*). Следовательно, невозможно однозначно предсказать ее поведение невозможно определить и траекторию ее движения

Классическая иллюстрация индетерминизма в квантовых явлениях — эксперимент с двумя щелями. Источник света направляется на светочувствительный экран. Перед экраном установлен щит с двумя вертикальными щелями так, что свет может попадать на светочувствительный экран, только пройдя через эти щели. Кванты света, фотоны, должны проходить через щели и достигать экрана в виде дискретных точечных порций энергии. Для начала закроем одну щель и включим источник света. На экране мы увидим след в виде яркой полосы прямо напротив щели. Результат вполне объясним: часть излучаемых фотонов попадает в щит и не достигает экрана. Фотоны, следующие по траектории через щель, оставляют на экране след в виде проекции щели. Теперь откроем вторую щель. Каков ожидаемый результат?

Можно предположить, что в результате бомбардировки экрана фотонами через щит с двумя щелями получатся проекции двух щелей, две яркие полосы на экране. Но при проверке выходит не так. На экране появляется так называемая *интерференционная картина*, т.е. регулярное чередование областей повышенной и пониженной яркости. На экране остаются не две, а множество полос. Но удивительному факту тоже находится объяснение. Если представить, что через щели проходят не корпускулы, частицы материи, а волны, эффект ожидаемый. Ведь волны при наложении одна на другую в зависимости от фазы либо гасят, либо усиливают друг друга. Таким образом, там, где гребни волн совпадают, на экране возникает яркая полоса, а где гребень приходится на ложбину — засветки совсем нет. Такая интерпретация предполагает, что свет проходит одновременно через две щели (образуются как бы две волны). Эта интерпретация заставляет склоняться от корпускулярной природы света к волновой. Но в целом объяснение пока звучит вполне логично. Правда, эта интерпретация успешна только отчасти.

Еще немного усложним эксперимент. Начнем бомбардировать экран не потоком, а поочередно, по одному фотону. В случае поочередной бомбардировки вроде бы нечему создавать интерференцию. Т.е. мы ожидаем две параллельные засвеченные линии. Но нас ждет еще более удивительный результат. Следы на экране действительно выглядят как результат воздействия частиц. Но и в этом случае в эксперименте получается интерференционная картина, т.е. множество световых полос, сочетающихся с абсолютно незасвеченными зонами. А это как объяснить? Что, один излученный фотон покидает источник как частица, становится волной, проходит через две щели, интерферирует сам с собой, и попадает в экран вновь как частица? Попробуем это экспериментально проверить. Рядом с одной из щелей установим детектор. Если фотон света проходит через щель, мы получим об этом данные. Итак, начинаем поочередную бомбардировку.

Детектор периодически срабатывает, регистрируя прохождения фотонов через щель. Но на экране вдруг вместо интерференционной картины получаются две полосы. Как будто волновое поведение света вдруг прекратилось под воздействием наблюдения. Фотон обнаружил вмешательство и изменил свое поведение? Но какое отношение к поведению фотона может иметь наблюдатель? Чем объяснить такое странное изменение? И что все-таки такое фотон: частица или волна? Все возможные интерпретации, способные объяснить эти явления, контринтуитивны. Но наиболее распространена такая. Согласно ей, поведение фотона вне наблюдения может быть описано как распределение вероятностей событий. С течением времени эти вероятности эволюционируют. И только в момент регистрации наблюдателем одна из них актуализируется. В квантовой физике этот эффект называется редукцией волновой функции. Без наблюдения не только невозможно определить точное положение частицы в определенной момент. Этого положения вообще нет. Только наблюдение и измерение приводят к его фиксации. Возможно предсказать, какой рисунок получится в

итоге на экране, но нельзя с определенностью сказать, где окажется каждая конкретная частица. И это не результат недостаточности знаний: так, видимо, устроен мир.

Странные явления в микроскопическом мире, по всей видимости, управляются вероятностными законами. Будущее элементарных частиц лишено определенности. Тогда что это означает? Что детерминизм ложен? Что найдено пространство для свободы, необходимое для моральной ответственности? И все же оптимизм здесь преждевременный. Только лишь ссылка на квантовую механику, скорее всего, не даст решения проблемы свободы воли. По крайней мере, есть несколько поводов так считать. Во-первых, в отношении самой теории нет консенсуса физиков. Существует несколько интерпретаций, которые в равной мере согласуются с экспериментальными данными. Действительно, самая распространенная, *копенгагенская интерпретация* является индетерминистической. Но есть и другие. Например, эвереттовская интерпретация (интерпретация многих миров) или теория де Бройля-Бома (интерпретация со скрытыми параметрами), напротив, являются детерминистическими. Во-вторых, даже если фундаментальная физика и показывает наличие индетерминизма в мире, специальные науки, напротив, все больше и больше подтверждают детерминизм. Биология, генетика, нейрофизиология, психология выявляют все больше регулярностей, которые определяют поведение живых организмов. Эти науки показывают, что реакции и свойства, кажущиеся спонтанными и первоначально необъяснимыми, имеют в качестве своих причин наследственные, биохимические, эволюционные основания. В таких исследованиях работает допущение, что поведение живых организмов определено, законосообразно и, следовательно, может быть алгоритмизировано. Т.е. в какой-то момент можно будет создать механизм, программу, способную симулировать все аспекты когнитивной деятельности и поведение. И в отношении нее можно будет делать точные предсказания.

Как, возникает вопрос, это сочетается с физикой? Что, не только философы, но и ученые не следят за трендами в фундаментальной науке? Или просто не существует единого поля научного знания?

Допустим, описываемые события в квантовом мире не детерминированы, как события в мире макрообъектов. В квантовом мире действуют статистические законы, а в мире макрообъектов — законы динамические. Означает ли это, что эти миры несовместимы друг с другом? Два мира можно согласовать, если одни описания касаются единичных событий, а другие — статистических. . Даже если в отношении единичных событий некоторого типа невозможны достоверные предсказания, они могут быть возможны по отношению к множеству таких событий. Скажем, если один раз бросить монетку, то нельзя с точностью определить, выпадет ли орёл или решка. Но если монетку бросать много раз, можно будет достоверно предсказать, что в половине случаев будет выпадать орёл, а в половине — решка. Поскольку само существование макрообъектов — это в каком-то смысле статистическое событие для микрообъектов, вывод кажется последовательным. Обычно считается, что квантовые неопределенности должны сглаживаться на макроуровне. И не должны вносить неопределенность и индетерминизм в поведение человека. Это — третье основание. Но и это — не последний повод считать индетерминизм в квантовой физике нерелевантным для решения проблемы свободы воли. Есть еще одно, уже четвертое основание считать, что квантовый индетерминизм не решает философскую проблему свободы воли.

Предположим даже, что квантовые эффекты могли бы усиливаться до событий в макром мире. И неопределенность эволюции состояний микрочастиц распространялась бы на изменения в макрообъектах. Как эта неопределенность соотносилась бы со свободой действия, необходимой для моральной ответственности? С одной стороны, у действий появились бы альтернативы. События в прошлом в совокупности с законами природы не

определяли бы будущее однозначно. Т.е. одно из условий свободы действий было бы удовлетворено — Принцип альтернативных возможностей. Но, с другой стороны, второе, обозначенное выше условие, вряд ли бы выполнялось. Ведь причиной, источником действия, кажется, не был бы сам агент. Реализация той или иной альтернативы в квантовом мире не зависит от воли агентов. Или... все-таки субъект играет роль?

В описанном выше эксперименте было показано, что наблюдения влияют на редукцию волновой функции. Может, и здесь агент имеет значение? Положительный ответ возможен, только если под агентом подразумевать нематериальную субстанцию. Но если агент — это всего лишь особым образом организованный макрообъект, состоящий, в свою очередь, из элементарных частиц, места для свободы воли все-таки не остается. Квантовая неопределенность — это совсем не та свобода, на которую стоит рассчитывать агенту. Фактором, определяющим результат, здесь является случай. Действия, основанные на квантовых эффектах, с точки зрения агента представляют собой непредсказуемые случайности, как конвульсии или судороги. Они неконтролируемы и не являются тем, что требуется для моральной ответственности.

## 2. Свобода воли и индетерминизм. Аргумент случайности

Оказывается, не только каузальный детерминизм представляет собой проблему для свободы воли и моральной ответственности, но и индетерминизм. Ведь индетерминизм, похоже, ведет к потере контроля. Это отражает следующий принцип: («Принцип случайности») если действия являются индетерминированными, то они являются результатом случайности. А случайные действия не могут быть контролируемыми, свободными, и за них агент не может нести ответственность. На этом принципе основаны аргументы, встречающиеся в философской традиции с давних времен. Они называются «Аргументами случайности». Классической формулировкой Аргумента случайности считается определение Юма.

Шотландский философ так выражает эту идею: «Поступки по самой своей природе суть нечто временное и преходящее; если их не вызывает какая-либо причина, коренящаяся в характере и склонностях совершившего их лица, они не могут ни способствовать его славе, будучи хороши, ни бесчестить его, будучи дурны. Поступки сами по себе могут заслуживать порицания, противоречить всем правилам нравственности и религии, но человек не ответствен за них: поскольку они не были вызваны каким-либо постоянным и непреходящим свойством характера и не оставляют такого же рода последствий, человек не может подвергаться за них наказанию или мщению. Таким образом, согласно принципу, отрицающему необходимость, а следовательно, и причинность, человек, совершив самое ужасное преступление, остается столь же чистым и незапятнанным, как и в момент своего рождения...»<sup>100</sup>. Итак, свобода воли предполагает устранение необходимости действия. По мнению Юма, необходимость есть неотъемлемая характеристика причинности. Устраняя необходимость, свобода устраняет и причинность, а вместе с ней — и связь с характером человека, делая действие случайным. Соответственно, человек не может и не должен нести ответственность за действия, определенные свободой воли.

В современной аналитической философии также используются аргументы, основывающиеся на Принципе случайности. В частности, они встречаются в работах И. Хайджи<sup>101</sup>, А. Миле<sup>102</sup>, П. ван Инвагена<sup>103</sup>. Различные формулировки отличаются в том, как они определяют случайность. Некоторые авторы считают, что «случайность» — это отсутствие возможности объяснить событие. Из этого они делают вывод, что индетерминированные события являются случайными потому, что у них отсутствует соответствующее объяснение. Другие философы считают, что

---

<sup>100</sup> Юм Д. Исследование о человеческом познании. Москва: Канон+РООИ «Реабилитация», 2009. С. 99.

<sup>101</sup> Haji I. Control Conundrums: Modest Libertarianism, Responsibility and Explanation // Pacific Philosophical Quarterly 2001. № 88. Vol. 82. Issue 2, P. 178-200.

<sup>102</sup> Mele A. Free Will and Luck. New York: Oxford University Press, 2006.

<sup>103</sup> Van Inwagen P. Free Will Remains a Mystery // Philosophical perspectives. 2000. №14. P. 1-19.



понятие «случайность» не требует редукции и само по себе очевидно. Эти авторы оставляют понятие «случайность» без расшифровки, а утверждение, что индетерминированные действия являются случайными, принимают как самоочевидное.

Приведу один из вариантов современного Аргумента случайности. Этот аргумент использует понятие возможных миров для интерпретации модальности. Согласно этой интерпретации, если некоторое событие возможно, оно «существует в ближайшем возможном мире». А если событие необходимо, то оно «существует во всех возможных мирах».

1. Свобода воли заключается в отсутствии детерминации предшествующими событиями последующих событий.
2. В реальном мире  $W$  человек, обладающий свободой воли, в момент  $t$  совершает поступок  $A$ .
3. В ближайшем возможном мире  $W_1$  (полностью идентичном  $W$  до момента  $t$ ) этот же человек, обладающий свободой воли, в момент  $t$  совершает поступок  $B$ .
4. Не существует причин, которые с необходимостью предопределяли бы действие  $A$  или  $B$ .
5. Ничто из тех способностей агента, его ментальных состояний, характеристик, диспозиций или мотивов, что предшествовали  $t$ , не является причиной действий  $A$  и  $B$  и не объясняет выбора действий.
6. Следовательно, осуществление  $A$  или  $B$  — дело случая. И человек не отвечает за то, что он в момент  $t$  сделал  $A$  или  $B$ , поскольку он не может нести ответственности за случайные события.

Этот аргумент можно проиллюстрировать гипотетической ситуацией. Представим, что Джон размышляет, где ему провести отпуск: на своей даче или в экспедиции на остров Тристан-да-Кунья. После значительных размышлений, во время которых он взвешивает все за и против, он в итоге выбирает поездку на собственную дачу. Из индетерминизма следует, что

некоторые события происходят без достаточных причин. Т.е. одни и те же события могут развиваться по-разному. Это удачно иллюстрирует метафора разветвляющихся тропинок. Соответственно, если выбор Джона был индетерминированным, после всех тех же самых размышлений он мог бы выбрать экспедицию. Одни и те же размышления могли привести его к совершенно противоположным действиям, на разные тропинки. Следовательно, Джон не отвечает за свой поступок. И никакие ссылки на размышления Джона не могут объяснить его выбор. Отсутствие детерминизма, исходя из этой ситуации, приводит к отсутствию ответственности.

Эта гипотетическая ситуация более опасна, чем Классический аргумент. Опасность в том, что в ней рассматриваются не просто поступки, а поступки намеренные. В этом мысленном эксперименте Джон всерьез размышляет над выбором, имеет рациональные основания в пользу альтернатив. И при этом он может решить поступить так или иначе. Т.е. за гипотетической ситуацией стоит аргумент, где посылки 2 и 3 оснащены дополнительными обстоятельствами.

2. В реальном мире  $W$  человек, обладающий свободой воли, в момент  $t$  *намеренно и по собственному желанию* совершает поступок  $A$ .
3. В ближайшем возможном мире  $W_1$  (полностью идентичном  $W$  до момента  $t$ ) этот же человек, обладающий свободой воли, в момент  $t$  *намеренно и по собственному желанию* совершает поступок  $B$ .

Эта модернизация делает аргумент более устойчивым для критики. Однако не будем забегать вперед. Пока рано говорить о специфике и преимуществах модернизированного аргумента. Даже классические, юмовские возражения против сочетаемости индетерминизма и ответственности представляют проблему.

### 3. Дилемма свободы воли: сочетание с детерминизмом и индетерминизмом

Из всего этого следует, что особенности законов квантовой физики, вероятно, не создают прочного фундамента для свободы и не спасают возможность моральной ответственности от детерминизма. Впрочем, рассуждения, представленные выше, также приводят к еще одному поразительному наблюдению: моральная ответственность, кажется, несовместима не только с детерминизмом, но и с индетерминизмом. Индетерминизм, похоже, превращает действие в случайное событие, трансформирует ментальное в эпифеноменальное. Индетерминизм предполагает, что действия не следуют из причин. Но действия, происходящие без достаточной причины или без причины вовсе — это события, обусловленные случаем. Если действия происходят случайно, то агент не является их источником. И потому он не может нести какую-либо ответственность. Ведь, как было отмечено выше, причинность и контроль — родственные понятия. Итак, обнаруживается дилемма: моральная ответственность предполагает свободное действие. Свободная воля (1) подразумевает возможность реальных альтернатив действий. А также, что (2) агент является исходной причиной действий. Согласно детерминизму агент, видимо, не может быть причиной действий в конечном счете, и у действий не может быть реальных альтернатив. Детерминизм, похоже, не совместим с моральной ответственностью. Но индетерминизм также не совместим с моральной ответственностью, т.к. при отсутствии каузальной необходимости агент перестает иметь связь со своим поведением. Он перестает быть исходной причиной поведения. Исчезает контроль как основание ответственности за поступки. Возможно, даже исчезает само действие вместе с ментальной каузальностью. Свобода трансформирует действие в случайное событие. Это — порочный круг. С одной стороны — необходимость, с другой — случайность. И то, и другое является препятствием для возможности свободы и ответственности. Как резюмирует известный российский этик А. Разин, «В истории философии в понимании свободы

издревле было выявлено определенное противоречие. Оно связано с вопросом об обусловленности человеческой воли. Если воля совершенно ничем не обусловлена, ...свобода в таком случае просто случайный выбор наугад. Но если воля сама по себе чем-то обусловлена, она перестает быть свободной»<sup>104</sup>. Если бы только агент мог быть *causa sui*, причиной самого себя, а его ментальные состояния — причиной действий, тогда моральная ответственность могла бы существовать. Но это не совместимо с современным представлением о каузальности.

#### 4. Альтернативные позиции по вопросу свободы воли в конце XX — начале XXI вв.

Впрочем, философы не представляют проблему нерешаемой. В современной философии представлены практически все альтернативы решений. Есть множество крайне уточненных вариантов распутывания этого концептуального узла. Современные варианты решений проистекают из философских позиций, оформившихся еще в Новое время. Но было бы совершенно неправильно говорить об отсутствии прогресса в этой области. Особенно с середины XX в. позиции по проблеме свободы воли были значительно усовершенствованы. Прогресс был связан в первую очередь с работами П. Стросона, Г. Франкфурта и П. ван Инвагена. И хотя консенсус так и не сложился, современные теории представляют явный шаг вперед в сравнении со своими предшественниками.

Демаркационной линией в философских позициях служит отношение к детерминизму. Часть философов считает мир детерминированным, а часть отстаивают индетерминизм. Вторая черта, в соответствии с которой разделяются позиции — отношение между детерминизмом и свободой воли. Компатибилистами (сторонниками теории совместимости) называют тех, кто считает свободу воли совместимой с детерминизмом; инкомпатибилистами — тех, кто убежден в их несовместимости. Компатибилизм можно также

---

<sup>104</sup> Разин А.В. Свобода и необходимость // Вестник российского университета дружбы народов. 2014. Серия: Философия, №2. 2015. СтрС. 9.

разделять на классический и современный. Классическим компатибилизмом называется позиция, основанная на определении свободного действия как действия, сообразного желанию. Т.е. классические компатибилисты считают, что для ответственности достаточно ментальной каузальности. Классический компатибилизм, как было показано, уязвим для контрпримеров, подобных тем, что приводит ван Инваген, Перебум и Миле. Не только препятствия желаниям, но и сами непреодолимые желания могут представлять ограничения для свободы. Поэтому современные компатибилисты (Дж. Фишер, Д. Деннет, Г. Франкфурт) повышают требования к свободному действию. Среди компатибилистов можно выделить также суперкомпатибилистов. Этот новый термин, который вводит в обращение М. Варгас<sup>105</sup>, обозначает очень распространенную позицию. Согласно этой позиции, свобода и ответственность совместимы как с детерминизмом, так и с индетерминизмом. Сторонники суперкомпатибилизма считают проблему Случайности в случае индетерминированных действий решаемой.

Либертарианство можно разделить на дуалистическое и натуралистическое или физикалистское. Дуалистическое либертарианство допускает отсутствие каузальной замкнутости физического мира и возможность воздействия со стороны нефизических сущностей. Такие теории менее всего распространены в настоящее время. Зато есть варианты натуралистического либертарианства, вдохновленного вероятностными законами квантовой физики. Выше я попытался показать, что на первый взгляд квантовая физика не дает «пространства для свободы агента». Но есть попытки построить теории свободы воли на базе вероятностных законов и обойти Аргумент случайности. Одной из наиболее успешных и проработанных в этом отношении является теория Роберта Кейна<sup>106</sup>.

---

<sup>105</sup> Vargas M. Why the Luck Problem Isn't // Philosophical issues. 2012. №22. P. 491-506.

<sup>106</sup> Среди современных аналитических философов либертарианскую позицию также занимают К. Франклин и М. Балагер.

## II. Решение проблемы свободы воли в либертарианстве Р. Кейна

### 1. Этапы либертарианской кампании

Как считает Р. Кейн, перед сторонником либертарианства стоят две основные задачи. Сначала ему нужно показать, что свобода, необходимая для моральной ответственности, несовместима с детерминизмом. (Эта задача, впрочем, общая как для либертарианца, так и для сторонника жесткого детерминизма). А потом философ-либертарианец должен продемонстрировать, как свобода и моральная ответственность сочетаются с индетерминизмом. Первая задача носит критический характер. Либертарианцу нужно показать, что существует как минимум один вид свободы (и, соответственно, один вид ответственности), который не может быть реализован в условиях детерминизма. И что именно этот вид свободы важен. Вторая задача — это развитие непротиворечивого представления об этой необходимой свободе и ответственности в условиях индетерминизма. Как было показано выше, простое отрицание необходимости не создаёт условий для реализации свободы. Чтобы либертарианская концепция была конструктивной, она должна демонстрировать, как свободные действия могут направляться и контролироваться агентом. Защитник либертарианской позиции должен пройти между Сциллой необходимости и Харибдой случайности.

Кейн сравнивает два этапа в развитии либертарианской теории с восхождением и спуском с горы. Недостаточно только подняться на вершину, нужно суметь с нее спуститься. Если решить только одну проблему, штурмом взять вершину инкомпатибилизма, но задержаться там наверху — понятие свободы останется чем-то мистическим, непроясненным. Разряженный воздух способствует иллюзиям и галлюцинациям. В истории некоторые философы уже становились жертвами этой «панической метафизики». Многих философов, как пишет Кейн, посещали «видения ноуменальных трансэмпирических личностей и неслучающихся причин»

(nonoccurrent causes)<sup>107</sup>. Одним из вариантов такой панической метафизики является апелляция к субстанциональной каузальности. Это — крайне экстравагантный вид либертарианства. Он приводит к отрицанию уже известного нам Принципа каузальной замкнутости физического мира, а это крайне нежелательно. Сам Кейн стремится избежать таких экстравагантных решений. Он рассчитывает не только подняться, но и успешно спуститься с горы, не став жертвой опасных иллюзий. И в этом походе у него есть проводники — это обыденные интуиции и современные физические теории.

Кейн считает, что бытовые интуиции об ответственности и свободе воли ближе всего к либертарианской позиции. Поэтому большинство людей, не задумывающихся всерьез над философскими проблемами, — либертарианцы. Размышляя о будущем, люди представляют альтернативы реальными. Совершая выбор, они верят, что могут решить и так, и иначе, т.е. представляют будущее как «сад расходящихся тропок». Но в детерминированном мире будущее напоминает, скорее, многополосное шоссе: множество непересекающихся между собой прямых линий. То есть интуитивные представления о свободе несовместимы с детерминизмом. Детерминизм также несовместим и с интуитивными представлениями об ответственности.

Интуитивно кажется, что для ответственности человек должен сам быть причиной поступков и создателем собственного характера. Как только мы узнаем все причины, обуславливающие поведение человека, мы снимаем с него ответственность. То есть обыденная точка зрения предполагает, что цепочка причинности должна где-то разрываться, что в детерминации существуют лакуны и агент способен как бы левитировать в каузальном мире. Поэтому Кейн может опираться на интуицию и бытовые представления. Но это не единственная его опора.

---

<sup>107</sup> Kane R. The Significance of Free Will. Oxford. N.Y.: Oxford University Press, 1996. P. 14.

Вторым важным союзником для Кейна является современная фундаментальная наука. Выше я уже отмечал, что квантовый мир, по крайней мере, в одной из теорий описания, не является детерминированным. Т.е. события в нем происходят только с определенной вероятностью, но не могут быть предсказаны наверняка. Анализ, однако, показал, что одни только положения квантовой физики не позволяют считать макромир индетерминистичным. К тому же встает вопрос о том, какое отношение агент имеет к квантовой неопределенности. Нельзя ведь отвечать за события, которые совершаются по случайности. И тем не менее, фундаментальная наука дает концептуальные инструменты для современного либертарианца. В первую очередь она позволяет говорить о вероятностной каузальности. А вероятностная каузальность совсем не то же, что простая случайность. Стороннику либертарианской позиции остается продумать связь между вероятностной каузальностью и агентом. И тогда решение проблемы совместимости свободы и индетерминизма может быть обнаружено.

## **2. Свобода воли и детерминизм несовместимы**

Полагаться на одни только интуиции или феноменальный опыт в вопросе о свободе воли ненадежно. Как и в случае с ментальной каузальностью, простой демонстрацией или ссылкой на очевидность здесь не обойдешься. Недостаточно просто поднять руку и сказать: «вот я поднял руку, а мог бы ее не поднимать. Значит, свободная воля существует». Несмотря на привлекательность поэтической метафоры сада расходящихся тропок по сравнению с индустриальным вариантом параллельных шоссе, предпочтение одного или другого варианта для описания процесса выбора не очевидно. Поэтому Кейн опирается на аргументы. Два основных пути аргументации в пользу инкопатибилизма были обозначены выше: первый — через Принцип альтернативных возможностей, второй — через Принцип автономности или «Принцип ответственности в конечном счете» (так называет его Кейн). Анализ свободы Кейн начинает с Принципа альтернативных возможностей.



## А. Альтернативные возможности или возможность поступить иначе

### а) Принцип альтернативных возможностей у Р. Кейна

«Мы начинаем с условия альтернативных возможностей (АВ) и перейдем к принципу ОКС («ответственности в конечном счете» — Д.В.) позже. АВ намного более распространенное из двух условия и больше обсуждается в связи с вопросом о совместимости. В отличие от ОКС, существует огромное количество современной литературы об отношении альтернативных возможностей или "способности поступить иначе" с детерминизмом...»<sup>108</sup>. Общий аргумент, построенный на Принципе альтернативных возможностей, из которого следует несовместимость детерминизма со свободой воли, Кейн формулирует так:

1. Если агент обладает свободой воли, значит, у него должны быть альтернативы действий.
  2. Если детерминизм верен, никто не может действовать иначе, чем он действует.
- В. Поэтому, если детерминизм верен, никто не действует в соответствии со своей свободной волей.

Однако философ признает, что это рассуждение в последнее время стало выглядеть спорным. Причем сомнения теперь вызывает каждая из двух посылок. Атака на первое утверждение началась в 1969г. в связи с публикацией ключевой статьи Г. Франкфурта «Альтернативные возможности и моральная ответственность»<sup>109</sup> и была подхвачена Д. Деннетом в монографии «Пространство для свободы: типы свободы воли, которые стоит желать»<sup>110</sup>. Франкфурт и Деннет поставили под сомнение то, что раньше вызывало общий консенсус — необходимость альтернативных возможностей как условия свободы. Они представили примеры реальных и гипотетических, но допустимых ситуаций, в которых альтернативные возможности

<sup>108</sup> Kane R. The Significance of Free Will. Oxford, N.Y.: Oxford University Press, 1996. P.37.

<sup>109</sup> Frankfurt H. Alternate Possibilities and Moral Responsibility // Journal of Philosophy. 1969. Vol. 66. No. 23. P. 829-839.

<sup>110</sup> Dennett D. Elbow Room: The Varieties of Free Will Worth Wanting. Cambridge, 1984.

отсутствуют, но при этом кажется, что агент поступает по собственной воле и несет ответственность за свои поступки. Наибольший резонанс получили примеры Франкфурта.

*б) Фальсификация Принципа альтернативных возможностей с помощью примеров Франкфурта и Деннета*

*Пример 1.* В первом примере Франкфурт предлагает вообразить ситуацию, в которой Джонс совершает поступок «А». По замыслу автора, Джонс самостоятельно принимает решение совершить «А». Однако существует и другая причина в пользу поступка «А» — внешняя угроза. После того, как Джонс принял решение делать «А», ему угрожают, что, если он не совершит «А», с ним, его семьей и близкими жестоко расправятся<sup>111</sup>. Фактически перед Джонсом нет альтернатив — он вынужден делать «А». При этом интуитивно кажется, что Джонс несет моральную ответственность за поступок «А». Ведь, несмотря на наличие угрозы, его действующим мотивом было собственное желание. И даже если бы угрозы со стороны не существовало, он все равно сделал бы «А». Франкфурт пишет: «Кажется очевидным, что в этих обстоятельствах тот факт, что Джонсу угрожали, никаким образом, не снимает с него моральной ответственности, которую он нес бы за свое действие»<sup>112</sup>. Можно, однако, возразить, что у Джонса все-таки была альтернатива, ведь так или иначе он мог не поддаться угрозам и сделать не «А», а «Б». Что ж, если в этом случае, наличие реальной альтернативы все-таки есть, можно далее усложнить ситуацию так, чтобы альтернатива полностью испарилась. Так сконструирован второй пример Франкфурта.

*Пример 2.* Блэк хочет, чтобы Джонс выполнил действие «А». Блэк имеет возможность заставить Джонса выполнить «А», но он не хочет вмешиваться раньше времени. Поэтому он ждет, что решит делать сам Джонс. Если Джонс сам выберет «А», Блэк не станет ничего предпринимать.

---

<sup>111</sup> Frankfurt H. The Importance of What We Care About: Philosophical Essays. Cambridge: Cambridge University Press, 1988. P. 2.

<sup>112</sup> Frankfurt H. The Importance of What We Care About: Philosophical Essays. Cambridge: Cambridge University Press, 1988. P. 2.

Если же Джонс попытается выбрать «Б», то Блэк вмешается и заставит его выполнить «А». Можно представить, что у Блэка для этого есть все необходимые условия. Так, он во время сна установил радиоуправляемые электроды в мозг Джонса, которые при необходимости простимулируют последнего сделать «А». У Джонса просто нет альтернатив. И при этом, если он сам выбирает «А», он, видимо, будет нести ответственность. Он несет ее в той же мере, как если бы Блэка вообще не существовало, и у него были бы реальные альтернативы. Джонс может являться морально ответственным агентом, объектом поощрения или порицания, хотя фактически альтернатив у него нет.

В то время как Франкфурт апеллирует к гипотетическим ситуациям, Деннет предлагает обратиться к историческому случаю — вспомнить знаменитое выступление Лютера, с которого началась протестантская реформация. В ходе выступления перед судом обвиняемый в еретичестве Лютер подтвердил свои убеждения, выраженные им в сочинениях, словами: «На том стою. И не могу иначе». Лютер утверждал, что не имеет альтернатив. Вполне вероятно, он выражался буквально. Но разве это была попытка уклониться от моральной ответственности за свой выбор или жалоба на отсутствие свободы? Разве этим высказыванием он ставил под сомнение наличие контроля? Наоборот, в этих высказываниях он принимал ответственность, выражал полное единство и последовательность между своей мыслью и действием. Его позиция детерминировала его решение, не оставляя альтернатив. Но необходимость и отсутствие других возможностей ни в какой мере не освобождала его от ответственности. Деннет утверждает: «Его декларация — это свидетельство того, что мы не освобождаем от ответственности человека за действие, у которого не было альтернатив»<sup>113</sup>.

Этот исторический пример — не исключение. Возьмем ситуацию более близкую, к примеру, мое решение не пытаться невинного человека за

---

<sup>113</sup> Dennett D. Elbow Room: The Varieties of Free Will Worth Wanting. Cambridge, 1984. P. 145.

вознаграждение в 1000 рублей. Я определенно не согласен пытаться человека за такое вознаграждение. У меня нет альтернатив, так как я не вижу никаких оснований совершать такой отвратительный поступок. В той ситуации, в которой я нахожусь сейчас, я не могу поступить иначе. Но разве отсутствие альтернатив здесь лишает меня ответственности за решение?

Примеры Франкфурта и Деннета предназначены для фальсификации Принципа альтернативных возможностей (ПАВ) — принципа, согласно которому «человек морально ответствен за свои действия только в том случае, если он мог поступить иначе». И эту роль они выполнили достаточно эффективно. Они заставили сомневаться в этом принципе, как и в связанном с ним рассуждении — Аргументе последствий.

*с) Аргумент последствий*

Р. Кейн называет Аргумент последствий «одним из самых обсуждаемых аргументов о несовместимости свободы и детерминизма за последние двадцать пять лет»<sup>114</sup>. Различные формулировки этого аргумента были даны К. Жине (1966), Д. Виггинсом (1973) и П. ван Инвагеном (1975). Классическая защита этого аргумента представлена в работе ван Инвагена «Эссе о свободе воли»<sup>115</sup>. В неформализованном виде Питер ван Инваген представляет его так: «Если детерминизм верен, тогда наши действия являются следствиями законов природы и событий в удаленном прошлом. Но от нас не зависело то, что произошло до нашего рождения; и от нас не зависят законы природы. Поэтому и следствия событий и законов (включая наши действия) не зависят от нас»<sup>116</sup>. Для удобства анализа это рассуждение можно разложить на несколько тезисов и заключение.

1. Мы ничего не можем сделать сейчас для изменения прошлого.
2. Мы ничего не можем сделать сейчас, чтобы изменить законы природы.

---

<sup>114</sup> Kane R. The Significance of Free Will. Oxford, N.Y.: Oxford University Press, 1996. P. 44.

<sup>115</sup> Van Inwagen P. Essay on Free Will. Oxford: Clarendon Press, 1983

<sup>116</sup> Van Inwagen P. A Modal Argument for Incompatibilism // Russel, P. and Deery, O. The Philosophy of Free Will. Oxford: Oxford University Press, 2013. P. 117.

3. Мы ничего не можем сделать, чтобы изменить законы природы и прошлое.

4. Если детерминизм верен, наши текущие действия являются необходимым последствием событий прошлого и законов природы.

В. Следовательно, мы ничего не можем сделать, чтобы наши текущие действия были другими.

Первые две посылки являются очевидными. Третья представляет простую конъюнкцию первых двух. Четвертая посылка опирается на утверждение о детерминизме. Вывод исключает наличие альтернатив и, по Принципу альтернативных возможностей, свободу воли. Согласно Аргументу последствий, в детерминистическом мире нельзя поступать иначе, то есть наши действия являются последствиями независящих от нас событий: необходимым результатом законов природы и событий прошлого, в том числе событий, произошедших до нашего рождения. И мы не можем делать что-то другое, нежели то, что делаем. У нас нет реальных альтернативных возможностей. Альтернативы могли бы существовать, если бы было другое прошлое или другие законы природы, или если бы мы могли влиять на законы природы или события прошлого. Но если мы не имеем альтернативных возможностей, а свобода воли предполагает их наличие, значит, либо свободы воли нет, либо наш мир не детерминистичен.

*d) Критика Аргумента последствий. Примеры Франкфурта и значение «иметь возможность».*

Большинство противников Аргумента последствий считают, что проблема — в заключении. По сути дела, они атакуют второй тезис Общего аргумента — утверждение о том, что из незыблемости детерминистических законов природы и прошлого следует отсутствие альтернативных возможностей. Альтернативные возможности в каком-то смысле есть даже в детерминированном мире. Это можно проследить на практическом примере. Я сейчас пишу, а не читаю. Но ведь я мог бы читать. Я мог бы сейчас читать,

писать, расхаживать по комнате или пить чай, несмотря на законы природы и прошлое. Все это является моими актуальными возможностями как полноценного агента. А если аргумент приводит к тезису о том, что это невозможно, скорее всего, аргумент ложный или понятие «иметь возможность» в аргументе задано в каком-то узком специализированном значении. Но тогда область его применения минимальна. Аргумент просто вводит такое значение понятия «иметь возможность», которое в действительности несовместимо с детерминизмом. Тем хуже для инкомпатибилиста. Детерминизм может быть несовместим с понятием «иметь возможность»<sub>1</sub>, но совместим с понятием «иметь возможность»<sub>2</sub>, и как теперь разобраться, какое из понятий необходимо для свободы воли и моральной ответственности? Кажется, впереди патовая ситуация. Одни настаивают на узком определении «возможности», а другие — на широком, опять прибегая к условной интерпретации. Где искать выход?

Р. Кейн считает, что пата в перспективе можно избежать. Но именно в связи с этой патовой ситуацией он предлагает ориентироваться в первую очередь не на Принцип альтернативных возможностей, а на ответственность в конечном счете, то есть на Принцип автономии. «Фокусирование только лишь на альтернативных возможностях (или "разветвляющихся тропках" в будущее) или "возможности поступать иначе", как в случае с Аргументом последствий, — слишком слабое основание, чтобы построить на нем доказательство несовместимости свободы воли и детерминизма»<sup>117</sup>.

В пользу такого решения, по его мнению, говорят также примеры Франкфурта и Деннета. «Так, я согласен с Франкфуртом, что ПАВ (Принцип альтернативных возможностей — *Д.В.*) ложен»<sup>118</sup>, — пишет он. Ошибка Принципа АВ — в том, что он требует, чтобы *каждое* свободное действие

---

<sup>117</sup> Kane, R. *Libertarianism* // Fischer J., Kane R., Pereboom D., Vargas M. *Four Views on Free Will*. New Jersey: Willey Blackwell, 2007. P. 13.

<sup>118</sup> Kane, R. *Rethinking Free Will: New Perspectives on an Ancient Problem* // *The Oxford Handbook of Free Will*. Oxford, 2011. P. 41.

имело альтернативы. Но, как показывают контрпримеры, этого не требуется. Франкфурт и Деннет, по мнению Кейна, действительно уловили важный момент. Но Деннет и Франкфурт неправы в том, что вообще отрицают необходимость альтернативных возможностей для свободы воли. Кейн считает, что они все равно нужны, хотя бы для некоторых решений. Эти решения определяют личность и делают личность свободной. Это он доказывает с помощью Принципа автономии.

#### **В. Принцип автономии и «ответственность в конечном счете»**

Вторым условием свободы, как было установлено выше, является автономия личности. Она выражается в возможности агента быть исходной причиной действий. Все обычные причины, являющиеся частью каузальных цепочек, не подходят для этого. Их истоки уходят в бесконечное прошлое. Для свободы, вероятно, необходимо, чтобы агент не зависел от причин, уходящих в бесконечное прошлое. Только в таком случае он сам устанавливает выбор и действительно контролирует поведение. Уже в ранних работах Кейна учитывается это условие свободы. Философ обозначает его как требование «абсолютного суверенитета» агента над действием (Sole или Ultimate Dominion). В более поздних работах Кейн называет это условие «ответственностью в конечном счете» (ОКС). По определению автора, согласно Принципу ОКС, «чтобы быть ответственным за действие в конечном счете, агент должен отвечать за все, что является достаточной причиной или мотивом действия»<sup>119</sup>. Так, если основанием и объяснением действия агента являются его интенции, агент должен нести ответственность за формирование интенций. Если основанием действия является характер, агент должен нести ответственность и за формирование своего характера. «Принцип ответственности в конечном счете» менее уязвим для критики. И его нельзя подорвать примерами Франкфурта или Деннета. Ведь этот принцип не требует, чтобы для каждого ответственного

---

<sup>119</sup> Kane, R. *Libertarianism* // Fischer J., Kane R., Pereboom D., Vargas M. *Four Views on Free Will*. New Jersey: Willey Blackwell, 2007. P. 14.

действия требовалось наличие альтернативы. Агент, по мнению Кейна, может отвечать за свои поступки даже тогда, когда перед ним — действительно единственно возможное действие. Важно только, чтобы он отвечал за все, что является достаточной причиной или мотивом этого действия.

Кейн объясняет ситуацию с высказыванием Лютера: когда Лютер говорил «На том стою. И не могу иначе», он действительно мог не иметь альтернатив, не быть в состоянии отказаться от собственных текстов и идей. Но он не снимал с себя ответственность. Его выбор был обусловлен твердой позицией, которая, в свою очередь, основывалась предыдущим мучительным выбором, сделанным в результате колоссальных сомнений и неопределенности. Высказывание Лютера — это выражение отрефлексированности и определенности позиции как результата мучительного и свободного выбора. И речь здесь не идет только лишь о внутренней психологической борьбе и когнитивной неопределенности, которую ему пришлось пройти. Речь идет о метафизической неопределенности и сознательном выборе в прошлом.

Именно моменты метафизической каузальной неопределенности Кейн считает основанием моральной ответственности личностей в дальнейшей их биографии. Они позволяют, по его мнению, остановить регресс от ближайших причин к причинам более ранним. И в жизни каждой личности такие моменты имеют особое значение. Они являются характерообразующими: «...свобода воли, в отличие от только свободы действия, заключается в формировании характера и ценностей, что является источником и основой похвалы и порицания, добродетельных и порочных деяний. Свобода воли относится к самосозиданию»<sup>120</sup>. Только в эти моменты агенту необходимо наличие доступных альтернатив.

---

<sup>120</sup> Kane, R. *Libertarianism* // Fischer J. Kane R. Pereboom D. Vargas M. *Four Views on Free Will*. New Jersey: Willey Blackwell, 2007. P. 16.



Таким образом, путь Кейна к инкомпатибилизму можно определить так:

1. *Ответственность в конечном счете* требует, чтобы агент отвечал за достаточные причины своих действий.
2. Достаточные причины действий агента могут уходить бесконечно в прошлое, если только не существует действий на жизненном пути агента, которые не имели достаточных причин.
3. Детерминизм — это положение о том, что все причины с необходимостью определяют действие или что из фактов прошлого и законов природы с необходимостью следуют все последующие события. (Определение детерминизма)
4. Если существуют действия без достаточных причин, то детерминизм ложен.

Этот путь следует называть *«Аргументом от ответственности в конечном счете»*. Его исходной посылкой является Принцип ОКС. Этот принцип приводит Кейна к каузальному регрессу: переходу от непосредственных причин действия к предшествующим причинам. Если допустить бесконечный регресс, ответственность в конечном счете невозможна, т.к. причины действий агентов будут уходить все дальше и дальше в прошлое до момента, когда агента не существовало. А агент, видимо, не может отвечать за события, которые были до его существования. Следовательно, регресс должен быть остановлен где-то в интервале жизни агента. И должна быть найдена исходная причина. Но такой причиной могут быть только действия агента, у которых не было достаточных причин. Это действия, которые не следуют с необходимостью из событий прошлого и законов природы. Но если подобные действия должны существовать, то детерминизм должен быть ложен<sup>121</sup>. Таким образом, Кейн, по его мнению,

---

<sup>121</sup> Аргумент в имплицитном виде содержится в Kane, R. *Libertarianism* // Fischer J. Kane R. Pereboom D. Vargas M. *Four Views on Free Will*. New Jersey: Willey Blackwell, 2007. P. 14-15.

доказывает, что детерминизм несовместим с ответственностью в конечном счете и свободой воли.

### С. Соотношение двух принципов свободы в теории Кейна

Из доказательства видно, что Принцип ответственности в конечном счете играет решающую роль в его рассуждениях. Поэтому Кейна причисляют к «инкомпатибилистам по источнику» (source incompatibilist). И в его аргументе в пользу инкомпатибилизма нет ссылок на альтернативные возможности. Но это не означает, что он отрицает и Принцип альтернативных возможностей как неотъемлемую часть свободы. Это отличает его от сторонников так называемого «узкого инкомпатибилизма» (narrow incompatibilism). Последние критерием свободы воли считают только один, а не два принципа. А Кейн утверждает, что оба принципа необходимы для реализации свободы воли и основания ответственности. Но при этом Принцип ответственности в конечном счете — это базовый принцип, а Принцип альтернативных возможностей — производный. Такую связь Кейн доказывает с помощью набора вымышленных ситуаций, которые называет *«примерами в стиле Остина»*. Эти ситуации противоположны примерам Франкфурта и Деннета. В них у агента есть альтернативные возможности, но нет свободы и ответственности. Они показывают, что, даже если ПАВ необходим для свободы воли, он явно недостаточен.

В первой ситуации предполагается, что для выигрыша чемпионата по гольфу игроку требуется выполнить патт с трех футов — катящий удар с близкого расстояния. Но в связи с нервной судорогой, игрок промахивается. Во второй ситуации киллер совершает покушение на премьер-министра, стреляя из снайперской винтовки, но в связи с конвульсивным движением промахивается и убивает помощника министра. В третьей ситуации Остин сам намерен выбрать в кофе-машине приготовление кофе без сливок, но в связи с неожиданной дисфункцией мозга случайно нажимает кнопку для приготовления кофе со сливками. Во всех трех случаях можно допустить, что

в цепочке событий присутствует настоящая случайность. Представим, что в нормальные каузальные процессы вторгается квантовая неопределенность. Т.е. все описанные события: промах, убийство помощника и приготовление кофе со сливками — недетерминированы, т.е. не имели достаточных причин. И эти события, соответственно, могли иметь реальные альтернативы. Игрок мог попасть в лунку, убийца — убить премьер-министра, а Остин, желавший кофе без сливок, мог бы получить кофе без сливок. Принцип альтернативных возможностей удовлетворен.

Но считаем ли мы, что все эти действия были свободными? — спрашивает Кейн. И были ли они добровольными? Хотел ли игрок в гольф промахнуться, убийца — убить помощника министра, а Остин — выпить кофе со сливками? Кажется, все эти действия были произвольными и несвободными. Агенты, совершившие эти действия, не выбирали их, а совершили их по случайности или ошибке. Следовательно, делает вывод философ, индетерминизма в сочетании с альтернативными возможностями недостаточно для свободы воли. Для свободы воли агенты должны иметь возможность самостоятельно управлять своей волей, быть её источником и исходной причиной. А это условие не выполняется в представленных ситуациях.

И все-таки альтернативные возможности необходимы для свободы. Кейн рассуждает так: Свобода воли предполагает ответственность в конечном счете за волевые акты и действия агентов. В свою очередь, как следовало из предыдущего аргумента, ответственность в конечном счете требует действий, не продиктованных необходимостью. Эти действия являются воле-устанавливающими (*will-setting actions*). Неопределенность, предшествующая воле-устанавливающим действиям, должна быть желаемой. Но если неопределенность должна быть желаемой, значит, отчасти желаемыми должны быть несколько альтернатив. Значит, неопределенность

должна состоять из желаемых реальных альтернатив (plurality conditions). И следовательно, реальные альтернативы должны существовать.

Шаги с первого по третий уже встречались выше. Новым является определение воле-устанавливающего действия. Кейн объясняет: «Действия без достаточных причин будут действиями, в которых волевые установки агента не были еще до конца определены до момента совершения действия. Именно в ходе этих действий агент установит свою волю окончательно»<sup>122</sup>. Эти действия отличаются от действий с реальными возможными альтернативами в «примерах Остина». Действия в этих примерах не являются воле-устанавливающими, так как воля агентов в них установлена заранее. Так, игрок в гольф до момента неопределенности уже определился в намерении попасть в лунку, убийца — убить премьер-министра, а Остин — выпить кофе без сливок. Поэтому их исход был просто случайностью. И примеры в стиле Остина не являются иллюстрациями воле-устанавливающих действий.

Другое дело, если бы желания агентов в примерах не были однозначными, если бы у них были конфликтующие желания. Воле-устанавливающие действия — это ситуации выбора между двумя и более конкурирующими желаниями, когда выбор этот неясен вплоть до момента осуществления действия. Тогда именно действие устанавливает волю. Так, если бы Остин хотел и кофе без сливок, и капучино, а убийца — и убить премьер-министра, и завязать со своей криминальной карьерой, не причиняя никому вреда, и эти желания конфликтовали бы вплоть до самого действия — вот тогда это были бы примеры устанавливающих волю действий.

Из этих же новых примеров следует и то, почему у воле-устанавливающих действий должны быть желаемые альтернативы. Ведь только тогда любой из желаемых исходов не будет случайным. Для

---

<sup>122</sup> Kane, R. *Libertarianism* // Fischer J.M., Kane R., Pereboom D., Vargas M. *Four Views on Free Will*. New Jersey: Willey Blackwell, 2007. P. 20.

остановки бесконечного каузального регресса нужны действия с недостаточными причинами. У этих недетерминированных действий должны быть конфликтующие основания. Эти конфликтующие основания — желаемые альтернативы действий. А если альтернативы желаемые, они получаются и реальными, т.к. в недетерминированной ситуации любой из сценариев может реализоваться. Таким образом, Кейн из Принципа ответственности в конечном счете выводит сначала индетерминизм, а затем — Принцип альтернативных возможностей. Эти два вывода он называет *двойным регрессом* свободы воли. Но допустим, что ему удастся провести этот двойной регресс. Допустим, его аргументы истинны, и Кейн поднялся на вершину инкомпатибилизма. Что дальше, ведь, как он сам утверждал, это только половина пути?

### **3. Свобода воли и индетерминизм совместимы**

#### **А. Контрпримеры к классическому Аргументу случайности**

Миры с индетерминизмом можно строить по-разному. Можно представить индетерминированный мир как череду случайных событий. В таком мире объекты будут возникать, пропадать и вновь появляться непредсказуемо в совершенно разных местах и конфигурациях. История такого мира будет представлять хаос, ленту из случайно подобранных кадров, персонажи которой перемешиваются, исчезают, преобразуются, превращаются в предметы. В подобной ситуации нельзя будет уловить сюжетную линию или предвидеть события. Нельзя будет предвидеть будущее даже на мгновение вперед. И объяснить наличное состояние также будет нельзя. Ведь там не будет регулярностей, рациональности, законосообразности и контроля. Конечно, в таком мире не будет места и ответственности. Но это не значит, что у сторонника либертарианского подхода нет выхода. Описанная ситуация — лишь карикатура на либертарианскую альтернативу мира с индетерминизмом. Существуют и другие, более реалистичные варианты.

Кейн конструирует осмысленный вариант мира, где индетерминизм существует в определенных рамках. Он предлагает рассматривать мир, где неопределенность только эпизодическая. Т.е. индетерминированные события происходят иногда наряду с событиями детерминированными. Но даже эти эпизодические индетерминированные события не являются беспричинными. Просто причины не с необходимостью определяют действия, а с некоторой вероятностью. Это — вероятностная каузальность. При этом в случае, если действие не случается, случается другое, но совсем не любое. Возможными событиями являются не любые случайные события, а ограниченное число альтернатив. В таких мирах нет хаоса или просто произвольного стечения событий. Они вполне регулярны и предсказуемы. Именно такое устройство мира представляется необходимым Кейну для существования свободы воли и ответственности. За иллюстрациями далеко ходить не придется.

Подходящие случаи с индетерминированными, но контролируемыми, предсказуемыми и объясняемыми событиями уже встречались в нашей работе. Речь идет о примерах «в стиле Остина». Кейн в данном случае использует их в качестве контрпримеров к Аргументу случайности. Напомню один из них. Наемный убийца совершает покушение на премьер-министра. Он направляет снайперскую винтовку на жертву. Но в его движениях есть неопределенность. И от этой неопределенности зависит, будет выстрел точным или нет. Эта неопределенность связана с квантовыми эффектами в его мозге. Вот происходит выстрел, пуля попадает в государственного чиновника, и премьер-министр погибает, а убийцу хватают агенты спецслужб. В этой ситуации убийца несет ответственность за преступление, даже несмотря на неопределенность в цепочке событий. И он не сможет объяснить свою непричастность тем, что между его решением выстрелить и выстрелом имели место индетерминированные нейронные процессы. Ведь его намерения были уже определены. Случайность могла ему только помешать. Для того чтобы вменить вину, есть достаточные основания

— mens rea и actus reus. Убийца действовал преднамеренно. И он виноват, он несет ответственность. Индетерминизм в данном случае, вероятно, сочетается с ответственностью.

Ну а вот другой контрпример. Математик пытается решить сложное уравнение. Его намерение определено, но результат вычислений зависит от квантовой неопределенности в активации нейронных сетей его мозга. Допустим, он решает уравнение. Можно ли считать его ответственным за решение? По всей видимости, да. Ведь он старался это сделать, и сделал это. Неопределенность была лишь помехой в работе его мозга, как шум, помехи, мешающие слушать радиопередачу. Результат его действия был объясним, рационален и отчасти предсказуем. Это — также случай совмещения индетерминизма с ответственностью.

На этих контрпримерах, «примерах в стиле Остина», Кейн показывает, что индетерминизм совместим с ответственностью. И что индетерминированное событие не всегда является событием случайным, произвольным. Оно может быть хотя бы отчасти контролируемым и отчасти предсказуемым. С помощью этих ситуаций Кейн также демонстрирует, что причина не всегда создает необходимость действия. Достаточно, чтобы причина создавала вероятность действия. Как намерение убийцы застрелить премьер-министра и намерение математика создают вероятность смерти чиновника и вероятность решения уравнения. Эти иллюстрации позволяют Кейну критиковать классический аргумент Юма и Принцип случайности. Однако их недостаточно, чтобы отразить модернизированную модель аргумента, ведь в контрпримерах только одно из альтернативных действий агента является намеренным и желаемым, и только одно действие является неслучайным.

#### **В. Ответ на модернизированный Аргумент случайности**

В приведенных выше гипотетических ситуациях границы индетерминизма слишком узки, чтобы обеспечить *ответственность* «в

*конечном счете*». Ответственность действует в рамках только одного замысла. Альтернативы у этого замысла есть: не попасть в цель или не решить задачу. Но эти альтернативы неумышленные. Поэтому они представляют собой случайность, погрешность, ошибку. Контроль в этих случаях у агентов тоже ограниченный. Убийца контролирует только свою попытку убить премьер-министра. Но у него нет контроля над другими, ошибочными альтернативами. Так, если он не попадет в премьер-министра, его все равно могут схватить и судить за покушение. Его намерение было установлено, и только случайность не позволила ему совершить преступление. Он не контролирует попадание пули в помощника, в автомобиль премьер-министра или промах. Убийцу можно осуждать за убийство премьер-министра, но нельзя благодарить за то, что он промазал или, скажем, спас премьер-министра, случайно убив другого снайпера, тоже целившегося в чиновника. Нельзя и рационализировать его промах исходными намерениями: «он промазал, потому что хотел попасть». Примерно такая же ситуация — с математиком. Он намерен решать уравнение. Если он решит его, он отвечает за успех вычислений. Но если ввиду индетерминизма решить он уравнение не смог или выдал какое-то другое, неправильное решение, это будет случайность, за которую он нести ответственность, кажется, не должен.

Для опровержения модернизированного аргумента нужно, чтобы в контрпримере все альтернативы были желаемы и соответствовали намерениям агента. И чтобы все альтернативы были неслучайными и контролируемы агентом. Чтобы сконструировать такую ситуацию, нужно, чтобы индетерминизм действовал не на этапе реализации одного какого-то замысла, а до этого момента — во время выбора между разными замыслами. Иными словами, нужно переместить зону неопределенности с периферии в центр, из области, где конкретное решение уже выполняется, в область, где только принимается решение. Тогда ситуация будет подпадать под



модернизированный Аргумент случайности. Назовем эти ситуации «примерами в стиле Кейна». Сохранится ли в таких примерах ответственность агента? Будут ли это контрпримеры или, напротив, иллюстрации к аргументу? Проведем испытание на двух гипотетических ситуациях.

*Случай «Ограбление»*<sup>123</sup>. Молодой человек во время прогулки в парке замечает, что происходит ограбление: преступники отбирают сумку у прохожей. Перед молодым человеком встает дилемма. С одной стороны, моральный долг призывает его вмешаться в конфликтную ситуацию, заступиться за женщину. С другой стороны, его пугает, что попытка защитить женщину может закончиться неприятностями для него самого. Молодому человеку нужно решиться: пройти мимо или попытаться помочь. Альтернативы действий продиктованы конфликтующими мотивами. Молодому человеку приходится выбирать между эгоцентрическими интересами и моральными альтруистическими установками. И этот выбор сопровождается внутренней борьбой. Внутренняя борьба приводит к индетерминизму. В результате этого индетерминистического процесса молодой человек делает выбор — вступить за прохожую. Он совершает поступок в соответствии со своими моральными установками.

*Случай «Выбор Ральфа»*<sup>124</sup>. Ральф совершает выбор жизненного пути: он решает, оставаться ему в Мэйберри или переезжать в Нью-Йорк. Мотивом переехать в Нью-Йорк служат его желания играть за футбольную команду Гигантов и быть звездой в бродвейском шоу. Мотивом остаться служат его желания взять замуж Робби Анну и управлять местным рестораном «ДерВиннершницель». Альтернативы несовместимы между собой, но в одинаковой степени желаемы для Ральфа. Выбор альтернативы

---

<sup>123</sup> Подобный пример приводит Р. Кейн в книге «Значимость свободы воли». Kane R. The Significance of Free Will. Oxford, N.Y.: Oxford University Press, 1996. P. 126.

<sup>124</sup> Подобный пример приводит М. Балагер, автор схожей с кейновской либертарианской позиции. Balaguer M. Free Will as an Open Scientific Problem. Cambridge: A Bradford Book, 2010. P. 72.

сопровождается внутренней борьбой, которая приводит к индетерминизму. В результате индетерминистического процесса Ральф принимает спорное решение ехать в Нью-Йорк. Он *просто решает* ехать в Нью-Йорк.

Несут ли ответственность молодые люди за принятые ими решения? Справедливо ли считать отважный выбор заступника его заслугой, или это — просто случайность? Отвечает ли Ральф перед Робби Анной за свой отъезд?

По мысли Кейна, оба молодых человека несут ответственность за свои действия. Более того, именно такие действия являются ключевыми для агентов. Именно их Кейн называет *воле-устанавливающими действиями (ВУД)*. Они совершаются, когда достаточные мотивы есть для реализации различных альтернатив, когда образуется конфликт из равносильных мотивов внутри агента, и этот конфликт создает каузальную индетерминированность. В этот момент перед агентом распаивается реальная возможность альтернатив. Если бы не эти действия, у агентов не было бы возможности быть иными. И в этот момент обрывается каузальная цепочка, разрушается бесконечный регресс причин и действий. Агент как бы левитирует над своей каузальной историей. Тогда возникает возможность ответственности. Тогда формируется личность.

Воле-устанавливающие действия определяют характер агента, который впоследствии будет детерминирующим для многих последующих действий. И именно они оправдывают ответственность и за все последующие детерминированные действия, за действия, у которых нет реальных альтернатив. Эти последующие действия являются результатом установок воли, сделанных во время воле-устанавливающих действий. Накопление ВУД приводит к накоплению оснований для ответственности. Чем больше таких действий произведено, тем больше оснований считать человека ответственным за свое текущее состояние, тем больше оснований считать его последующие действия зависимыми именно от него, а его за них — *ответственным «в конечном счете»*.

Именно так, по мнению Кейна, обстояло дело и в примере с Лютером, когда он говорил: «На том стою. И не могу иначе». Лютер в момент этого выступления действительно мог не иметь альтернатив. Его поведение могло быть детерминировано. И он это констатировал, но ответственность за свою позицию он приобрел в силу того, что альтернативы у него были раньше, во время мучительных размышлений и свободного выбора, в то время, когда он совершал воле-устанавливающие действия. Его ответственность базировалась на множестве произведенных ранее индетерминированных тяжелых решениях.

Возможность осуществления ВУД в мозговых процессах, по мнению Кейна, не противоречит физике и нейрофизиологии. Об индетерминистических процессах на квантовом уровне было уже сказано. Как я говорил выше, квантовая неопределенность сглаживается на макроуровне. Тем не менее, Кейн считает возможным периодическое распространение этой неопределенности и на макроуровень. Эта возможность основывается им на способности мозга к параллельным вычислениям. Широко известно, что мозг, в отличие от однопроцессорного компьютера, способен выполнять когнитивные задачи сразу по множеству потоков. В нем нет узкого горла, через которое должны пройти все вычисления. Эта особенность мозга и создает механизм усиления неопределенности. Это происходит, когда есть внутренний конфликт, неопределенность в уме, — пишет Кейн, — когда напряженная работа мозга приводит к нарушению термодинамического равновесия — возникновению «хаотических процессов в мозге, которые делают его чувствительным к индетерминированным микропроцессам на нейронном уровне»<sup>125</sup>.

Допустим, нейрофизиологическая возможность ВУД доказана. Но почему мы должны все-таки верить интуициям Кейна? Почему должны

---

<sup>125</sup> Kane R. Responsibility, Luck and Chance: Reflections on Free Will and Indeterminism // Russel P. and Deery O. (eds.). The Philosophy of Free Will. Oxford, New York: Oxford University Press, 2013. P. 199.

считать молодого человека в ситуации индетерминированного выбора ответственным за принимаемое решение? У либертарианца есть два основания считать так. **Во-первых, выбор, сделанный агентом, оценивается им самим как его собственный (феноменологическая основа).** Очевидно, что в подобной ситуации сам агент принял бы этот выбор. Ведь агент сам напряженно размышляет над альтернативами, эти альтернативы продиктованы его собственными ценностями, желаниями и убеждениями, и хотя в выборе присутствовал индетерминизм, это был его выбор, а не кого-то другого. Он не может пожаловаться, что выбор случился сам по себе. И что это неожиданное для него решение. Дело было бы совсем иначе, если бы выбор пал на альтернативу, которую агент даже не обдумывал. Тогда, возможно, он бы не принял этот выбор, а удивился ему. Но в нашей ситуации это был выбор его, по крайней мере, из феноменологической перспективы.

**Во-вторых, о выборе, описанном выше, нельзя сказать, что это выбор совсем случайный.** Когда говорят о случайном действии, говорят обычно об ошибке, недоразумении, произвольном, непреднамеренном событии. Разве выбор защищать прохожую был случайным, ошибочным недоразумением? Он был обоснованным, его можно полностью объяснить мотивами агента. Ситуация в целом похожа на ту, когда убийца, намеревающийся убить премьер-министра, попадает в цель. И не похожа на ту ситуацию, когда убийца промахивается. Ведь действия агента соответствуют его намерениям. Специфика ситуации с ВУД, однако, в том, что намерениям соответствуют оба исхода, обе альтернативы. В обоих случаях агент успешно совершает намеренные действия, в обоих случаях ему удается сделать то, что он хочет. Даже если бы молодой человек не бросился на помощь прохожей, а скрылся бы — это тоже было бы успешным выполнением его замысла. В этом смысле он имеет контроль над обеими перспективами. Кейн называет это *множественным волевым контролем*.

Обладать «множественным волевым контролем» над набором альтернатив — значит иметь возможность реализовать любую из альтернатив, которую желаешь, или больше всего желаешь, по своим предпочтениям и мотивам, без внешнего принуждения<sup>126</sup>. Множественный волевой контроль над действиями является основанием для ответственности. И он является вариантом расширенного контроля, так как предполагает контроль над альтернативами.

#### 4. Возражения к теории Кейна

Теория Кейна в целом продуктивна. С помощью индетерминизма философу удастся наделить агента альтернативными возможностями действий и прервать бесконечный регресс каузальности, уходящий в далекое прошлое. С помощью «случаев в стиле Остина» ему удастся показать, что индетерминизм не всегда оборачивается случайностью. Кейн также демонстрирует возможность контроля агентом в индетерминистической ситуации, в частности двойного волевого контроля во время волеустанавливающих действий. Теория согласуется с физикой и нейрофизиологией. Все это, безусловно, преимущества теории. Но у нее есть и существенные недостатки. На мой взгляд, Кейну (1) не удастся показать, как индетерминизм создает принципиальную возможность *ответственности в конечном счете*. Но даже если такая принципиальная возможность существовала бы в идеальном мире, (2) философ не обосновывает практическую реализуемость ее в реальном мире.

Главной целью Кейна является демонстрация возможности *ответственности в конечном счете*. «Чтобы быть ответственным в конечном счете за действие, агент должен отвечать за все, что является достаточной причиной или мотивом действия»<sup>127</sup>. Случаи индетерминизма, подобные случаю с убийцей и математиком, не являются актами

---

<sup>126</sup> Kane R. Responsibility, Luck and Chance: Reflections on Free Will and Indeterminism // Russel P. and Deery O. The Philosophy of Free Will. Oxford, New York: Oxford University Press, 2013. P. 209-210.

<sup>127</sup> Kane, R. Libertarianism // Fischer J., Kane R., Pereboom D., Vargas M. Four Views on Free Will. New Jersey: Willey Blackwell, 2007. P. 14.

самосозидания агента. Для того, чтобы агент был ответственен в конечном счете за действие, он должен быть ответственен за формирование своего замысла. Это возможно только в случае, если он ответственен за свои воле-устанавливающие действия, которые определили его замысел (*случаи в стиле Кейна*). Но, по моему мнению, доказать, что агент ответственен в конечном счете за воле-устанавливающие действия, Кейну все-таки не удастся. Т.е. Кейну не удастся убедить, что в «случаях в стиле Кейна» агент является ответственным в конечном счете за действие.

#### **А. Отсутствие ответственности в конечном счете в случае ВУД**

«Случаи в стиле Кейна» изначально рождают противоречивые интуиции в отношении наличия ответственности. Конечно, после рассуждений Кейна все-таки склоняешься к тому, что молодой человек, решающий вступить за девушку, действительно отвечает за свой выбор. Но тот же самый случай можно прокомментировать иначе, и тогда интуиции будут противоположными. Можно обратить внимание на то, что молодой человек в двух соседних возможных мирах — там, где он в момент  $t$  помогает прохожей, и там, где решает пройти мимо — имеет совершенно одинаковые предшествующие состояния. Ничто в его убеждениях и желаниях, размышлениях и усилиях, характере и достоинствах не отличается в двух соседних мирах. И тем не менее, в одном случае он выходит из ситуации трусом, а в другой — героем. При этом объяснить разницу невозможно. Разве тогда исход — не случайность? Разве нельзя сказать, что ему просто повезло в одном случае оказаться героем и не повезло в другом, где он повел себя как трус?

Интерпретации «случаев в стиле Кейна» как ответственных действий зависит от их связи с «примерами в стиле Остина». Именно на этой аналогии строит заключения Кейн. Но связь между этими типами случаев в отношении ответственности проблематична. Ведь зона индетерминизма в них расположена в разных местах, а это принципиально важно. Существует

несколько вариантов расположения зоны индетерминизма в функциональном устройстве агента. Во-первых, действие может начинаться индетерминированным образом. Т.е. альтернативы действий могут формироваться в ходе индетерминистического процесса. Во-вторых, выбор из набора альтернатив может осуществляться индетерминированным образом. И, наконец, в-третьих, сама реализация действия может проходить с элементами индетерминизма. Кейн сначала приводит случаи «в стиле Остина», относящиеся к третьей категории, для доказательства совместимости индетерминизма с ответственностью. А потом делает экстраполяцию на вторую категорию, на «случаи в стиле Кейна». Мне кажется, что это неоправданный шаг. Ответственность, очевидно присутствующая в случаях первой категории, убывает в случаях, относящихся ко второй категории. Перенесение зоны индетерминизма с периферии действия к центру принятия решений значительным образом снижает ответственность. Здесь подойдет такая иллюстрация.

Предположим, студент решает математическую задачу, в которой есть один правильный ответ. В случае А, он дает один ответ. В случае В, он дает несколько ответов. При этом он не может сделать выбор между ними, и настаивает на правильности каждого варианта. В третьем случае, С, студент использует уже готовые варианты ответов и выбирает один из них. В каком случае он несет большую ответственность за решение? В случае А, как мне кажется, ответственность за решение не вызывает сомнения. Либо решение правильное, и студент должен получить высокую оценку, либо решение неправильное, и оценка должна быть низкая. В случае Б, ответственность, кажется, меньше. Позиция студента неопределена. Оценить его работу очень сложно. Так как даже если среди его решений есть правильные, он не определился до конца с выбором. Тем не менее, если среди его ответов нет правильного, можно сказать, что его решение точно неверное, и он заслуживает отрицательной оценки. А как быть в третьем случае? Мне

кажется, что в случае С ответственность студента за решение еще меньше. Если ответ правильный, нельзя сказать, насколько студент на это повлиял. Возможно, что ему были предложены все варианты только с одним правильным ответом. Аналогично, если он представил неправильный вариант, он может быть к этому непричастным, если выбор был только между неправильными вариантами. Думаю, иллюстрация показывает, что передвижение зоны индетерминизма с периферии в центр влечет убывание ответственности и даже ее исчезновение. Следовательно, оснований распространять выводы, относящиеся к случаям «в стиле Остина», на случаи «в стиле Кейна» у философа нет.

Важно отметить, что я здесь не утверждаю, что в случаях «в стиле Кейна» ответственности совсем нет. Некоторая ответственность, по моему мнению, может существовать. Но она меньше, чем в случаях «в стиле Остина». А из этого следует, что накопление ВУД в личной истории агентов ведет не к увеличению их ответственности за действия, а к снижению. Это прямо противоречит утверждениям Кейна. Ведь Кейн считает, что чем больше ВУД в истории человека, тем ближе он к *ответственности в конечном счете*.

Если в этом все еще остаются сомнения, я предлагаю представить следующую ситуацию. Снайпер готовится к покушению на премьер-министра. Но в это время его терзают сомнения. С одной стороны, к убийству его подталкивают корыстные мотивы. Он представляет, как заказчики выплатят ему солидную премию за убийство и он сможет обеспечить безбедное существование детям. С другой стороны, его мучает совесть. Он понимает, что может совершить преступление, в результате которого пострадает множество невинных людей. Это случай ВУД, т.е. его выбор не детерминирован. Но допустим, что, неведомо для него, за состоянием его мозга следит заказчик преступления. Некоторое время назад он установил в мозг снайпера радиопередатчики, которые позволяют



наблюдать за его состояниями и влиять на них. Передатчики этого устройства не очень надежны, т.е. в их действии имеется индетерминизм. С некоторой вероятностью они могут определить выбор снайпера. Обнаружив, что снайпер сомневается и покушение может не состояться, заказчик пытается вмешаться в работу мозга снайпера. Устройство успешно срабатывает, и выбор снайпера определяется в пользу покушения. Несет ли снайпер полную ответственность за покушение и убийство? Является ли это решение контролируемым самим агентом, или оно продиктовано чем-то внешним? Думаю, здесь читатель сочтет ответственность снайпера ограниченной. В каком-то смысле он виноват за то, что допустил вариант с преступлением. Но ведь он реально мог поступить иначе, если бы не вмешательство заказчика покушения. Значит, его контроль и ответственность редуцированы. И агент не имеет ответственности в конечном счете. Хотя при этом выполнены условия множественного волевого контроля.

Сторонник Кейна мог бы возразить: но в описанном выше случае очевидно вмешательство стороннего агента! Это случай манипуляции. Естественно сложится впечатление, что контроль редуцирован. На это я отвечаю: да, конечно, в этом конкретном случае есть вмешательство стороннего агента. И это — действительно манипуляция. Но заказчика с его радиопередатчиком могло бы не быть, а результат остался бы тем же. Просто роль заказчика выполнил бы случай. Случай тоже может быть манипулятором. Теперь стороннику Кейна нужно показать, чем отличается ситуация, когда заказчик с помощью индетерминированного процесса управляет решением снайпера, от ситуации, когда решением снайпера управляет случай. А до тех пор, пока он этого не сделал, можно утверждать, что индетерминированность событий уменьшает контроль, уменьшает

ответственность<sup>128</sup>. И наличие некоторого количества таких событий исключает возможность ответственности в конечном счете.

Перед тем, как перейти к другому возражению, хочу предложить одно направление мысли, которое могло бы улучшить позицию Кейна. Речь идет об оснащении «случаев в стиле Кейна» дополнительными условиями. Обновленные случаи я назову «случаями в стиле Васильева». Эти случаи навеяны критикой Аргумента случайности, предложенной российским философом в одной из частных бесед. Представим, что перед певицей стоит хрустальный бокал. То, что она возьмет ноту в третьей октаве, является необходимым условием того, чтобы бокал разбился. Т.е. если она не возьмет ноту, бокал точно не разобьется. Но связь между этими событиями индетерминистическая: если она возьмет ноту, бокал может разбиться, а может и нет. Вот она берет ноту в третьей октаве, и бокал разлетается вдребезги. Очевидно, что певица виновата в бое посуды. Но чем, вы скажите, это отличается от «случаев Остина»? Васильев подчеркивает, что когда действие агента является необходимым условием события, ответственность будет очевидна.

Думаю, если оснастить ВУД Кейна умышленными событиями, которые бы являлись необходимыми условиями для обеих альтернатив, агент будет более ответственным. Примеры, которые подходят под такие условия, могут выглядеть немного искусственными. Но в принципе они реализуемы. Вот хотя бы ситуация с молодым человеком, решающим помогать или не помогать жертве грабежа. Предположим, что необходимым условием для эгоистической альтернативы является ускорение шага, а для альтруистической — поиск баллончика с газом в сумке. Если молодой человек совершил оба этих необходимых действия, а потом в результате индетерминистического процесса выбрал одну из альтернатив — он будет в

---

<sup>128</sup> Похожий пример предлагает Д. Перебум в параграфе Randomizing Manipulators // Pereboom D. Living Without Free Will. Cambridge, 2001. P. 50.

большой степени ответственен за поступок. Это и будет «примером в стиле Васильева». Я считаю, что ответственность из случаев в стиле Остина нельзя экстраполировать на случаи в стиле Кейна, но можно — на «случаи в стиле Васильева». Однако случаи последнего типа вряд ли часто встречаются. Вряд ли можно ожидать, что человек часто будет вести себя несколько непоследовательным образом. Впрочем, и случаи в стиле Кейна маловероятны. Это приводит нас ко второму возражению.

#### **В. Малая вероятность реализуемости ответственности в конечном счете в реальности**

Второе возражение относится к практической возможности ответственности агентов. Кейн утверждает, что *ответственность в конечном счете* возникает тогда, когда накапливаются волеустанавливающие действия. Они формируют ответственную личность. Однако для того, чтобы такие действия реализовались, необходимо, чтобы в определенный момент у агента сформировались конфликтующие желания, и при этом сила этих желаний была одинаковой. Но какова вероятность такого стечения обстоятельств? По мнению В. Васильева, вероятность совпадения силы желаний практически ничтожна. А значит, и ответственные личности в нашем мире — крайняя редкость. Думаю, на это есть возражение. Совпадение силы желаний — редкость только в том случае, если мы представляем, что желания формируются сразу определенной силы или если их сила изменяется скачкообразно. Если же сила желаний возрастает и падает плавно, вероятность совпадения не так мала.

К примеру, если взять все сосуды, в которых есть жидкость в настоящее время, практически невероятно найти сосуд, заполненный точно наполовину. Но если взять один сосуд и постепенно наполнять его водой, то в какой-то момент он точно будет заполнен наполовину. Если учесть, что мозг — это параллельный процессор и в нем могут реализовываться и изменяться сразу множество желаний, вероятность совпадения должна быть

существенной. Впрочем, так это или нет, очевидно, что этот момент недостаточно освещен Кейном, и это является недостатком теории.

Таким образом, я считаю, что либертарианство Кейна не достигает своей цели. Индетерминизм в этой теории не создает условий для *ответственности в конечном счете*. На место бесконечного регресса вступает все больше господство случайностей. И теперь не цепочка причин, но случай ограждает агента от ответственности в конечном счете. Что остается? Признать, что моральная ответственность несовместима как с детерминизмом, так и с индетерминизмом? Такая позиция называется твердым инкомпатибилизмом. Одним из самых сильных сторонников современного *твердого инкомпатибилизма* является Д. Перебум.

### **III. Свобода воли и моральная ответственность в твердом инкомпатибилизме Д. Перебума**

По мнению Р. Кейна, моральная ответственность несовместима с детерминизмом, и Д. Перебум с этим согласен. Поэтому обе позиции являются вариантами инкомпатибилизма. При этом Кейн считает, что свобода воли и моральная ответственность совместимы с индетерминизмом, а по убеждениям Перебума — нет. Моральная ответственность и свобода воли несовместимы как с детерминизмом, так и с индетерминизмом, — считает Перебум. Поэтому его позиция называется твердым инкомпатибилизмом. Согласно этой теории, в любом из миров, где каузальность связывает события необходимостью или вероятностью, ответственность исключена. Ответственность, по его мнению, существовала бы, только если бы каузальность связывала субстанции, не детерминированные причинами, и события. Но такое положение дел, вероятно, противоречит науке. Следовательно, в нашем мире, скорее всего, свободы и моральной ответственности нет.

Скептический вывод относительно свободы, однако, не подрывает основания морали и смысла жизни. По мнению философа, он только способствует эволюции этих представлений. Классическая концепция ответственности во многом опирается на идею базовой заслуги<sup>129</sup>, но это не единственно возможная опора. Можно отказаться от идеи заслуги и сохранить некоторый вариант ответственности. В соответствии с этой обновленной идеей ответственности, поощрять людей будет все равно оправдано, так как поощрение способствует распространению положительных примеров в обществе. Изменениям подвергнется практика наказаний — она приблизится к практике «лечения» и «карантина». Такого рода наказания сохранят свою основную функцию: они будут оберегать общество от криминального поведения и препятствовать его распространению. При этом они будут более гуманны в отношении преступников. «Эти предложения не исключают из нашей практики применения... ответственности даже в криминологическом аспекте. Скорее, они приведут к пересмотру практики с учетом уже существующих правил. Поэтому предложения, которые представляет скептик в отношении свободы воли, не являются принципиальными по отношению к нашей практике... применения моральной ответственности»<sup>130</sup>.

В этих утверждениях — суть позиции твердого инкомпатибилизма Перебума. Теперь рассмотрим основные аргументы в их пользу. Сначала стоит обратиться к аргументам Перебума против совместимости свободы воли и индетерминизма, то есть против либертарианства, в частности против позиции, которую отстаивает Кейн. Следующим шагом я приведу анализ аргументов против совместимости свободы воли и ответственности, то есть доводы Перебума против компатибилизма. Моя собственная позиция и критические замечания в отношении позиции Перебума будут представлены в последней, четвертой части данной главы.

---

<sup>129</sup> Pereboom D. Free Will, Agency, and Meaning in Life. Oxford, New York: Oxford University Press, 2014. P. 2.

<sup>130</sup> Pereboom D. Free Will, Agency, and Meaning in Life. Oxford, New York: Oxford University Press, 2014. P. 174.

## 1. Свобода воли и индетерминизм несовместимы

### А. Индетерминизм не расширяет контроль

Кейн считает, что индетерминизм становится основанием моральной ответственности, когда конфликтующие мотивы достигают равной силы и агент вынужден принимать одно из решений («случаи в стиле Кейна»). В такой ситуации перед агентом раскрываются альтернативы. В силу индетерминизма агент может поступить так или иначе, но выбор при этом не перестает быть выбором самого агента. Контроль над выбором в таких ситуациях не теряется. Напомню, Кейн представляет два основания считать так. Во-первых, решение объективно не перестает быть желаемым, намеренным, понятным, целенаправленным. Оно направлено мотивами агента и является результатом его усилий. Нельзя назвать это решение случайным или ошибочным. Во-вторых, агент субъективно считает эти решения собственными. Т.е. с феноменологической точки зрения они представляются контролируруемыми решениями. Но эти основания присутствуют и в арсенале компатибилиста. Они существовали бы даже в том случае, если бы «трудный выбор» агента происходил в детерминистическом мире: с феноменологической стороны детерминированный выбор также представлялся бы агенту не случайным, а намеренным, желанным, ожидаемым. И объективно этот выбор соответствовал бы мотивам агента. Тогда в чем тут преимущество индетерминированного выбора?

Кейн считает, что преимущество заключается в доступности альтернативных возможностей. По его мнению, альтернативные возможности создают условия для самоопределения агента. Они позволяют осуществить множественный волевой контроль в случае волеустанавливающих действий. А такой вид контроля не может присутствовать в детерминистическом мире. Перебум, однако, не видит в нем ценности. С его точки зрения доступ агента к альтернативным возможностям ничего принципиально не меняет. Агент мог бы быть ответственным, даже если бы

их не было. Позиция Перебума в данном случае основана на уже упомянутых примерах Франкфурта. Перебум выступает защитником этих примеров и даже придумывает улучшенный собственный пример «в стиле Франкфурта».

*а) Примеры «в стиле Франкфурта»*

Напомню, в примерах Франкфурта у агента отсутствуют альтернативные возможности, при этом интуиции подсказывают, что он несет моральную ответственность. Примеры направлены против принципа альтернативных возможностей. Выше я приводил, в частности, такую гипотетическую ситуацию: Блэк (контролер) хочет, чтобы Джонс выполнил действие «А». Для этого он установил радиоуправляемые электроды в мозг Джонса, которые при необходимости простимулируют последнего сделать «А». Но в момент выбора Джонс совершает действие «А» по собственному желанию, без вмешательства Блэка. По всей видимости, Джонс сам несет ответственность за «А». Ведь Блэк так ничего и не предпринял. Все произошло, как если бы Блэка и его устройства вообще не было. При этом Джонс не мог поступить иначе. Ведь если бы он попытался поступить иначе, Блэк вынудил бы его выполнить действие «А» с помощью радиоуправляемого механизма.

На первый взгляд, этот пример Франкфурта вполне убедителен. Может показаться, что он достигает поставленных автором целей: демонстрирует ситуацию, в которой у агента отсутствуют альтернативы, но имеется ответственность. Однако, как и в случае с большинством других мысленных экспериментов, здесь многое зависит от деталей. Этому примеру Франкфурта нужны уточнения, которые могут оказаться принципиальными. Необходимо определить, как именно Блэк узнает, что ему не требуется вмешиваться. Что должно произойти, чтобы он точно понял, что не нужно действовать? Ведь если Блэк не будет иметь ориентиров, воображаемая ситуация невозможна.

В исходной статье «Альтернативные возможности и моральная ответственность» Франкфурт предлагает несколько вариантов того, как Блэк

может влиять на выбор Джонса: это и угрозы, и специальное снабдьё, и гипноз, и радиоуправляемые электроды, воздействующие на нейронные сети Джонса. Но он почти ничего не говорит о том, на какой сигнал Блэк должен ориентироваться. Франкфурт ограничивается только уверением в том, что Блэк — «превосходный эксперт в этих вопросах»<sup>131</sup>. Остается додумывать самим. Можно предположить, что наиболее вероятным ориентиром для действий Блэка будет состояние мозга Джонса, предшествующее решению. Так, предположим, что выбору в пользу действия «А» в момент  $t_2$  предшествует состояние «NA» в момент  $t_1$ . Если Блэк видит, что в момент  $t_1$  мозг Джонса находится в этом состоянии, он заключает, что Джонс сам сделает «А», и не вмешивается. Если же в  $t_1$  у Джонса — какое-то другое состояние, Блэк будет действовать. Но теперь неясно, действительно ли у Джонса нет альтернатив. Удастся ли Франкфурту действительно совместить ответственность и отсутствие альтернатив у агента?

Если принять, что в мысленном эксперименте Франкфурта Блэку необходим некоторый сигнал, чтобы действовать, необходимо принять и возможность различных сигналов. Так, Джонс, очевидно, в момент  $t_1$  может иметь состояние NA, а может и не иметь. Несмотря на то, что Джонс не может поступить иначе, он может оказаться в том или другом предшествующем состоянии. Назовем это «проблесками свободы». Сторонник альтернативных возможностей как условия моральной ответственности может теперь зацепиться за это. И настаивать на том, что именно «проблески свободы» в примерах Франкфурта обосновывают интуиции о моральной ответственности агента.

Но, думаю, это слишком ненадежная опора. Ведь, напомним, речь идет только о непреднамеренных неосознанных промежуточных состояниях. Эти состояния равносильны рефлексам, случайным вздрагиваниям, оговоркам,

---

<sup>131</sup> Frankfurt G. Alternate Possibilities and Moral Responsibility. // Russel P. and Deery O (eds.). The Philosophy of Free Will. Oxford, New York: Oxford University Press, 2013. P. 144.



тикам и т.п. Они не выражают осознанный выбор агента. Вряд ли в возможности подобных произвольных событий состоит основание для ответственности. Абсурдно же утверждать, что основание ответственности за убийство лежит в возможности промахнуться. Впрочем, защита «проблесков свободы» как альтернативных возможностей, на которых основывается моральная ответственность, — не единственный путь оспорить примеры в «стиле Франкфурта». Философы Р. Кейн, Д. Вайдеркер и К. Жине предлагают иную линию критики. Они считают, что в релевантных примерах Франкфурта скрыты не «проблески свободы», а реальные альтернативы.

Доказательство начинается с определения, в каком мире реализуются сценарии Франкфурта. Сначала предположим, что это — детерминистический мир. В детерминистическом мире по предшествующему сигналу Блэк может достоверно определить, нужно ли ему действовать. Но проблема в том, что детерминистический мир в представлениях инкомпатибилиста не дает возможности для моральной ответственности. В таком мире, согласно аргументу, представленному выше, агент не является «ответственным в конечном счете». Настаивать на моральной ответственности агента в этом случае — значит просто бездоказательно отрицать позицию инкомпатибилиста.

Другое дело, если мир, в котором реализуются примеры, индетерминистический. В таком мире, по мнению инкомпатибилиста-либертарианца, ответственность может существовать. Но тогда и возникает сомнение в отсутствии альтернативных возможностей для агента. Представим, что имеет место воле-устанавливающее действие. Между предварительным сигналом о состоянии агента в момент  $t_1$  и его выбором в момент  $t_2$  отсутствует детерминистическая связь. У контролера-манипулятора нет достаточного основания вмешиваться. Если он вмешивается, то агент ответственности, очевидно, не несет. Выбор за него осуществил контролер. А если контролер не вмешался, у агента в момент  $t_2$

остаётся возможность сделать тот или иной выбор. Ведь индетерминизм предполагает, что при одних и тех же исходных условиях какое-то действие случается только с определенной вероятностью. С некоторой вероятностью может случиться другое действие. Так образом, без вмешательства контролера агент имеет альтернативные возможности и несет ответственность.

*b) Решение дилеммы контролера*

Перед защитником примеров «в стиле Франкфурта» встает дилемма: либо в примере неочевидно наличие моральной ответственности агента, либо неочевидно отсутствие альтернативных возможностей. Назовем это *дилеммой контролера* в примерах в стиле Франкфурта. «Если агент ответственен в конечном счете, так как контролер не вмешался, агент мог поступить иначе; но если агент не мог поступить иначе, потому что контролер вмешался, контролер ответственен в конечном счете, а не агент»<sup>132</sup>. Без решения этой дилеммы опровержение принципа альтернативных возможностей не удастся. Чтобы ее решить, требуется модификация исходного примера Франкфурта. И такую модификацию предлагает Д. Перебум. Его пример «в стиле Франкфурта» очень изобретателен. Но автор убеждает, что взят этот пример частично из собственного опыта.

*Пример 3. Уклонение от налога*<sup>133</sup>. Джо размышляет о том, воспользоваться ли ему налоговой льготой в связи с дорогостоящей покупкой. Он понимает, что в данном случае льгота не предусмотрена законом. Но, скорее всего, его ходатайство никто не оспорит. В любом случае он сможет оправдаться тем, что плохо разбирается в налоговом законодательстве. У Джо есть сильное желание воспользоваться льготой и не платить налог. Принятие эгоистических решений, даже если они ведут к нарушению некоторых законов, для Джо приемлемо. Единственное, что

---

<sup>132</sup> Kane R. The Significance of Free Will. Oxford, New York: Oxford University Press, 1996. P. 142.

<sup>133</sup> Pereboom D. Free Will, Agency, and Meaning in Life. Oxford, New York: Oxford University Press, 2014. P. 15.

могло бы привести Джо к решению не пользоваться льготой и заплатить все налоги, был бы определенный уровень внимания к моральным доводам.

Определенный уровень внимания к моральным доводам есть **необходимое, но не достаточное** условие выбора в пользу уплаты налога. Если Джо достигает определенного уровня внимания к моральным доводам, он, обладая либертарианской свободой воли, может сделать выбор как в пользу уплаты налога, так и в пользу уклонения от него. Блэк хочет, чтобы Джо уклонился от уплаты налогов. Поэтому втайне от Джо он установил в его мозг устройство, которое в случае достижения определенного уровня внимания к моральным доводам автоматически вынудит Джо выбрать уклонение от налогов. Джо размышляет над решением, он не достигает необходимого уровня внимания к моральным доводам и выбирает уклонение от налогов. Устройство остается бездейственным. Вероятно, в такой ситуации Джо несет полную ответственность за свой выбор.

Рассмотрим схему на Рис. 4, которая отражает обстановку мысленного эксперимента Перебума. Ромбиками обозначены состояния агента. Стрелками с пунктиром — индетерминированные связи, а сплошными линиями — детерминированные. Красным цветом выделен случай интервенции контролера.

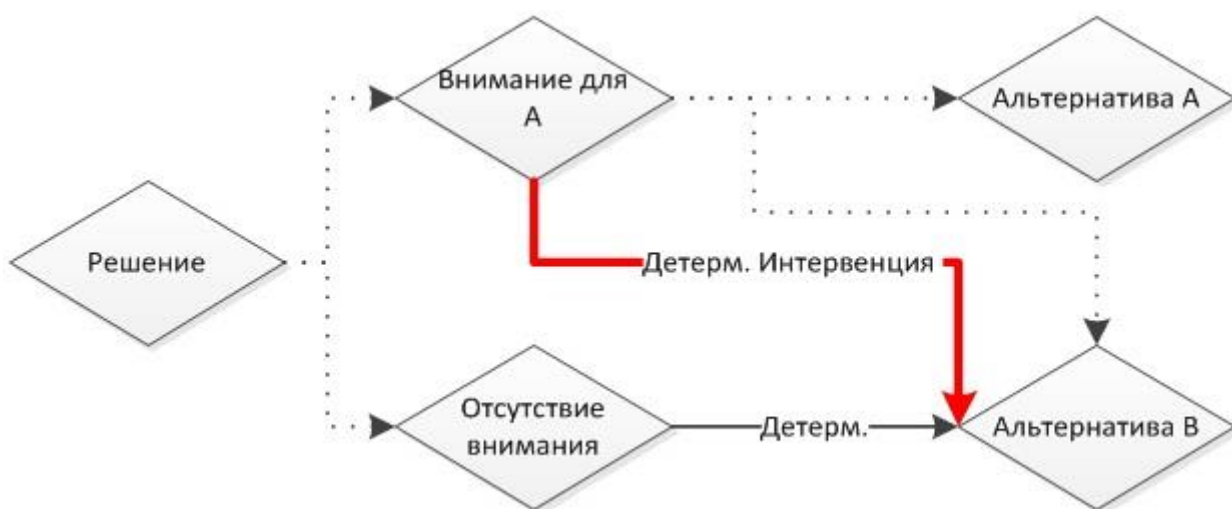


Рис. 4. Схема примера Д. Перебума «Уклонение от налога»

Согласно описанной ситуации, агент находится в мире, где есть индетерминированные события. В силу неопределенности агенту доступны две альтернативы: сконцентрироваться на моральных доводах или нет. Концентрация внимания является необходимым условием решения платить все налоги. Но это — не достаточное условие. После концентрации агент в силу неопределенности все еще может принять любое из двух решений. Однако для контролера это — уже достаточное основание вмешаться, ведь на этом этапе возникает риск, что Джо все-таки решит платить налог. Зато отсутствие концентрации — верный признак того, что все само пойдет как надо Блэку и ему не придется вмешиваться. В эксперименте есть надежный сигнал для вмешательства контролера и неопределенность, необходимая для либертарианства. Такова обстановка мысленного эксперимента.

Теперь мы можем протестировать работу мысленного эксперимента. Проиграем воображаемую ситуацию. Джо задумывается о выборе. Через некоторое время он начинает концентрироваться на моральных доводах. Тут же вмешивается Блэк. Он заставляет Джо принять решение в пользу налогового вычета. Ответственности у Джо нет, и нет альтернатив. Воспроизведем эксперимент еще раз. Джо опять концентрируется на моральных доводах, и опять Блэк вмешивается, заставляя последнего делать все тот же эгоистический выбор. У Джо просто нет шансов принять иное решение. Сколько бы раз эксперимент ни развертывался по этому сценарию, итог будет одинаковым. А что будет, если реализуется другой сценарий? Предположим, Джо опускает моральные соображения и сам выбирает уклонение от налогов. Тогда Блэк остается бездействовать. Решение Джо индетерминировано, и, следовательно, даже по мнению либертарианца это будет ответственное решение. При этом у Джо нет альтернативы выплатить налог. Значит ли это, что *дилемма контролера* преодолена?

У противника франкфуртских примеров есть еще один доступный шаг — показать, что у Джо все-таки есть какие-то альтернативы. Хотя Джо не

может не уклониться от налога, ему все же доступны разные варианты поведения. Он может сконцентрироваться или не сконцентрироваться на моральных соображениях. Это и есть альтернативы Джо. Противник франкфурстских примеров будет доказывать, что именно наличие этих альтернатив составляет основание ответственности. Если это так, то обстановка мысленного эксперимента не соответствует задачам. Она не позволяет показать совместимость ответственности и отсутствия альтернатив. Перебум с этим выводом не согласен. Он доказывает, что альтернативы, доступные Джо в его обновленном примере, не являются *релевантными и существенными* (robust), чтобы служить основанием ответственности. Они — не более чем «проблески свободы». Они не удовлетворяют критериям существенных релевантных альтернатив. Чтобы альтернатива была существенной и релевантной, она, во-первых, должна быть осознанной. Ведь абсурдно утверждать, что основание ответственности за поступок лежит в возможности случайно его не совершить. Агент должен желать совершить поступок. Во-вторых, альтернативный поступок должен вести к другим последствиям: в результате выбора этой второй альтернативы агент не должен нести того же морального бремени, что и в случае выбора первой. Иначе в чем заключается ее альтернативность, если она ведет к тем же последствиям? В-третьих, релевантная и существенная альтернатива должна удовлетворять эпистемическим критериям. Агент должен хотя бы отчасти понимать, что выбор в пользу второй альтернативы избавит его от той моральной ответственности, которая наступит в первом случае.

Перебум приводит пример, который иллюстрирует работу этих трех критериев. Назовем его «Кофе налогоплательщика». Предположим, Джо может заплатить или не платить налоги. Единственным случаем, когда Джо платит налог, может быть случай, когда он перед принятием этого решения выпивает чашку кофе. В чашку без его ведома добавлено вещество, заставляющее его действовать в соответствии с налоговым кодексом. Во всех

остальных случаях Джо не платит налогов. Итак, Джо не пьет кофе, и не платит налог. Является ли единственная имевшаяся альтернатива основанием считать, что Джо ответственен за свое фактическое действие? Альтернатива Джо представляла (1) сознательное действие, (2) действие, ведущее к соблюдению Джо закона. Джо мог выпить кофе осознанно. И результатом этого действия было бы выполнение налоговых требований. Но это была бы не релевантная и не существенная альтернатива. Ведь Джо не мог знать, что это действие приведет к другому действию, к выплате налога. Эта альтернатива не удовлетворяет третьему критерию существенности и релевантности. Следовательно, она не может служить основанием для наличия ответственности.

Альтернативой такого же рода является и возможность Джо в примере «Уклонение от налогов» сконцентрироваться на моральных доводах. Джо не знал, что эти размышления могут привести его к уплате налогов и другим моральным последствиям. Поэтому концентрация внимания не является релевантной и существенной альтернативой. И, соответственно, она не является основанием для возникновения моральной ответственности. Если эти рассуждения верны, дилемма контролера решена. И из примеров в стиле Франкфурта можно сделать вывод, что основанием ответственности служит не наличие альтернатив, а способность агента быть «причиной в конечном счете». «Возможно, что альтернативные возможности — не релевантные и не существенные — является следствием того, что агент является источником действий. Но эти альтернативные возможности не играют первостепенной роли в объяснении моральной ответственности. Скорее, они могут быть только следствием первостепенного фактора: того, что агент особым образом является источником своих действий»<sup>134</sup>. Таким образом, Перебум, в отличие от Кейна, считает, что альтернативные возможности не являются существенными для ответственности и не добавляют контроля агенту. И в

---

<sup>134</sup> Pereboom D. Optimistic Skepticism about Free Will // Russell P. and Deery O. (eds.). *The Philosophy of Free Will: Selected Contemporary Readings*, New York: Oxford University Press, 2013. P. 431.

этом — одно из важных отличий теорий Кейна и Перебума. Но самое принципиальное их отличие — в том, что Перебум считает свободу воли и индетерминизм несовместимыми. В пользу этого вывода он приводит Аргумент исчезающего агента.

#### В. Аргумент исчезающего агента

Перебум считает, что его аргумент относится к семейству Аргументов случайности, о которых речь уже шла выше. Однако в отличие от классического аргумента Юма, это рассуждение нацелено на современные варианты либертарианства. В нашем тексте это — уже вторая модификация аргумента. Вывод этого рассуждения касается ситуации, парадигмальной для современных сторонников либертарианства и, в частности, для Кейна. Он опровергает наличие достаточного контроля во время воле-устанавливающих действий.

Представим, что агент делает выбор между двумя альтернативами поступков. Моральные соображения склоняют его выбрать первую альтернативу, а эгоистические соображения — вторую. Альтернативы действий исключают друг друга, а мотивы в пользу каждой из них обладают одинаковой силой для агента. Если считать, что только события имеют каузальную силу, и принять условие, что релевантные antecedentes события, в которых задействован агент, не устанавливают (settle), какое из решений будет принято, но только делают каждое из решений возможным на 50%, то ничто не устанавливает, какое решение будет принято. Соответственно, ни сам агент, ни что-либо, относящееся к агенту, не устанавливает, какое решение будет принято. Поэтому агент не имеет контроля, необходимого для моральной ответственности в смысле базовой заслуги (basic desert).

Сделаем теперь три прояснения к Аргументу исчезающего агента.

a) *Мишенью аргумента является не существование агента, а существование моральной ответственности*

Название, которое автор дал аргументу, может сбить с толку: может показаться, что, согласно рассуждению Перебума, в индетерминированной предшествующими событиями ситуации в либертарианстве событий неизбежно **исчезает агент** и что именно это исчезновение является выводом аргумента. Тогда для того, чтобы доказать, что индетерминизм ведет к отсутствию моральной ответственности, нужно было бы указать, что исчезновение агента, или исчезновение *того самого* агента, что совершил выбор, есть основание для исчезновения ответственности. Но мишенью для критики в данном аргументе служит все-таки не существование агента, а наличие у него (расширенного) контроля. Перебум считает, что в условиях, когда мотивация агента к альтернативным поступкам равносильна, **теряется именно контроль**. Поэтому правильной было бы назвать его Аргументом исчезающего контроля. «Проблема, которая выявляется аргументом, в том, что агенты либертарианцев событий не будут иметь возможности установить, какое из решений сработает, они не выполняют роли, которая обеспечивала бы контроль, требуемый этим типом моральной ответственности»<sup>135</sup>. Индетерминизм в ситуации спорного выбора подрывает контроль агента, но без контроля выбор становится случайным, а за случайный выбор ответственность не наступает, по крайней мере, в смысле *базовой заслуги*. И это — тоже важное уточнение.

b) *Аргумент критикует возможность не всякой моральной ответственности в условиях индетерминизма, а только возможность ответственности в смысле базовой заслуги*

Перебум считает, что только ответственность в смысле **базовой заслуги проблематична** в данном случае и в дебатах о свободе воли в целом. Другие типы ответственности могут быть совмещены даже с вынужденными действиями или действиями случайными. Базовая заслуга предполагает приписывание ответственности агенту исходя лишь из того, что

---

<sup>135</sup> Pereboom D. Free Will, Agency, and Meaning in Life. NY: Oxford University press, 2016. P. 32.



он совершил действие и понимал его моральное значение, без учета соображений о будущих событиях. «Чтобы агент был морально ответственным за действие в этом смысле, оно должно принадлежать ему таким образом, чтобы он заслуживал порицания, если понимал, что оно было отрицательным с моральной точки зрения, и заслуживал похвалу, если понимал, что оно было положительным с моральной точки зрения. Заслуга в данном случае является базовой в том смысле, что агент заслуживает порицание или похвалу только из-за того, что совершил этот поступок, при условии, конечно, что он понимал моральный статус поступка, а не в связи с одними только соображениями консеквенциалистского или конвенционального характера. Такая характеристика оставляет возможность считать агента морально ответственным [в другом смысле — Д.В.] даже, если он не заслуживает порицания или похвалы...»<sup>136</sup>. Только ответственность в смысле базовой заслуги в случаях индетерминации ставится под сомнение с помощью Аргумента исчезающего агента. Более того, такая ответственность, по мнению Перебума, не может существовать только в *либертарианстве событий* (event-causal libertarianism).

с) По замыслу Перебума, аргумент критикует только либертарианство событий

Среди современных каузальных либертарианских теорий можно выделить как минимум два типа. Они различаются по сущностям, которые теории допускают к участию в каузальных связях. К *либертарианству событий* относятся теории, в которых каузальные связи существуют только между событиями. Согласно этим теориям, действия, анализируемые как события, в которых задействован агент, в качестве вероятностной недетерминирующей причины имеют желания и убеждения, которые также анализируются как события, в которых задействован агент. К сторонникам этого типа либертарианства относятся Кейн, Балагер, Франклин. Ко второму

---

<sup>136</sup> Pereboom D. Free Will, Agency, and Meaning in Life. NY: Oxford University press, 2016. P. 2.

типу, *агентскому либертарианству*, относятся теории, в которых каузальные связи могут существовать не только между событиями, но и между субстанциями. Согласно агентскому либертарианству, свобода воли, необходимая для моральной ответственности (в смысле базовой заслуги) обеспечивается каузальной связью между агентами, которые несводимы к событиям, и действиями. При этом сами агенты каузально индетерминированны, то есть обладают свободной волей в либертарианском смысле. К сторонникам этого типа либертарианства относятся О'Коннор, Кларк, Гриффит.

Оба типа либертарианства предполагают необходимость расширенного контроля для ответственности в смысле базовой заслуги. В обоих типах либертарианства такой расширенный контроль обеспечивается индетерминизмом. Но агентское либертарианство, по мнению Перебума, не может быть мишенью для Аргумента исчезающего агента. Этот аргумент критикует наличие расширенного контроля только в теориях **либертарианства событий**, и в первую очередь — в специальных случаях, когда силы конфликтующих мотивов агента находятся в равновесии. Аргумент по замыслу обнаруживает внутреннее противоречие этого либертарианства в данных случаях. Это — противоречие между утверждением необходимости расширенного контроля агента над собственными действиями и фактическим отсутствием такого контроля в индетерминистических ситуациях в связи со случайностью выбора.

Из уточнений, сделанных выше, может сложиться впечатление, что аргумент Перебума имеет узкую направленность: его вывод относится к одному типу случаев, он опровергает только частный вариант либертарианства и критикует возможность моральной ответственности только в смысле базовой заслуги. Но это не так. Во-первых, случаи равновесных конфликтующих мотивов являются парадигмальными для либертарианцев. Вероятней всего, выводы, относящиеся к этим случаям, могут быть распространены на все остальные случаи вероятностной

каузальности, то есть на все теории, основывающие свободу в индетерминизме. Во-вторых, либертарианство событий, которое в этом аргументе критикует Перебум, является, по мнению большинства, самым перспективным вариантом либертарианства. Агентское либертарианство многим кажется слишком экстравагантным. В-третьих, возможно, что аргумент, использованный в данном случае против либертарианства событий, может также быть использован против агентского либертарианства. Таким образом, Аргумент исчезающего агента представляет проблему для совместимости индетерминизма со свободой воли в целом. Но, по мнению Перебума, у детерминизма не больше перспектив.

## **2. Свобода воли и детерминизм несовместимы: Аргумент манипуляций**

Перебум, как и Кейн, считает, что детерминизм несовместим со свободой воли. Но основания этого тезиса у двух философов отличаются. Кейн, напомним, считает, что моральная ответственность предполагает и наличие альтернативных возможностей, и то, что агент должен обладать «ответственностью в конечном счете», быть единственным источником своих действий. А детерминизм несовместим с двумя этими условиями. Перебум делает вывод о несовместимости детерминизма и моральной ответственности лишь на основании несовместимости детерминизма с условием «ответственности в конечном счете». Он видит несовместимость только «по источнику». Его вывод основывается на «Аргументе манипуляций»<sup>137</sup>. Этот аргумент является разновидностью Классического аргумента в пользу инкомпатибилизма. Он построен на интуиции об ответственности и свободе в случае манипуляции. Согласно этой интуиции, если действия агента являются результатом манипулирования, то агент не несет ответственности за действия. Но философ считает, что манипулирование и детерминация сходны между собой. А это значит, что как в случае с манипуляцией, так и в случаях с каузальной детерминацией

---

<sup>137</sup> Одно из изложений этого аргумента можно найти в книге Д. Перебума «Свобода воли, агентность и смысл жизни». Pereboom D. *Free Will, Agency, and Meaning in Life*. Oxford, New York: Oxford University Press, 2014. P. 74.

поведение агента лишено необходимого контроля и, следовательно, моральной ответственности.

Перебум начинает исследовать интуиции с умозраительного, но очевидного, по его мнению, случая, где агент является марионеткой в руках ученых. В этой гипотетической ситуации ученые управляют каждым действием агента, и потому агент, кажется, не несет ответственности. Далее Перебум приводит другие случаи манипуляций, все более и более близкие к реальности. Наконец, он доводит примеры до обычного случая, где агентом «манипулируют» не другие агенты, а каузальные связи и физические законы. Философ утверждает, что существенной разницы между всеми этими случаями нет и что экстраполяция интуиции на обыденные примеры должна убедить нас в том, что ответственности нет и в этих, обыкновенных условиях. Аргумент манипуляций Перебума строится на анализе четырех ситуаций. Все четыре ситуации задуманы так, чтобы в них удовлетворялись компатибилистские условия свободы, а интуиции склоняли к суждению об отсутствии моральной ответственности. Во всех ситуациях профессор Плам совершает убийство Уайта из корыстных соображений. Итак, первая ситуация — случай, где Плам является марионеткой ученых:

*Сценарий 1. Радиоуправляемый агент.* Команда нейроученых с помощью радиопередатчиков управляет мозгом и, соответственно, ментальными состояниями Плама. Перед тем, как Плам начинает размышлять о действиях, нейроученые нажимают кнопку. Это приводит к тому, что рассуждения Плама принимают крайне эгоистический характер. В результате этих рассуждений Плам решает убить Уайта. Плам не убил бы Уайта, не будь он под управлением нейроученых. Тем не менее, эгоистическое решение Плама не противоречит склонностям его характера, так как его характер в целом располагает к эгоистическим соображениям. Желание совершить убийство не противоречит более высокопорядковым желаниям Плама. Т.е. Плам желает убить Уайта и желает иметь это желание.

Желание совершить убийство основано на рациональных мотивах. И будь у него другие рациональные мотивы, он бы воздержался от преступления. Не все рассуждения Плама исключительно эгоистические. Плам также не испытывает «непреодолимых желаний». Нейроученые не внушают ему желаний такого рода<sup>138</sup>.

По замыслу Перебума, радиоуправляемый агент должен представляться читателям как агент, лишенный моральной ответственности. Ведь его действия — результат манипуляций. При этом манипулируемый Плам по основным параметрам удовлетворяет критериям моральной ответственности, выдвигаемым компатибилистами. В частности, он удовлетворяет критериям, которые выдвигают Юм, Айер, Франкфурт, Уоллас, Фишер и Равицца. Это должно продемонстрировать, что эти компатибилистские теории не предлагают достаточных условий моральной ответственности. Но могут ли вообще компатибилистские теории предоставить такие критерии? Для ответа на этот вопрос Перебум предлагает рассмотреть сходные, по его мнению, ситуации.

*Сценарий 2. Запрограммированный агент.* Плам во всех отношениях — обычный агент, кроме одного обстоятельства. В начале его жизни команда нейроученых запрограммировала его мозг так, что его рассуждения принимают крайне эгоистический характер. Неизбежным следствием этих рассуждений является решение Плама убить Уайта. Их нейрофизиологическая реализация и история состояний Плама полностью соответствуют Сценарию 1. Эгоистическое решение Плама не противоречит в целом склонностям его характера. Желание совершить убийство соответствует более высокопорядковым желаниям Плама. Желание совершить убийство основано на рациональных мотивах. И будь у него другие рациональные мотивы, он бы воздержался от преступления. Не все

---

<sup>138</sup> Pereboom D. Optimistic Skepticism About Free Will // Russell P. and Deery O. (eds.). *The Philosophy of Free Will: Selected Contemporary Readings*, New York: Oxford University Press, 2013. P. 424.

рассуждения Плама исключительно эгоистические. Плам также не испытывает «непреодолимых желаний».

Второй сценарий во многом близок первому. Все выдвинутые ранее компатибилистские критерии свободы воли удовлетворены. А интуиции, тем не менее, склоняют к отрицанию моральной ответственности. Разве повинен запрограммированный агент в преступлении? Как и в предыдущем случае, разве эта вина не должна быть переложена на злоумышленников-экспериментаторов? Разница между Сценарием 1 и Сценарием 2 — во времени, когда происходит манипуляция. Радиоуправляемым агентом ученые манипулируют в реальном времени. А запрограммированный агент изначально создан так, чтобы совершить преступление. Но эти различия в отношении моральной ответственности, по мнению Перебума, ничего не меняют. Как, впрочем, ничего не должны менять и различия между вторым и третьим Сценариями.

*Сценарий 3. Обусловленный агент.* Плам — нормальный человек, кроме одной особенности. В детстве он подвергся специальному воспитанию, которое каузально детерминировало его склонность к рассуждениям крайне эгоистического характера. Это воспитание было проведено до того, как Плам стал самостоятельным в своих рассуждениях. Неизбежным следствием эгоистических рассуждений Плама стало его решение убить Уайта. Во всех остальных характеристиках Плам в Сценарии 3 соответствует Пламу в Сценариях 1-2. Плам может руководствоваться моральными доводами, его первопорядковое желание соответствует его второпорядковым желаниям. Эгоистические склонности не противоречат в целом характеру Плама. И у него нет непреодолимых желаний.

Третий сценарий еще ближе к реальным условиям. Он вполне соответствует ситуации, которая, по представлениям некоторых утопистов, может сложиться в обществе будущего. Предположительно, он близок и реальной ситуации, в которой воспитывались мальчики из аристократических семейств в XIV и XV вв. Нет ничего удивительного в том,

что воспитание в детском возрасте формирует ценности и склонности взрослых. При таком специфическом воспитании ответственность, кажется, переходит от воспитанника к воспитателям. Т.е. и в третьем Сценарии, несмотря на удовлетворение компатибилистских критериев свободы воли, ответственность у агента отсутствует. Перебум считает, что такие интуиции должны распространиться и на четвертый, заключительный случай:

*Сценарий 4. Детерминированный агент.* Все, что происходит в мире, каузально детерминировано событиями прошлого и законами природы. Плам — это обычный человек, родившийся и воспитанный в обычных условиях. Его рассуждения часто (но не всегда) носят крайне эгоистический характер. Его решение убить Уайта происходит из склонностей его характера и формируется в ходе рациональных рассуждений. Это его желание полностью соответствует его второпорядковому желанию иметь такое первопорядковое желание. Нейрофизиологическая реализация и история состояний Плама полностью соответствуют Сценарию 1-3. Плам способен в целом регулировать свое поведение моральными принципами и не действует из непреодолимых желаний.

Если рассуждать последовательно, то и в четвертом случае моральная ответственность отсутствует. Ведь этот случай во многом аналогичен предыдущим. Сценарии подобраны таким образом, что каждый следующий только незначительно отличается от предыдущего. Перебум сначала представляет оригинальный контрпример на компатибилистские критерии свободы воли, а затем обобщает его до случая со стандартными, нормальными условиями. В процессе этого обобщения, выстраивая все новые и новые примеры, Перебум выявляет ключевую причину отсутствия моральной ответственности. Она заключается, по его мнению, в каузальной детерминации действий агента. Каузальная детерминация, действующая во всех четырех случаях, снимает ответственность с агента, т.к. она находится вне зоны его контроля. Агент не может контролировать ни предшествующие события, ни законы природы, и именно поэтому не несет ответственности за

производимые действия. Из этого следует, что не только учтенные компатибилистские теории не могут предоставить достаточных условий свободы воли и ответственности, но и любые компатибилистские теории в принципе не могут представить таких критериев. Свобода воли и моральная ответственность несовместимы с каузальной детерминацией. Компатибилизм ложен.

Аргумент Перебума можно представить в более формальном виде. Для удобства введем аббревиатуру «ККС» — компатибилистские критерии структуры свободного агента. Эта аббревиатура будет обозначать все необходимые условия, которым, в соответствии с компатибилистскими теориями, должно отвечать действие, чтобы его можно было признать свободным действием агента. В число этих критериев входят соответствие действия склонностям характера, отсутствие непреодолимых желаний, соответствие первопорядковых желаний второпорядковым желаниям, согласованность желаний с рациональными мотивами агента. С использованием этих критериев, аргумент будет выглядеть так:

1. Если агентом  $S$  успешно манипулируют, чтобы он сделал  $A$ , тогда  $S$  делает  $A$  не по своей воле, и потому он не несет моральную ответственность за действие  $A$ . ( $M \rightarrow \neg R$ )
2. Манипулируемый агент, выполняющий действие  $A$  с удовлетворением ККС, не отличается в релевантном смысле от обычного каузально детерминированного агента. ( $M = D$ )
3. Следовательно, обычный агент, детерминированный и удовлетворяющий ККС, не действует по своей воле, и потому не несет моральной ответственности. ( $D \rightarrow \neg R$ )

Первая посылка кажется интуитивно верной. К тому же она иллюстрируется первым сценарием. Вторая посылка опирается на обобщение от Сценариев 1-3 к Сценарию 4. При истинности посылок, вывод тоже кажется верным. Аргумент выглядит изобретательным и правдоподобным. Означает ли это, что свобода воли и ответственность несовместимы с



детерминизмом? Более того, в совокупности с Аргументом исчезающего агента это может означать, что свобода воли и ответственность вообще несовместимы ни с детерминизмом, ни с индетерминизмом.

Я не согласен с этим выводом и собираюсь доказать, что свобода воли и ответственность совместимы как с детерминизмом, так и с индетерминизмом. Эта позиция называется *суперкомпатибилизмом*. Часть работы уже проделана — были представлены соображения против Аргумента последствий в отношении «способности поступить иначе». Я также показал способ защитить «примеры в стиле Франкфурта». Эти примеры доказывают отсутствие необходимости доступа к альтернативным возможностям для свободы и ответственности. Теперь, чтобы отстоять позицию суперкомпатибилизма, осталось показать, что Аргумент манипуляций, направленный против совместимости детерминизма и свободы, и Аргумент исчезающего агента, обращенный против совместимости индетерминизма и свободы, несостоятельны. Это я намерен сделать в заключительной, четвертой части этой главы.

#### IV. Обоснование суперкомпатибилизма

##### 1. Возражение на Аргумент манипуляций

Чтобы проверить Аргумент манипуляций нужно разъяснить некоторые моменты. Во-первых, нужно убедиться в правомерности обобщения всех случаев с манипуляциями и обыденного случая. Действительно ли сценарии в релевантном смысле сходны между собой и в особенности с четвертым сценарием? В чем это сходство, а в чем — отличия? Во-вторых, нужно убедиться, что случаи с манипуляциями действительно снимают моральную ответственность с агента. И в чем, наконец, основание интуиций о моральной ответственности? Для того чтобы все это выяснить, нужно действовать как экспериментаторы в лаборатории. Нужно провести несколько преобразований со сценариями Перебума, поочередно внося в них

изменения, и смотреть, что получится. Определим для начала, какими параметрами можно варьировать.

В сценариях Аргумента манипуляций есть несколько ключевых параметров. На первый взгляд это: (1) *тип манипуляторов*; (2) *время манипуляции*; (3) *компатибилистические критерии свободы 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. Проверим это, включая и включая тумблеры по очередности.

В Сценариях с первого по третий действия Плама контролируют *другие агенты (I)*, а в четвертом их вовсе нет. В первом Сценарии Пламом манипулируют нейроученые, которые с помощью радиосигналов управляют его мозгом. Во втором — это программисты мозга Плама, в третьем — воспитатели, обуславливающие его поведение. В четвертом манипуляторов нет, контроль, кажется, остается за законами природы. Можно предположить, что именно наличие других управляющих агентов освобождает Плама от ответственности. Другие агенты, как магнит, перетягивают моральную ответственность в свою сторону. И это и есть отвечающий за интуицию тумблер. Если это так, то аргумент Перебума обречен на неудачу с самого начала, ведь даже если ответственность действительно отсутствует в первых трех случаях, её обобщение на четвертый несостоятельно.

Проверим это: уберем агентов из первых трех сценариев, выключим тумблер «манипуляторы» и... увидим, что радикальных изменений не происходит. Если вместо ученых поставить случайную работу машины, не имеющей специальной цели, или случайное воздействие электромагнитного поля, интуиции сильно не изменятся. По мнению Перебума, эффект, отсутствие моральной ответственности, точно останется. Плам совершит убийство, но моральной ответственности за него не должен будет нести. В этом пункте можно согласиться с Перебумом. Интуиции о наличии моральной ответственности скорее всего не возникнут. Хотя в этом случае интуиции могут быть уже не так однозначны. В любом случае, очевидно, что причина не в том, что Пламом управляют другие агенты, точнее, по крайней мере, не только в этом, но и, очевидно, в чем-то другом.

Другой параметр — *время манипуляции (2)*. В первом случае Плам управляется в режиме реального времени, т.е. момент манипуляции непосредственно предшествует действию. В третьем Сценарии манипуляция осуществляется в детстве Плама, за много лет до совершения убийства. Во втором Сценарии воздействие происходит еще раньше, до рождения Плама. Четвертый сценарий предполагает «манипуляцию» в бесконечно далеком прошлом: в момент, когда законы природы были сформированы. Т.е. манипуляция может отодвигаться на сколь угодно длительное время от действия. Но имеет ли значение время манипуляции в детерминированном мире? Представим, что манипулятором в первом сценарии является бог, который создает мир, в котором через тысячелетия будет рожден и воспитан Плам. Если при этом все каузальные связи сохранятся, и история развития Плама будет совпадать с историей в первом сценарии, время манипуляции не будет, кажется, иметь значения. Плам все равно не будет нести ответственность за действия, предопределенные в далеком прошлом. Выходит, тумблер «время» не имеет принципиального значения. Хотя в этом случае у меня интуиция уже не так однозначна. В любом случае, очевидно,

что причина не в том, что Пламом управляют в режиме реального времени, точнее, по крайней мере, не только в этом, но и в чем-то другом.

Возможно, тогда дело — в *конкретных компатибилистских критериях* (3) свободного ответственного действия? Попытаемся варьировать этот параметр. Перебум использует критерии свободного ответственного действия, которые фигурируют в наиболее развитых компатибилистских теориях. Более того, он использует их в сочетании. В сумме они могли бы обеспечить необходимые и достаточные условия свободы действия. Но, кажется, они в лучшем случае обеспечивают только необходимые условия. Даже в самом первом сценарии эти условия могут быть полностью удовлетворены, но при этом моральной ответственности нет. Может, тогда стоит пополнить или изменить их список? По мнению Перебума и даже некоторых компатибилистов, пополнение ККС не изменит интуиций принципиально. Почему? Допустим, существует «лучшая компатибилистская теория», в которой учтено условие X, являющееся лучшим кандидатом на достаточное условие свободного ответственного действия. Все равно кажется возможным модифицировать все сценарии Аргумента манипуляций так, чтобы они учитывали это условие X, но чтобы при этом агент в них все равно был лишен ответственности. Просто в таком случае манипуляция будет более сложная. С этим тоже в принципе можно согласиться. Тумблер ККС не выключает лампочку.

Следующим идет тумблер «тип поступка» (4). Философия свободы воли изобилует примерами тяжелых преступлений. Первое место, пожалуй, занимают случаи с убийствами. Поэтому философские произведения выглядят как гибриды интеллектуальной прозы и жестоких второсортных детективов. Объяснение у этого простое. Философам нужны примеры, которые вызывали бы однозначную реакцию. В этом смысле убийства подходят. Но что произойдет, если убийство Уайта заменить другим действием — например, действием, заслуживающим одобрения и поощрения? Предлагаю новый сценарий с другим типом действия.

*Сценарий 5. Радиоуправляемый герой.* Команда нейроученых с помощью радиопередатчиков управляют ментальными состояниями Плама. Плам видит, что из окон соседнего дома идет дым. От соседей он узнает, что в горящем доме — 4-х летний ребенок. Перед тем, как Плам начинает думать о том, как поступить, нейроученые нажимают кнопку. Это приводит к тому, что рассуждения Плама принимают альтруистичный характер. Плам решает спасти ребенка. Он разбивает окно и, прикрываясь от огня и дыма курткой, попадает в дом. В доме горят вещи, воздух раскален и пропитан едким дымом. Плам обжигает руки и лицо. Но это его не останавливает. В спальне он находит маленького испуганного мальчика, сидящего на полу. Плам берет его на руки и выносит на улицу. Спасенный мальчик обнимает Плама, родители благодарят его и восхищены мужественным поступком. Плам отказывается принимать благодарности и скромно прощается с новыми знакомыми. Он чуть не погиб сам, спасая ребенка, но если бы аналогичный случай произошел еще раз, Плам бы поступил точно так же.

Во всех остальных вещах этот сценарий похож на Сценарий 1 с радиоуправляемым агентом. Поступок Плама не противоречит в целом склонностям его характера. Его желание не противоречит его второпорядковым желаниям и основано на рациональных мотивах — спасти ребенка. Несет ли альтруистичный Плам моральную ответственность? Заслуживает ли он похвалы, благодарности?

Кому-то, возможно, покажется, что нет. Но у меня — в большей степени другие интуиции. Мне кажется, что с этой модификацией Сценариев обнаруживается любопытная особенность. При одном и том же механизме манипуляций, оценки будут разные в отношении Сценариев с положительными и отрицательными поступками. Да, характер Плама стал более альтруистическим в результате манипуляции, но ведь он все-таки стал таковым, и Плам совершил хороший поступок, имея добрые мотивы. Если Плама вообще можно считать агентом, и этот поступок можно считать его поступком, он, с моей точки зрения, заслужил уважения и благодарности.

Подобные же интуиции разделяют и американские философы-компатибилисты С. Вульф и Д. Нелкин. Это — странная ситуация. Выходит, изменение характера поступка даже в случаях с манипуляциями влияет на признание наличия моральной ответственности? Но к какому выводу это нас толкает? К убеждению, что иммунитетом в отношении аргумента манипуляций обладают только положительные поступки и что только моральная ответственность за положительные действия совместима с каузальным детерминизмом? Или вообще, что моральная ответственность оправдана только в случае с положительными поступками? Не странно ли это?

У Перебума есть на это ответ. «Я согласен, что интуитивно более уместно хвалить Плама в этом примере [с положительным поступком — *Д.В.*], чем в исходном Сценарии... Но для меня не очевидно, что эта похвала и моральное одобрение являются признаками моральной ответственности в значении базовой заслуги...»<sup>139</sup>. Людям в принципе свойственно хвалить других за врожденный интеллект, внешнюю привлекательность, таланты. Хотя эти качества — вовсе не заслуга агентов. Хвала в данном случае, по мнению философа, — это выражение восторга, радости, позитивного отношения, а не признак моральной ответственности. Возможно, и в случае с модифицированным сценарием реакция на поступок Плама — это просто выражение радости в отношении спасенного ребенка. Другими словами Перебум считает, что этот тумблер не является искомым, что он не влияет на лампочку моральной ответственности. Это — ложноположительный результат. Допустим. Тогда переключим следующий тумблер.

Поэкспериментируем с еще одним параметром: (5) *последовательностью Сценариев*. Перебум утверждает, что все ситуации аналогичны между собой в отношении моральной ответственности. Следовательно, их можно чередовать в любом порядке. Тогда разместим их в обратной последовательности. Начнем анализ со случая самого обычного —

---

<sup>139</sup> Pereboom D. *Free Will, Agency, and Meaning in Life*. Oxford, New York: Oxford University Press, 2014. P. 102.

*Детерминированного агента*. Какие интуиции возникают: несет ли агент моральную ответственность? Многие могут решить, что в этом случае моральная ответственность существует. Такое отношение к данному случаю исходит из нормальной практики человеческих отношений. Но даже более скептическая позиция приемлема. Допустимо считать, что в ситуации с детерминированным агентом неясно, является ли этот агент свободным или нет и, соответственно, несет ли он ответственность. Теперь развернём второй случай, случай с *Обусловленным агентом*. Если тактика обобщения работает, и Сценарии не отличаются в релевантном смысле друг от друга, случай с Обусловленным агентом должен опять привести к агностическому суждению. Но если продолжать дальше, то и в отношении запрограммированного и даже манипулируемого агента будет сохраняться агностическая интуиция. Такую инверсию Аргумента манипуляций предлагает другой компатибилист, М. МакКенна. Его вывод нейтрален: «Я пытаюсь продемонстрировать, что из Сценариев... не следует, что Плам не свободен и не имеет моральной ответственности»<sup>140</sup>. Из одного только детерминизма вывести отсутствие свободы воли и ответственности нельзя. Но если тактика обобщения верна, следовательно, агностический вывод достоверен.

Однако в чём достоинство агностической позиции? Предположим, из аргумента следует, что моральная ответственность может быть, а может и не быть совместимой с детерминизмом. Тогда что? МакКенна считает, что уже это — частичная победа. Ведь задача Аргумента манипуляции — именно в том, чтобы представить основания для твердого инкомпатибилизма. А если вывод из аргумента агностический, то этих оснований нет. Аргумент перестает выполнять необходимую роль. Естественно, Перебум с этим не может согласиться.

---

<sup>140</sup> McKenna M. Hard-line Reply to Pereboom's Four-case Manipulation Argument // Philosophy and Phenomenological Research. 2008. July. Vol. LXVII, No.1. P. 153.

По его мнению, результат анализа аргумента и его модифицированных инвертированных вариантов зависит от исходных установок. Так, некоторые начинают рассуждения с позиции убежденного агностика. Сторонники этой позиции считают, что в отношении моральной ответственности неясно, совместима она с детерминизмом или нет. Но убежденные агностики не видят возможности эту позицию дальше прояснить. Для них на этом вопрос уже закрыт. Можно также начинать с позиций убежденного компатибилиста или убежденного инкомпатибилиста. Но и эти позиции не предполагают продуктивной дискуссии. Более продуктивны открытые исследовательские позиции, ищущие согласование с новыми доводами и фактами. Перебум считает, что агностический вывод из аргумента МакКенны следует только из позиции убежденного компатибилиста, закрытого для принятия новых фактов. А новые факты здесь имеют значение. Эти факты свидетельствуют о том, что основания для поведения агентов лежат за пределами самих агентов, даже за пределами их жизненных историй. В связи с этим Перебум неоднократно ссылается на Спинозу и на его объяснение иллюзии свободы: люди в опыте находят подтверждение своей уверенности в свободе, т.к. они осведомлены о своих желаниях и действиях, но не знают о причинах, их вызывающих. Именно эти новые факты, свидетельствующие о детерминирующих обстоятельствах, должны склонить чаши весов в сторону инкомпатибилизма. И это как раз должен продемонстрировать Аргумент манипуляций.

МакКенна апеллирует к компатибилистическим интуициям в обыденной жизни. И предлагает больше обращать внимание на агентские способности Плама в Сценариях. Перебум настаивает на том, что нужно сфокусироваться на внешних обстоятельствах: на манипуляторах, программистах поведения и цепочке причин. И действительно, смена фокуса влияет на восприятие Сценариев. Интуиции, как зрительные образы в случае с кубом Неккера, изменяются в зависимости от исходных установок и



акцентов. Кажется, что Аргумент манипуляций заводит в тупик. Как разрешить эту ситуацию? Пробуем переключить последний, шестой тумблер.

Это *тип детерминации (б)*. Напомню, во всех Сценариях аргумента причины с необходимостью определяют результат, т.е. для агента доступно только одно будущее. Агент и его манипуляторы существуют в мире с каузальным детерминизмом. Что произойдет, если в Сценариях ввести индетерминированные события? Появится ли интуиция о моральной ответственности? Как ни удивительно, последствий в отношении моральной ответственности, кажется, тоже не будет. Даже если манипуляторы Плама только с вероятностью определяют его действия, интуиции в отношении его ответственности не будут поколеблены. Пусть в одном случае манипуляторы управляют со стопроцентным успехом, в другом — только с некоторой вероятностью — для Плама это ничего не меняет.

Это — любопытная ситуация. Она, казалось бы, подрывает аргументацию Перебума. Ведь в конечном итоге с помощью Аргумента манипуляций он критикует компатибилизм. К тому же философ, как я уже говорил, сам утверждает, что причиной отсутствия моральной ответственности как в Сценариях 1-3, так и в четвертом, стандартом случае, является каузальная детерминация. Но вот мы убираем каузальную детерминацию, вносим индетерминизм, и ничего не меняется. Как такое может быть? На это, однако, у Перебума есть ловкий и вполне убедительный ответ. Каузальная детерминация в данных конкретных случаях действительно является причиной отсутствия моральной ответственности. Но это — только в конкретных случаях. Есть еще и более глубокая, общая причина, которая подрывает ответственность. Это — присутствие в каузальной истории событий, над которыми у агента нет контроля. Эта более общая причина объясняет отсутствие моральной ответственности в случаях с каузальной детерминацией, а также и в случаях с индетерминированными событиями. Для того чтобы понять соотношение между двумя этими причинами, Перебум приводит такую аналогию. Предположим, где-то

прорвало плотину. Инженера спрашивают, в чем причина. Инженер объясняет, что плотину прорвало, так как резервуар был наполнен более чем одним миллионом тонн воды. На это его начальник возражает: вода здесь ни при чем. Представь, что резервуар был бы наполнен более чем одним миллионом тонн нефти. Ведь тогда бы произошло то же самое? Значит, дело не в воде. Судя по всему, начальник в этом примере не понимает, что здесь нет противоречия. В конкретном случае плотина прорвалась от переполнения резервуара водой. Но она действительно прорвалась бы, даже если резервуар переполнился бы нефтью. Ведь есть еще более глубокая, общая причина прорыва: превышение силы давления на стену плотины. Прорыв плотины, как и отсутствие моральной ответственности, имеет причины разных уровней общности и глубины. Таким образом, в описанных ситуациях причиной является наличие каузальной детерминации, в других ситуациях ею мог быть индетерминизм, но в любом случае общей причиной является отсутствие необходимого контроля.

Думаю, с этим вряд ли можно согласиться. И тумблер, отвечающий за исчезновение интуиции о моральной ответственности в сценариях — это не (6) «каузальная детерминация», а (7) «личность» Плама. **Интуиция о моральной ответственности возникает тогда, когда мы представляем Плама единой целостной личностью, проживающей последовательную, достаточно полную и целенаправленную историю, и исчезает, когда этой целостности и последовательности нет.** На мысль об этом наводит уже проделанный эксперимент с типом поступка. Напомню, Перебум говорит, что благодарность в случае альтруистического поступка — это не свидетельство интуиции о наличии моральной ответственности, а просто реакция на стечение обстоятельств. Но почему нужно так считать? Мне кажется, что это именно реакция в отношении поступка Плама. И она справедлива. Одно дело — выражать радость в отношении каких-то обстоятельств, но совсем другое — реагировать, в частности с благодарностью, на поступки агента. Реакция в отношении Плама

положительная, и эта реакция направлена на него, на его поступок, а не на обстоятельства ситуации. И это — показатель наличия моральной ответственности в смысле «заслуги». Здесь я согласен с Вульф и Нелкин.

Но мне кажется неверной их интерпретация причины. Эти философы предполагают, что моральная ответственность асимметрична. Т.е. что каузальная детерминация несовместима с отрицательными реакциями, но совместима с положительными. Именно поэтому моральная ответственность в случаях с отрицательными действиями отсутствует, а в случаях с положительными — присутствует. По моему мнению, оснований для такого экстравагантного решения, утверждения асимметрии, недостаточно. Есть другое, более простое объяснение. Когда манипулируемый Плам поступает альтруистично, он переживает последствия собственных действий. Бросаясь спасать ребенка, он оказывается в горящем помещении. Он рискует задохнуться, сгореть, получить ранение, чтобы совершить альтруистический поступок. Решение согласовано с его желаниями, его характером (пусть даже «новым»), жизненной историей и последствиями. И потому он заслуживает похвалы и благодарности. Он — целостная личность, совершающая поступки и испытывающая их последствия.

Проблема с Аргументом манипуляций в том, что в сценариях Перебума все заканчивается убийством, которое совершает Плам. Но для точной установки интуиции не менее важно, что происходит потом: как реализуются его цели, как он переживает результаты своих действий. Что за личность Плам после поступка? Остается ли он той же личностью или изменяется?

Сценарий *Радиоуправляемого героя* отличается от других Сценариев тем, что читатель представляет, как агент переживает последствия своих действий. Плам рискует, спасая ребенка, и сам чуть не погибает. Это позволяет выявить, какого рода личностью он является. Во всех остальных Сценариях эти последствия скрыты. В Сценариях, описываемых Перебумом, нет явного указания на переживание последствий агентом. Нет возможности наблюдать характер агента, нет ощущения целенаправленности его

взаимодействий с внешним миром и его соответствия стандарту рациональности. Предполагаю, что интуиции в отношении ответственности Плама-убийцы, могут пошатнуться, если подробней описать его дальнейшее поведение и обстоятельства жизни.

*Продолжение к Сценарию 1. Закоренелый радиоуправляемый преступник:* В результате сдвига системы ценностей под воздействием манипуляции Плам совершает преступление. Он сначала похищает Уайта, пытается его, заставляя подписать бумаги, а потом жестоко убивает. Тело Уайта Плам сбрасывает в реку. Покончив с уликами, Плам оформляет имущество Уайта на свое имя. Он переезжает в усадьбу Уайта и начинает наслаждаться жизнью. Он снимает деньги с банковского счета Уайта, тратит их на наркотики и азартные игры. А когда кто-то из родственников Уайта начинает догадываться об убийстве, Плам совершает еще одно преступление. Ему не нужны свидетели и конкуренты. Это происходит, конечно, уже без каких-то дополнительных манипуляций со стороны нейроученых. Они ведь еще до первого убийства сдвинули его шкалу ценностей в сторону крайней эгоистичности. Дальше все разворачивается как естественное продолжение.

Неужели и теперь у нас должно сохраниться ощущение невинности Плама? Да, Плам в некотором смысле — жертва обстоятельств. Но он уже стал отвратительной личностью и продолжает ею оставаться. Его поступки сообразны его личным качествам, его характеру и ценностям. Их непосредственно относят к Пламу. Разве по справедливости он не должен отвечать за преступления? Другое дело, если бы после убийства Уайта Плам оказался морально раздавленным. Если бы он винил себя, отчаянно пытался разобраться, как все это произошло, и отправился бы с признанием в полицию. В таком случае, зная об особых обстоятельствах, мы оценили бы поступок как результат манипуляции, как случайность, как ситуацию, «в которой Плам был сам не свой». Тогда бы в исторической целостности Плама был перебой. И тогда бы мы лишили Плама моральной

ответственности за содеянное. В общем, описания ситуаций Перебумом в целом не дают достаточно информации о том, какой личностью после поступка остается Плам. Остается ли он после манипуляции все таким же эгоистичным, беспринципным, жестоким, корыстным? Или после манипуляции он резко меняется, горько сожалеет, корит себя за внезапный порыв жестокости? Мы также не знаем, как действуют дальше нейроученые: продолжают ли они свои манипуляции или довольствуются первым изменением. Если с их стороны больше воздействий нет, и Плам сохраняет целостность личности и когерентность личностной истории, моральная ответственность, по моему мнению, сохраняется. А если воздействия продолжаются, или Плам резко меняется сам, возвращаясь к исходной системе ценностей, моральной ответственности быть не должно.

Таким образом, из этих рассуждений можно сделать следующий вывод: Аргумент манипуляций внушает интуиции об отсутствии моральной ответственности не из-за детерминизма, а из-за информационных лакун, допущенных в мысленном эксперименте. Именно эти лакуны 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.

## 2. Возражение на Аргумент исчезающего агента

С моей точки зрения не обладает убедительностью и Аргумент исчезающего агента. Проблемы в этом рассуждении обнаруживают оппоненты Перебума: А. Миле и К. Франклин. Их возражения мне кажутся очень полезными, однако ниже я предложу свой вариант критики, основанный на одной из возможных интерпретаций понятия личности. Этот вариант мне представляется наиболее перспективным.

А. А. Миле

- а) Аргумент против либертарианства событий действует также против агентского либертарианства

Перебум утверждает, что его аргумент направлен только против либертарианства событий. «Мои возражения в адрес агентского либертарианства важным образом отличаются [от возражений в адрес либертарианства событий — *Д.В.*]. Я утверждаю, что внутренняя противоречивость этой позиции еще не показана...»<sup>141</sup>. Таким образом, автор утверждает, что Аргумент исчезающего агента не может опровергнуть агентское либертарианство. Это связано с тем, что в агентском либертарианстве, в отличие от либертарианства событий, агент, по мнению Перебума, **может устанавливать**, какое решение будет принято, так как обладает расширенным контролем. Расширенный контроль появляется у агента в силу субстанциональной каузальности. Даже несмотря на то, что события, предшествующие решению, в обстоятельствах индетерминизма не способны установить, какое решение будет принято, агент с помощью агентской каузальности это сделать **может**. Перебум указывает на это преимущество: «То, что необходимо добавить в либертарианство событий — это участие агента в принятии решений, которое бы смогло увеличить его контроль, чтобы он мог установить, какое решение будет принято... Агентское либертарианство удовлетворяет этому требованию, представляя агента как причину, но не через события, относящиеся к нему, а фундаментально, как субстанцию»<sup>142</sup>. Перебум утверждает, что в либертарианстве событий заложено внутреннее противоречие: как и любая инкомпатибилистская теория, либертарианство событий требует наличия расширенного контроля для моральной ответственности, но в рамках этой теории такого контроля быть не может. А в агентском либертарианстве этого недостатка нет. Так ли это в действительности?

<sup>141</sup> Pereboom D. Free Will, Agency, and Meaning in Life. NY: Oxford University Press, 2016. P. 50.

<sup>142</sup> Pereboom D. Free Will, Agency, and Meaning in Life. NY: Oxford University Press, 2016. P. 51.

Один из давних оппонентов Перебума, А. Миле, оспаривает это утверждение. Философ считает, что если проблема случайности выбора есть у либертарианцев событий, она есть и у сторонников агентского либертарианства<sup>143</sup>. Миле предлагает свой анализ ситуации спорного решения уже в системе агентского либертарианства. Предположим, — пишет Миле, агент каузально определяет решение. Но в ситуации спорного выбора (например, в описанных выше Случаях Кейна «Ограбление» или «Выбор Ральфа»), где каждое из альтернативных решений может быть принято с вероятностью 0,5, в половине случаев агент сделает выбор в пользу альтернативы А, а в другой половине случаев — в пользу альтернативы Б. На чем будет основываться различие в выборе?

Ментальные состояния агента по определению одинаковы, мотивы — тоже. Ничто не отличается в ситуациях с разным исходом. Что тогда определяет, какой выбор сделает агент? Сам агент? Но чтобы выбор был различным, различия должны быть в самом агенте, а их по условиям нет. Кажется, ничто не может объяснить разницу в исходах. Но если нет ничего, что могло бы объяснить эту разницу, выбор представляется случайным. Следовательно, если Аргумент исчезающего агента опасен для либертарианства событий, он опасен и для агентского либертарианства. Агентское либертарианство, видимо, тоже не дает расширенного контроля. Но есть и другая интерпретация: мы просто не очень хорошо понимаем теорию агентского либертарианства — в ней есть скрытые аксиомы.

*b) Агентское либертарианство предполагает полный контроль над выбором, но полный контроль над индетерминистическим событием невозможен*

В целом кажется, что агентское либертарианство — очень смутная теория. Об этом говорят даже сами либертарианцы, такие как Р. Кейн. Тем не менее, в качестве гипотезы можно представить, как это готов сделать Перебум, что в агентском либертарианстве агентская каузальность не

---

<sup>143</sup> Mele A. On Pereboom's Disappearing Agent Argument // Criminal Law and philosophy, 2015.

допускает случайности. Тогда, видимо, агентская каузальность должна давать агенту полный контроль над исходом решений. Для того чтобы решение не было случайным, чтобы агент был свободен в той степени, что необходима для моральной ответственности, у него должен быть полный контроль над исходом выбора. Но не слишком ли это высокие требования? Неужели мы всегда ожидаем, что агент обладает полным контролем, когда приписываем ему ответственность?

Предположим, профессиональный баскетболист имеет 90% шанс забросить штрафной. У него есть большой контроль над тем, чтобы забросить мяч, — продолжает Миле. «Утверждение, что, несмотря на это, он не может свободно забросить любой из штрафных мячей потому, что у него нет полного контроля... устанавливает слишком высокую планку...»<sup>144</sup>. Возможно, в случае с либертарианским агентом полный контроль нужен не над исходом выбора, а над попыткой сделать выбор? Тогда представим другой сценарий: баскетболист имеет 90% шанс *сделать попытку* забросить мяч и 10% шанс психологического срыва — в его голове установлен индетерминистический генератор случайных чисел с такими вероятностями исхода. Вот он забрасывает мяч. Можно ли сказать, что он забросил этот мяч совершенно случайно? Миле считает, что и в этом случае требование полного контроля — чрезвычайно строгое требование. К тому же философ видит проблему в представлении о полном контроле индетерминистического выбора.

Полный контроль над выбором А, видимо (по аналогии с полным контролем над действием), предполагает, что у агента есть намерение сделать выбор А, что он пытается сделать выбор А и не может не преуспеть, в результате делая выбор А. Но в действительности агент почти никогда не имеет «намерения сделать выбор А» и не «пытается сделать выбор А». Агент сразу делает выбор А. Это более экономичный вариант мышления. Было бы странно, если бы наличие моральной ответственности предполагало бы такой

---

<sup>144</sup> Mele A. On Pereboom's Disappearing Agent Argument // Criminal Law and philosophy, 2015.



неэффективный способ мышления. Кроме того, полный контроль у агента должен быть над выбором, который при этом должен быть индетерминистическим, то есть который может быть не сделан. В этой ситуации невозможно определить, что такое полный контроль или, скорее, он просто невозможен.

*В. К. Франклин: Усовершенствованный вариант либертарианства событий позволяет устанавливать выбор*

Аргумент исчезающего агента обнаруживает потенциальные проблемы в либертарианстве событий, — соглашается другой комментатор Перебума К. Франклин. Этот философ тоже считает, что теории многих современных либертарианцев, в частности М. Балагера и Р. Кейна, оставляют выбор агента в условиях индетерминизма случайным. Но он уверен, что эти теории можно усовершенствовать. «Либертарианцы событий не могут предоставить удовлетворительную теорию свободы воли, просто дополняя каузальную теорию действия требованием индетерминизма. [Редуктивный анализ свободы воли — Д.В.] они должны дополнить редуктивным анализом способности *самоопределения*»<sup>145</sup>. Это дополнение необходимо, чтобы решить проблему с исчезновением ответственности и свободы. Под *способностью самоопределения* Франклин понимает возможность агента устанавливать или разрешать какой-то спор внутри себя. Способность самоопределения позволяет агенту управлять своими мотивами и делать выбор в пользу какой-то одной альтернативы действий. Эту способность Франклин (вслед за другими теоретиками свободы воли, Уотсоном, Шумейкером) считает необходимым условием свободы агента. К чему и каким образом можно её редуцировать?

Поскольку концепция самоопределяющегося агента является концепцией функциональной, Франклин считает необходимым применение функциональной редукции. Функциональная редукция в данном случае

---

<sup>145</sup> Franklin C. Event-Causal Libertarianism, Functional Reduction and the Disappearing Agent Argument // Philosophical studies. 2014. Vol. 170. Issue 3. P. 413–432.

заключается в сведении активности самоопределяющегося агента к состояниям и событиями, которые играют его функциональную роль, а также могут быть идентифицированы с агентом. Чтобы агент мог быть идентифицирован с состояниями и событиями, они должны выразить авторитет агента, «и когда у этих состояний есть авторитет агента, выполнение этими состояниями функциональной роли агента может считаться выполнением *им* функциональной роли»<sup>146</sup>. По мнению Франклина, агент в большей степени может быть идентифицирован со своей рациональностью. В этом Франклин следует многовековой философской традиции. Исходя из этого, состоянием, которое может быть идентифицировано с агентом в конкретный момент, Франклин считает состояние **«желания действовать в соответствии со своими самыми здравыми соображениями»** (desire to act for what the agent takes to be his strongest reasons). Это желание свойственно только самоопределяющимся агентам — более простые агенты, животные или младенцы, не обладают такими желаниями. И, соответственно, оно не может быть отчуждено от самоопределяющегося агента без того, чтобы он перестал быть таковым. Впрочем, даже если усомниться, что именно это состояние является сущностным для агента, можно предположить, что другие сходные состояния играют функциональную роль агента. Например, таким состоянием может быть желание действовать в соответствии с самыми сильными чувствами или заботами. Франклин предполагает, что состояния, выражающие самоопределяющегося агента, могут иметь *множественную реализацию*, то есть у разных агентов разные состояния могут играть роль функции самоопределения. В любом случае наличие такого состояния в дополнение к основным мотивам важным образом меняет обстоятельства выбора. Это можно обнаружить, если применить описанную выше концептуальную схему, например, к ситуации с Выбором Ральфа.

---

<sup>146</sup> Franklin C. Event-Causal Libertarianism, Functional Reduction and the Disappearing Agent Argument // Philosophical studies. 2014, Volume 170, Issue 3. P. 420.

Представим, что в результате размышления над собственными мотивами Ральф приходит к выводу, что его амбиции стать звездой в Нью-Йорке являются **самыми здоровыми соображениями**. Ральф решает действовать в соответствии со своими самыми здоровыми соображениями — он совершает самоопределяющее действие. В таком случае мотивационная сила амбиций Ральфа дополняется мотивационной силой действовать в соответствии с самыми здоровыми соображениями. Это — **новый каузальный фактор**, определяющий действия агента. Это желание можно идентифицировать с самим агентом, оно представляет самого агента. В нем агент реализует каузальную роль в добавление к каузальной роли своих желаний. Это каузальное влияние устанавливает, в дополнение к базовым мотивам, выбор. Соответственно, по мнению Франклина, этот выбор и последующие действия не являются случайным при общем сохранении индетерминизма в результате выбора. Таким образом, как считает философ, недостаток, выявленный Перебумом, устранен, а Аргумент исчезающего агента опровергнут.

*С. Д. Волков: Личность как пердурантная сущность устанавливает выбор*

Частично мое мнение совпадает с мнением критиков Перебума. Так, я согласен с Миле в том, что индетерминистические обстоятельства в интерпретации Перебума приводят к дилемме: либо в индетерминистических ситуациях для агентской каузальности тоже возникает случайность выбора, либо агентская каузальность внутренне противоречива. С одной стороны, если не допускать наличия полного контроля у субстанционального агента в агентском либертарианстве, выбор в спорных ситуациях всё равно остается случайным. Добавление агента в качестве вспомогательного каузального фактора не позволяет объяснить, что обуславливает именно этот, а не другой выбор. С другой стороны, в принципе полный контроль задает слишком высокую планку для условий моральной ответственности. А в условиях индетерминизма полный контроль вообще, кажется, невозможен.

Я также отчасти согласен с Франклином. Существует возможность усовершенствовать либертарианскую позицию так, чтобы обойти Аргумент исчезающего агента. Даже если либертарианские теории Кейна и Балагера уязвимы для аргумента Перебума, можно представить улучшенную альтернативу. Можно усовершенствовать условия даже индетерминистических ситуаций так, чтобы агент действительно устанавливал выбор и совершал ответственные действия. Для этого необходимо ввести редуктивный анализ способности самоопределения, как предлагает Франклин. В этом я тоже согласен. Однако сама редукция, проведенная Франклином, кажется мне неудачной. По моему мнению, она контринтуитивна и не решает поставленные автором задачи.

Франклин вводит дополнительный каузальный фактор, *метажелание*, которое, по его мнению, в ситуациях самоопределения должно представлять агента. Но это все равно не исключает различных исходов, ведь даже с этим фактором, учитывая индетерминистические обстоятельства, возможно принятие как одного, так и другого решения. При этом мы не можем объяснить, почему агент делает такой выбор, а не иной, а без этого объяснения выбор кажется случайным. И следовательно, агент не устанавливает, какой выбор будет сделан. Но что тогда необходимо, чтобы избежать случайности? Как нужно дополнить описание индетерминистической ситуации, чтобы агент устанавливал выбор? По моему мнению, критерием способности устанавливать выбор является наличие **контрастного объяснения, отсылающее к самому агенту**.

Контрастное объяснение — это такое описание, которое объясняет, почему в результате выбора случилось именно такое, а не другое событие. То есть отсылка к агенту и его свойствам должны объяснять, почему был сделан именно этот, а не другой выбор. Допустим, мы могли бы проиграть ситуацию с Ральфом бесконечное число раз. Удовлетворительный анализ агентского контроля должен представить объяснение, почему выбор Ральфа различался в 50% случаев. С моей точки зрения, ни агентское либертарианство, ни

анализ Франклина в данном случае не позволяют представить контрастное объяснение. Поэтому эти теории все равно не решают главной потенциальной проблемы индетерминированного выбора — случайности.

Другая общая проблема для агентского либертарианства и либертарианства Франклина — сама концепция агента. Агентское либертарианство предполагает, что агент — это субстанция. Это, пожалуй, наименее удачное решение, которое достаточно критиковалось в истории философии. Идея же Франклина об отождествлении агента с метажеланием кажется мне контринтуитивной. Даже если допустить, как предлагает Франклин, что метажелание, отождествляемое с агентом, имеет множественную реализацию, сомнительно, чтобы одна какая-то установка представляла агента. Это кажется очень узкой базой для функциональной редукции агента. Такая трактовка не дает основания для достаточно четкого различения и индивидуализации агентов и их состояний, поэтому она вряд ли подходит для данных обстоятельств. Для преодоления проблемы случайности в индетерминистических сценариях необходимо представить более состоятельный анализ концепции агента и его самоопределения в ситуации выбора. В качестве фундамента для такой концепции я предлагаю использовать разновидность нарративного подхода к идентификации личности.

Сторонники нарративной концепции считают, что личность может быть редуцируема к цельной, связанной, последовательной истории, биографическому рассказу. «Вероятно, основа нарративного подхода — в ... утверждении, что наши личности по сути своей являются нарративными структурами»<sup>147</sup>. Это утверждение разделяют философы А. Макинтайр, Ч. Тейлор, Д. Деннет, М. Шехтман. Связь свойств, событий и действий основывается на возможности непротиворечивого включения их в биографическую историю личности. То есть психологические

---

<sup>147</sup> Schechtman M. The Narrative Self // The Oxford Handbook of the Self. Oxford: Oxford University Press, 2013. P. 395.

характеристики, ментальные события и физические действия принадлежат личности и являются её частью тогда, когда они могут быть включены в её биографическую структуру. Биографический рассказ, нарратив, охватывает целую жизнь личности и распределен во времени. Его части расположены в разных временных эпизодах, как части материальных объектов расположены в различных пространственных местах. Значение и осмысленность поступки приобретают только в контексте нарратива. Вне нарратива не существует действия и рациональной личности.

Рациональная личность не присутствует целиком в одном моменте времени, в отличие от трехмерных объектов, а имеет четыре измерения: три пространственных и одно временное. Агент, взятый в какой-то отдельный момент, — только часть этой протяженной во времени сущности. Такую теорию существования в философии называют *пердурантизмом*, в противоположность эндурантизму — теории, согласно которой объект целиком представлен в один момент времени. Эндурантисты описывают, какие свойства объект имеет в прошлом, настоящем, будущем. Пердурантисты — какие свойства объект имеет безотносительно ко времени. Пердурантизм и нарративная концепция личности, по моему мнению, могут предложить надежный критерий тождества личности во времени и предоставить основание моральной ответственности. Они, в частности, позволяют решить вопрос с ответственностью в случае с индетерминистическим выбором, как он представлен в Аргументе исчезающего агента. Вот моя интерпретация случая с Ральфом.

Ральф выбирает между отъездом в Нью-Йорк и альтернативой остаться в Мейберри. У него есть мотивы в пользу обоих решений, и выбор его индетерминирован. Альтернативы действий исключают друг друга, а мотивы обладают одинаковой силой. Антецедентные события, являющиеся каузальными факторами, не могут окончательно определить исход, и если ситуация будет воспроизведена множество раз, то выбор Ральфа от случая к случаю будет различным. То есть каузальные факторы, желания и убеждения

Ральфа не устанавливают, какой выбор он сделает, а только делают каждое из решений на 50% возможным. Но, согласно предложенной интерпретации, сам агент, Ральф как пердуратная сущность, **устанавливает**, какой выбор будет сделан, так как выбор конкретной альтернативы является частью (гетерофеноменологического) нарратива о всей жизни Ральфа, включающего в себя и эпизод с этим самым выбором. Если бы была другая рациональная личность, был бы другой выбор. Поэтому Ральф имеет достаточно контроля, необходимого для моральной ответственности в смысле базовой заслуги.

Нарратив играет роль определяющего фактора по отношению к действию. Это — самоопределение, так как речь идет об определении целого своих составляющих частей. При этом в данном случае самоопределение не имеет каузального характера (как считает Франклин), ведь причины, согласно традиционной точке зрения, должны предшествовать действию. Отношение между нарративом и выбором — это, скорее, отношение *реализации*. Биографический нарратив реализуется в выборе агента, то есть состоит из множества сделанных выборов и действий. В свою очередь, нарративный агент *супервентен* на действиях и выборе. То есть невозможно различие в личности без различия в действии и выборе. Ральф, оставшийся в Мейберри, и Ральф, уехавший в Нью-Йорк — это две разные личности.

Нарративы, представляющие каждую из них, различаются и, в свою очередь, определяют выбор. Соответственно, каждая из таких личностей несет ответственность за соответствующий выбор, входящий в конкретный нарратив. Таким образом, нарративы представляют условия для контрастного объяснения выбора. А это, как я уже писал, является критерием способности самоопределения и установления выбора. Следовательно, данное дополнение (которое я бы назвал Аргументом четырехмерного агента) снимает потенциальное возражение о случайности некоторых индетерминированных действий, строящееся на основе Аргумента исчезающего агента, и возражение о невозможности моральной ответственности в условиях индетерминизма. Насколько сильна такая позиция?

Защитники аргумента Перебума могли бы попытаться доказать, что нарративный подход не является лучшей концепцией личности или что личность не является пердурантной сущностью. Это, конечно, важные вопросы. Существуют альтернативные подходы к концепции и идентификации личности: психологический, биологический, субстанциальный. Также часть философов считает, что личность не представляет собой пердурнатную сущность и что таких сущностей вообще нет. Но это — тема для отдельной дискуссии. Здесь я не берусь защищать эти тезисы. Укажу лишь на то, что среди современных философов есть сторонники предложенных мной взглядов. Достаточно допустить истинность нарративной концепции личности, чтобы признать Аргумент исчезающего агента ложным. Тем самым мяч переходит к Перебуму — это ему теперь необходимо доказать, что нарративная теория ложна, чтобы показать, что его аргумент работает.

Критики моей позиции также могли бы указать на то, что понятие «устанавливать» не подходит для характеристики связи между пердурантной сущностью и спорным выбором: во-первых, потому что пердурантная сущность не антецедентна выбору и, во-вторых, потому что каузально с ним не связана. В защиту Аргумента четырехмерного агента следует указать на то, что такая характеристика не противоречит интерпретации Перебума. В статье «Ответы Тадросу, Смилански, МакКенне и Миле» Перебум разъясняет, что синонимом глагола «устанавливать» в своем аргументе он считает слово «определять». «Устанавливать, какое действие произвести, означает определять, **не обязательно каузально**, какую из альтернатив действий выбрать», — пишет Перебум (жирный шрифт — мой)<sup>148</sup>. То есть сам автор не ограничивает возможность установления выбора только каузальными факторами. А понятие «определять» подходит для характеристики связи между пердурантной сущностью в целом и ее

---

<sup>148</sup> Pereboom D. Replies to Victor Tadros, Saul Smilansky, Michael McKenna, and Al Mele // Forthcoming in *Criminal Law and Philosophy*. 2016



темпоральной частью. Также в этой статье автор соглашается, что требование, заключающееся в том, чтобы устанавливающие факторы предшествовали выбору, излишне. «Миле подсказывает, что ошибочно утверждать, что либертарианство событий допускает, чтобы только предшествующие выбору события устанавливали, какое решение будет сделано... И я с радостью принимаю предложение Миле и снимаю требование о [возможности контроля — *Д.В.*] только над будущим»<sup>149</sup>. Таким образом, кажется, что моя интерпретация понятия «устанавливать» соответствует представлениям Перебума. А это значит, что мой аргумент не подменяет понятий и действительно бьет в цель.

Критик предлагаемого аргумента мог бы возразить, что описанные обстоятельства все равно не ведут к моральной ответственности в смысле базовой заслуги. Это связано с тем, что агент как нарративная пердурантная сущность устанавливает выбор в отношении конкретного действия, но не устанавливает нарратив в целом. То есть сам биографический нарратив является предустановленным для данного агента, и потому агент не может нести за него ответственность. Эта линия критики может быть отнесена к традиции критики с помощью Аргумента манипуляций. Как было показано выше, эта линия критики тоже несостоятельна. Таким образом, я считаю доказанным то, что свобода воли и моральная ответственность совместимы как с детерминизмом, так и с индетерминизмом, и, соответственно, суперкомпатибилизм верен. Это — важное достижение, так как суперкомпатибилизм означает, что вопросы о свободе воли и ответственности не являются эмпирическими и целиком находятся в проблемном поле философии. Возможность свободы воли и ответственности не зависит от открытий нейронауки или физики.

---

<sup>149</sup> Ibid.

## V. Итоги второй главы

В начале главы было обнаружено, что суждения о свободе агентов иногда оказываются в тупике: интуиции о наличии или отсутствии свободы расходятся. Было, однако, выявлено, что суждения о наличии свободы проясняются большинством современных аналитиков с помощью понятия ответственности. Моральная ответственность служит индикатором наличия необходимого контроля над действием у агента. Этот контроль обеспечивается в том числе наличием метафизической свободы.

Препятствием на пути реализации такого контроля и свободы оказался детерминизм. Два принципа свободы и ответственности: Принцип альтернативных возможностей и Принцип автономии — представляются плохо совместимыми с каузальным детерминизмом. Это стало очевидным из аргументов, построенных на этих принципах: классическом аргументе и Аргументе последствий. Впрочем, опасность для свободы и ответственности представляет также индетерминизм. Это было продемонстрировано с помощью Аргумента случайности, восходящего еще к Д. Юму.

Взгляды аналитических философов конца XX — начала XXI вв. на эти проблемы были показаны на примере двух авторитетных инкомпатибилистских позиций: либертарианской теории Р. Кейна и твердого инкомпатибилизма Д. Перебума. Эти позиции наиболее репрезентативно показывают доступные для современных философов альтернативы. Был проведен историко-философский и сравнительный анализ этих позиций, выявлены их общности и различия, слабые и сильные стороны. В ходе исследования было показано, что сходство этих теорий в том, что они отрицают совместимость свободы с детерминизмом. Отличием этих теорий является то, что в твердом инкомпатибилизме Перебума отрицается совместимость свободы с индетерминизмом. Перебум отрицает также необходимость альтернативных возможностей для свободы на основании

примеров «в стиле Г. Франкфурта». Кейн, со своей стороны, альтернативные возможности считает необходимыми для свободы и ответственности.

В ходе исследования была представлена критика основных аргументов против современного компатибилизма. Так, было показано, что критика Аргумента последствий является успешной. В этом аргументе значение понятия «иметь возможность» используется слишком узко. В этой главе было также продемонстрировано, что Аргумент манипуляций в формулировке Перебума вызывает противоположные интуиции. Для этого автором был предложен новый Сценарий — «Радиоуправляемый герой». В ходе работы были также представлены основания в пользу совместимости свободы, ответственности и индетерминизма. Для этого были проанализированы возражения на Аргумент исчезающего агента А. Миле и К. Франклина, а также представлен авторский довод против соображений Перебума Аргумент четырехмерного агента. Итоговым выводом главы стало утверждение суперкомпатибилизма, наиболее распространенной современной компатибилистской позиции. Из суперкомпатибилизма следует, что вопрос о свободе воли и моральной ответственности не является эмпирическим, а поиск ответа на него является собственно философской задачей.

Тем не менее, остается еще одна важная проблема: проблема существования агента. Аргумент четырехмерного агента и Сценарий «Радиоуправляемый герой» указали на то, что возможность свободы и ответственности непосредственно связана с возможностью существования личности. К анализу современных теорий личности мы приступаем в третьей, заключительной главе диссертации.

## ГЛАВА 3. Теории тождества личности и моральная ответственность

### I. Связь теории личности с проблемой моральной ответственности и свободы

Во второй главе была рассмотрена проблема совместимости свободы и ответственности с детерминизмом. В третьей главе мы собираемся изучить вопрос о совместимости свободы воли, моральной ответственности и событийного каузализма. В центре внимания главы будет вопрос существования и асинхронного тождества личности в контексте событийного каузализма. В начале главы я объясню, что следует называть событийным каузализмом и почему эта концепция важна для современной метафизики. Следующим шагом будет обозначена проблема для свободы воли и моральной ответственности, вызванная событийным каузализмом. Это проблема существования агента, а также его связи с действием — Проблема исчезающей личности (она отличается от проблемы контроля, которая обозначена в Аргументе исчезающего агента Перебума). Эта проблема признается, в частности, Д. Парфитом, Д. Шумейкером, Т. Нагелем. Далее в тексте третьей главы будут представлены подходы современных аналитических философов к решению этой проблемы.

Автором проведен анализ трех классических подходов: психологического, биологического и субстанционального. В современной аналитической философии эти подходы представлены соответственно теориями Д. Перри и С. Шумейкера, Э. Олсона, Р. Суинбёрна. В ходе исследования будет прослежено, что в традиционных подходах сохранило свою актуальность, а что утратило значение. Автором также предпринята попытка выяснить, какими аргументами пользуются современные теоретики тождества личности и какие в этих аргументах есть недостатки.

Первым будет рассмотрен психологический подход. Он является наиболее популярным. Более того, остальные подходы появляются во многом благодаря его критике. В параграфе о психологическом подходе

будет прослежена эволюция позиций от более ранних к современным. Далее будут представлены Аргумент редупликации Б. Вильямса и мысленный эксперимент «Телетранспортер» Д. Парфита. Эти рассуждения показывают уязвимость психологического подхода.

В следующем параграфе будут описаны некоторые преимущества биологического подхода Э. Олсона. Однако автором будет показано, что постановка проблемы в биологическом подходе ведет к отдельным трудностям. Далее будет проанализирован субстанциональный подход Р. Суинбёрна. Основным недостатком этого подхода, по мнению автора, является уязвимость для принципа верификации. Общий вывод об уязвимости классических подходов к проблеме тождества будет резюмирован Аргументом пяти трансплантаций. На этом заканчивается критическая часть главы. Вторая часть главы будет позитивной.

Наиболее перспективным решением проблемы тождества личности, по мнению автора, является теория личности Д. Деннета. В этой теории, по мнению автора, успешно решен вопрос о совместимости событийного каузализма, свободы и ответственности личности. Теория личности Д. Деннета относится «неклассическому» нарративному подходу. Сторонником этого подхода, помимо Д. Деннета, является М. Шехтман. Внимание к этому подходу обуславливается также следующими факторами: во-первых, нарративный подход является нетрадиционным и потому малоизученным в современной истории философии как в России<sup>150</sup>, так и за рубежом. Во-вторых, в текстах Деннета этот подход представлен в основном в виде мысленных экспериментов и художественных иллюстраций. Это вынуждает провести подробный анализ и попытаться расшифровать ключевые идеи автора. В-третьих, нарративный подход наилучшим образом позволяет

---

<sup>150</sup> Среди редких работ, посвященных этому подходу, следует упомянуть статью Левина С. Моральная ответственность и личность корпораций: нарративный подход // Мысль: Журнал Петербургского философского общества. 2016. № 21. С. 74.

объяснить, почему личность несет моральную ответственность за некоторые свои поступки. В ходе исследования будут предложены аргументы в пользу нарративного подхода и сделана попытка выявить его новизну и преимущества по отношению к другим теориям.

Исследование нарративного подхода начнется с анализа мысленного эксперимента «Где Я?» Д. Деннета. В ходе исследования будут выявлены критерии личностного, предлагаемые американским философом. Для анализа их эффективности эти критерии будут применены к одному из психологических феноменов — психопатии. Далее будет предпринята попытка выделить общие элементы в теориях сторонников нарративного подхода. В заключительной части будет представлена идея личности как центра нарративной гравитации.

Концепция личности как центра нарративной гравитации позволяет обосновать моральную ответственность личности и согласовать интуитивные и научные представления о личности. Она будет сопоставлена с примерами, используемыми Деннетом. В конце главы будут приведены психологические феномены, такие как диссоциативное расстройство личности (ДРЛ), работа разделенных полушарий мозга, а также феномен «вымышленных агентов» и их анализ в работах Д. Вегнера, М. Газзаниги и других исследователей. Анализ этих феноменов служит дополнительным доводом в пользу концепции Деннета. Этот анализ проясняет, каким образом возникает и функционирует ответственное «Я» человека. Результатом анализа будет обоснование нарративного подхода как теории личности и атрибуции ответственности за действия.

## **1. Событийный каузализм**

Каузализм — это убеждение в том, что у каждого события есть причина. Каузализм в некотором смысле является синонимом детерминизма, а также обычно выражает натуралистическое представление о природе действия. Это понятие активно использует А. Миле в статье «Каузальность,

действие и свобода воли»<sup>151</sup>. Стандартным каузализмом Миле называет теорию действия Дэвидсона, о которой мы говорили в первой главе. Событийным каузализмом я предлагаю называть теорию действия, в которой причиной действий (или любых других событий) может быть только событие. Этот вид каузализма совместим как с позицией Дэвидсона, так и позициями его оппонентов. Но он несовместим с позицией агентского каузализма, то есть с убеждением в том, что не только события, но агенты как субстанции, могут быть причинами. Современными агентскими каузалистами являются некоторые либертарианцы, такие как К. Лерер, Т. О'Коннор, Р. Кларк. Тем не менее, эта позиция является наиболее экстравагантной и не принимается большинством философов. Общепринятой теорией является, наоборот, событийный каузализм. Это подтверждает Д. Эринг в статье «Каузальные объекты отношений»: «Согласно наиболее популярному взгляду, каузальными объектами отношений являются неповторяющиеся события, в частности заключающиеся в изменениях, но не исключающие неизменные состояния»<sup>152</sup>.

Тем не менее, событийный каузализм влечет за собой проблему. Этой проблемой является Исчезновение агента. У нее — два аспекта. Во-первых, когда только события (пусть даже ментальные) являются причинами событий, остается непоясненным, какое отношение действие имеет к агенту, какую роль он выполняет в действиях. Беспокойство об этом аспекте выражает, в частности, философ сознания Т. Нагель: «Важный источник проблемы — это представление о личностях и их действиях как части природного порядка... Эта концепция... ведет к ощущению, что мы — вообще не агенты... Кажется, что нет места агентам в таком мире»<sup>153</sup>. Отсутствие каузально эффективного агента может казаться основанием

---

<sup>151</sup> Mele A. Causation, Action, and Free Will // The Oxford Handbook of Causation. Oxford, New York: Oxford University Press, 2009. P. 555-572.

<sup>152</sup> Ehring D. Causal Relata // The Oxford Handbook of Causation. Oxford: Oxford University Press, 2009. P. 367.

<sup>153</sup> Nagel T. The View From Nowhere. New York: Oxford University Press, 1986. P. 110-111.

отрицать существование агента вообще. Но если агента не существует, кто может обладать свободой и кто должен нести моральную ответственность?

Вторым аспектом проблемы является возможность длительного существования агента. Согласно общим представлениям, чтобы нести ответственность за действие, агент должен находиться в каком-то очень близком отношении к этому действию. Общепринятым считается утверждение, что агент «может нести моральную ответственность только за собственные действия». Это утверждение фиксирует Д. Шумейкер в статье «Моральная ответственность и личность»<sup>154</sup>. Но что подразумевает это убеждение? Для многих теоретиков свободы воли и моральной ответственности это убеждение означает, что «моральная ответственность предполагает тождество личности [во времени — *Д.В.*]»<sup>155</sup>. Иными словами, агент, совершивший действие в момент  $t_1$ , чтобы нести за него ответственность, должен быть тем же самым агентом в момент  $t_x$ . Но в условиях событийного каузализма возникновение и длительное существование тождественного агента проблематично. Это означает, что событийный каузализм, наряду с физикализмом и детерминизмом, представляет новую опасность для свободы и ответственности. «Чтобы прояснить ключевые компоненты моральной ответственности, ...нам нужно исследовать природу тождества личности»<sup>156</sup>.

Вопрос о моральной ответственности тесно связан с вопросом о тождества и единстве личности. Более того, я считаю, что этот первый вопрос разъясняет важный для человека практический смысл тождества.

---

<sup>154</sup> Shoemaker D. Moral Responsibility and the Self // The Oxford Handbook of the Self. Ed. Gallagher. Oxford: Oxford University Press, 2011. P. 488.

<sup>155</sup> Shoemaker D. Moral Responsibility and the Self // The Oxford Handbook of the Self. Ed. Gallagher. Oxford: Oxford University Press, 2011. P. 488.

<sup>156</sup> Shoemaker D. Moral Responsibility and the Self // The Oxford Handbook of the Self. Ed. Gallagher. Oxford: Oxford University Press, 2011. P. 488. С этим утверждением согласна, в частности, М. Секацкая: «Уйти от детерминизма и прийти к свободе невозможно без носителя свободы воли, без субъекта, свободным волеизъявлением предопределяющего собственные действия». Секацкая М.А. Этические идеалы, логические ограничения и проблема свободы // Вопросы философии. 2012. №2. С. 9. А также С. М. Левин. Физикализм, субъект и метафизическая свобода воли // Вестник ленинградского государственного университета им. А.С. Пушкина. 2013. №2. С. 67-77.



Ответственность является индикатором как для свободы воли, так и для тождества личности. С этой позицией согласен Д. Парфит. Он утверждает: «Некоторые важные вопросы (вопрос о том, что необходимо для выживания, воспоминаний и ответственности — *Д.В.*) предполагают вопрос о тождестве личности. Но они могут быть изолированы от этой связи. И когда они изолированы, вопрос о тождестве теряет значение»<sup>157</sup>. Вполне вероятно, что существуют какие-то другие смыслы тождества личности, но споры о них могут оказаться совсем пустыми. Именно тот, кто несет ответственность за мои поступки в прошлом, является тождественным мне в практически значимом смысле.

## II. Проблема единства и тождества личности

Проблема единства и тождества личности — классическая философская загадка. Она может быть обнаружена при сопоставлении обыденных представлений и философского анализа личности. Во-первых, согласно обыденным представлениям, личность — это простая, единая, конкретная и неделимая сущность. У нее есть пространственное расположение — место, в которое, как в «командный пункт», стекается вся информация из внешнего мира, поступающая через органы чувств, где она объединяется и где выносятся решения действовать. Все действия личности являются результатом решений, поступающих из этого «командного и волевого центра». Это — первая интуиция. Во-вторых, предполагается, что в обычных условиях с течением времени личность остается той же самой. Личность существует на протяжении жизни человека и в разные моменты времени это одна и та же личность. Личность может сохраниться или исчезнуть, но исчезла она или сохранилась — это объективный факт. И, наконец, в-третьих, кажется, что личность — это что-то уникальное. Может существовать только одна такая-то личность. Личность не может размножиться. Т.е. личность — это нечто простое, целое, неделимое,

---

<sup>157</sup> Parfit D. Personal Identity // The Philosophical Review. 1971. Vol. 80, No. 1. Jan. P. 4.

уникальное и дискретное. Кажется непредставимым, чтобы некоторый человек был частично «мною» и «не мною». Допустимо, чтобы мы не знали, что личность X в прошлом — это личность Y в настоящем. Но недопустимо, чтобы это было принципиально неопределимо.

С другой стороны, поиски простой и неделимой сущности кажутся обреченными. При исследовании устройства человека, выясняется, что оно отчасти напоминает часовой механизм. В нем множество важных частей, которые выполняют каждая свою функцию, и нет главной точки, единого физического центра личности. Нет «точки сбора» ментальных содержаний. О том, что *собственное Я* нельзя обнаружить, наблюдая человека через увеличительные приборы, размышлял Лейбниц еще в начале XVIII в. Он предлагал вообразить себя входящим в «машину», производящую мысль и восприятия. Во время осмотра этой «машины», — говорил он, — нельзя обнаружить ничего, кроме частей, толкающих друг друга. И трудно представить, как взаимодействие этих частей приводит к мыслям, восприятиям и решениям<sup>158</sup>.

Сегодня убеждение в отсутствии единой точки контроля и волевого управления получило дополнительные нейрофизиологические подтверждения. По мнению современных нейрофизиологов, «Я» — «это набор параллельно работающих и распределенных по разным частям мозга процессов без центрального пункта управления»<sup>159</sup>. В мозге есть тысячи мест, где принимаются решения и одновременно совершается выбор. Но в мозге, в этой биологической машине, нет духа, нет субстанционального «Я». Это — первая значительная проблема.

Трудности возникают и при попытке обнаружить длящуюся во времени часть личности. Клетки организма регулярно обновляются: появляются новые, старые отмирают. За несколько лет обновляется весь организм. Что

---

<sup>158</sup> Лейбниц Г.В. Сочинения в четырех томах. Т.1. М., 1982. С. 415.

<sup>159</sup> Gazzaniga M. Who's in Charge? Free Will and the Science of the Brain. New York, 2011. kindle. P. 107.

тогда составляет эту неизменную часть? Найти длящуюся часть личности нельзя как объективными методами, так и с помощью интроспекции. Стоит вспомнить рассуждения Юма о поиске *Я* во внутреннем опыте. «Что касается меня, то, когда я самым интимным образом вникаю в нечто, именуемое мной своим *я*, я всегда наталкиваюсь на то или иное единичное восприятие тепла или холода, света или тени, любви или ненависти, страдания или наслаждения. Я никак не могу уловить свое *я* как нечто существующее помимо восприятий»<sup>160</sup>. Восприятия сменяют друг друга, мысли и воспоминания следуют поочередно, и в них нельзя наблюдать, с точки зрения Юма, что-то, что сохраняло бы тождественность в ходе этих изменений. Как же решить эту головоломку?

Неуловимость и принципиальная невидимость *Я* становится основанием для различных заключений. Лейбниц приходит к выводу о том, что личность — простая духовная субстанция, неделимая монада. А Юм определяет *Я* как связку или пучок различных восприятий, находящихся в постоянном течении. Но ни первое, ни второе решение не является до конца удовлетворительным. Решение Лейбница отличается крайней метафизической экстравагантностью. Допущение мира духовных субстанций плохо увязывается с принципом бритвы Оккама и требованиями к научной теории. К тому же возникает почти вопрос взаимодействия духовных субстанций и материального мира, с моей точки зрения, это нерешаемый вопрос. Вдобавок к этому монадическая концепция делает проблематичным идентификацию личности.

В обыденной жизни мы не теряем ориентира, кто есть кто. И без труда различаем людей. На что мы полагаемся? Согласно «монадической» теории личности мы могли бы узнавать людей по их духовной субстанции. Но даже если бы она существовала, как возможно различить такие монады? По определению, монада не имеет физических качеств, не имеет протяженности

---

<sup>160</sup> Юм Д. Сочинения в двух томах. Т.1. М.: Мысль, 1996. С. 345.

и неспособна отражать свет, а значит — невидима. Критерий, который используется для различения людей — различия их внешности. И он является достаточно эффективными, так что, встретив «знакомое лицо», не приходится лишний раз проверять: а вдруг души поменялись друг с другом местами. Можно, конечно, попытаться выяснить, является ли это простым совпадением. Может, соотношение «одна и та же душа — одно и то же тело» проявляется с высокой степенью регулярности, и мы, привыкнув к ней, больше не утруждаем себя проверками. Да, но как вообще можно было бы помыслить такую проверку? Качества уникальности, простоты и длительности не дают нам способа идентификации ни со стороны, ни даже изнутри.

Но и альтернатива Юма не многим лучше. Представление о личности как о потоке последовательных восприятий выглядит действительно трудно совместимой с обыденными интуициями. Да и сама логическая конструкция кажется не до конца проясненной. Восприятия сменяются в постоянном потоке, а не существуют сами по себе, отдельно и изолировано. Значит, есть некоторая связь, природу которой остается объяснить. Юм сам признавался позже в неудовлетворительности своих выводов и отказывается от обозначенной позиции<sup>161</sup>. К тому же она, на мой взгляд, не согласуется с представлением о моральной ответственности. Кажется, проблема Исчезающей личности действительно актуальна. Но ее решение важно для возможности моральной ответственности. Без единой длящейся во времени личности ответственность невозможна.

### III. Современные подходы к решению вопроса о тождестве личности

Вопрос о тождественности личности во времени можно сформулировать так: каковы необходимые и достаточные условия, чтобы один человек, одна личность в момент  $t_1$  был тем же человеком, той же

---

<sup>161</sup> Юм Д. Сочинения в двух томах. Т.1. М.: Мысль, 1996. С. 379.

личностью в момент  $t_2$ ?<sup>162</sup> В этом вопросе свойство быть «тем же» может обозначать две разные характеристики: нумерическую и качественную тождественность. Когда говорят о нумерическом (метафизическом) тождестве вещей, имеют в виду то, что эти две вещи являются фактически одной, одной и той же вещью. Когда говорят о качественной тождественности, говорят, что вещи полностью схожи, абсолютно одинаковы. Но качественно тождественных вещей может быть нумерически две или больше. Две элементарные частицы могут быть абсолютно качественно тождественны, т.е. не будет способа их различить между собой, но они будут двумя отдельными частицами. При этом один и тот же человек или одна и та же вещь на разных этапах может быть качественно различной, но нумерически тождественной. Так, работающие и сломанные часы различаются между собой качественно, но могут являться одними и теми же часами в разные моменты времени. Какой из вариантов тождественности необходим для атрибуции моральной ответственности, если необходимо вообще, нам еще предстоит выяснить.

Современные аналитические философы разрабатывают три классических подхода к вопросу о единстве и тождестве личности. Это субстанциональный, психологический и биологический подходы. Психологический и биологический являются *реляционными подходами*. Они предполагают, что между личностью в момент  $t^2$  и личностью в момент  $t^1$  существует какое-то отношение. Реляционные подходы отличаются от более экзотического для философского сообщества и наиболее уязвимо для критики *субстанционального подхода*. Но именно субстанциональный подход к тождественности личности распространен больше всего за пределами профессионального дискурса. Субстанциональный подход предполагает, что тождественность между двумя фазами личности обуславливается тем, что обе фазы реализуют одну и ту же субстанцию.

---

<sup>162</sup> Аналогичным образом проблему асинхронного тождества личности формулирует Д. Парфит. Parfit D. *Reasons and Persons*. Oxford, 1986. P. 202.

Реляционные подходы являются *редукционистскими*, т.к. вопрос о тождестве личности они определяют на основе фактов о мозге, физическом теле, связи физических и ментальных событий. Нередукционистским является субстанциональный подход, т.к. вопрос о тождестве личности он ставит в зависимость от «прочих фактов», не связанных с физическими или ментальными событиями. Обычно это факты о картезианском «Я» или душе. Все описанные выше подходы представляют способы отождествления личности в различные моменты времени и направления решения обозначенной проблемы реидентификации.

### 1. Психологический подход

В аналитической философии конца XX в. — начала XXI в. доминирует психологическая теория<sup>163</sup>. Этот историко-философский факт подтверждает, в частности, М. Шехтман. Она пишет: «Последнюю половину столетия существовал значительный интерес к разработкам и защите психологических подходов к тождеству личности... Существуют десятки версий непрерывности психологических черт и множество талантливых сторонников этого общего подхода»<sup>164</sup>. Согласно психологическому подходу тождество личности определяется психологическими характеристиками. Личность продолжает существовать тогда, когда происходит наследование психологических характеристик: воспоминаний, убеждений, желаний, характера; и исчезает, когда наследников этих качеств не остается. Одним из первых сторонников этого подхода был Дж. Локк.

#### А. Дж. Локк: память как критерий тождества личности

С именем Локка связывают в целом постановку проблемы тождества личности в современной философии. В «Опыте о человеческом разумении» он формулирует проблему и предлагает собственное решение, которое

---

<sup>163</sup> В опросе, проведенном в 2009г. среди 900 англоговорящих профессоров философии, 34% утверждали, что разделяют тот или иной вариант психологического подхода, 17% — физиологический подход, 12% — субстанциональный. Остальные 37% не имели определенного мнения. Kind A. *Persons and Personal Identity*. Malden, 2015. P. 102.

<sup>164</sup> Schechtman M. *Staying Alive: Personal Identity, Practical Concerns, and the Unity of a Life*. Oxford: Oxford University Press. 2014. P. 20-21.

можно считать исходной психологической теорией тождества личности. В этой теории ключевым критерием тождества является память. «...Так что все то, что имеет сознание настоящих и прошедших действий, есть одна и та же личность, к которой относятся и те, и другие действия»<sup>165</sup>. Локк считал, что существование личности простирается в прошлое настолько, насколько прошлое может быть доступно в воспоминаниях, и в будущее настолько, насколько будет распространяться индивидуальное сознание. Этот подход обладал множеством достоинств.

Прежде всего, кажется, что именно на памяти основано представление человека о длительности собственного существования. Мы помним события прошлого и из этого, видимо, заключаем, что присутствовали в прошлом. Помимо этого, подход Локка построен на интуитивно достоверном убеждении, что сущность человека заключается в его личностных психологических качествах. «Для того чтобы выяснить, в чем заключается тождество личности, — пишет Локк, — мы должны учитывать, что такое личность; личность, я думаю, — это думающая, разумная сущность, имеющая разум и рефлексию, которая способна считать себя собой, одной и той же думающей, разумной сущностью в разное время и в разных местах. И это она делает посредством сознания, которое, по моим представлениям неотделимо от мышления...»<sup>166</sup>. Приоритетное значение психологических свойств, в частности памяти, в вопросе об идентификации подтверждается гипотетической ситуацией о принце и сапожнике.

Локк предлагает вообразить ситуацию, в которой принц и сапожник меняются воспоминаниями: все воспоминания принца передаются сапожнику, а мысли и воспоминания сапожника — принцу. Кто из них кто? — спрашивает философ. Разумное большинство сочтет, что принц оказался в теле сапожника, а сапожник — в теле принца. Из этого утверждения Локк делает вывод, что идея тождества памяти лежит в основе обыденных

---

<sup>165</sup> Локк Дж. Сочинения в 3-х т. Т. 1. Опыт о человеческом разумении. М.: Мысль, 1985. С. 393.

<sup>166</sup> Locke J. Of Identity and Diversity // Perry J. (ed.). Personal identity. P. 39.

интуиций о тождестве личности. Именно память, а не тождество, например, тела, определяет, кто есть кто.

Подтверждением приоритетности психологических критериев также является особая озабоченность многих людей существованием после физической смерти. Этому не мешает знание о том, что биологическая жизнь конечна и что тело рано или поздно перестанет существовать. Пусть человеческое тело погибает, уничтожается, — считают они, — но личность все равно может быть сохранена. Интуитивно кажется, что для продолжения существования важно сохранение сознания, мышления, воспоминаний и других психологических характеристик. Наличие идентичного тела при этом не играет существенной роли.

Идеи Локка согласуются с обыденными интуициями и в других случаях, например, в отношении поступков, совершенных в состоянии невменяемости или опьянения. Так, его психологический подход к идентификации объясняет, почему личность не должна нести полной ответственности в таких случаях. Интуитивно кажется, что пьяный человек — не тот же самый человек, что человек трезвый. И критерий доступности в памяти это вполне оправдывает. Поступки, совершенные в состоянии невменяемости или опьянения, часто недоступны памяти человека. А, значит, согласно психологическому критерию, их нельзя приписывать той же личности. Здесь вполне уместно говорить: «он был сам не свой». Пьяный или невменяемый человек, в соответствии с психологическим критерием, — не та же самая личность, что трезвый и рациональный человек, даже если они представляют одно и то же биологическое существо. А значит, на этом основании трезвого и рационального человека можно освобождать от ответственности за поступки человека невменяемого.

Описанный выше подход дает средства связать между собой различные асинхронные «фазы» существования личности и намечает путь синхронного объединения личности. Память является одним из компонентов сознания. А сознание может являться основанием для синхронного объединения



личности. Таким образом, подход Локка заслуживает внимания и перспективен. Но решение это в своем исходном варианте все-таки неверное. Уже ближайшие оппоненты обратили внимание на явные недостатки памяти как критерия тождества личности и атрибуции поступков. Дж. Батлер и Т. Рид, сторонники альтернативного, субстанционального критерия, выступили с острой критикой исходного варианта психологического подхода.

### В. Возражения к теории памяти Дж. Локка

Батлер и Рид разделяли убеждение Локка в том, что для ответственности нужна тождественность личности. В частности, Рид писал, что «Тождественность [личности — *Д.В.*]... является фундаментом всех прав и обязанностей...»<sup>167</sup>. Но оба философа были полностью не согласны с его основанием тождества. Батлер указывал, что нельзя принимать сознание за критерий тождества, так как содержание сознания постоянно изменяется. По мнению Батлера, из идей Локка следует, что никто не может отвечать за «свои» поступки в прошлом, так как никто не является тем же человеком сегодня (в плане сознания), что был вчера. А это следствие противоречит обычной практике, что является основанием для сомнений. Помимо этого, Батлер находил *логический круг* в локковских построениях. По его мнению, доступность сознанию каких-то действий в воспоминаниях уже предполагает установление тождественности личности, совершавшей действия, и потому сознание и память не могут быть основанием тождественности. Т.е. никто не может действительно *помнить* поступки, предварительно не идентифицировав как-то поступающего с самим собой.

Предположим мы хотим узнать, является ли Блотт, который существует сейчас, Клоттом, о котором мы знаем, что он существовал раньше. Из критерия памяти следует, что Блотт — это Клотт, если Блотт может в памяти иметь доступ к состояниям Клотта, которые имели место в

---

<sup>167</sup> Reid T. Essay on the Intellectual Power of Man. Boston: MIT Press, 1969. P. 112.

прошлом. Но «воспоминания» Блотта о состояниях Клотта будут только в том случае воспоминаниями, если Блотт — в действительности Клотт. Поэтому, чтобы использовать критерий памяти, нам нужно сначала идентифицировать Блотта с Клоттом. Конечно, Блотт может воображать или ошибочно думать, что он «вспоминает» состояния Клотта. Но такие воображаемые или ошибочные «воспоминания» не будут действительно воспоминаниями. Это будут иллюзии или фантазии. Следовательно, не память делает поступки собственными, а что-то еще.

Батлер считал позицию Локка недоразумением. А Рид дополнял критику яркими контрпримерами. Он обращал внимание на то, что у этой теории — странные, подозрительные следствия: так, согласно теории Локка, один и тот же человек может быть и не быть субъектом одного и того же действия. Рид демонстрирует это с помощью гипотетической ситуации с отважным офицером. Предположим, храбрый офицер в детстве воровал яблоки из чужого сада. Потом он похитил вражеский флаг во время своей первой военной кампании, а ещё позже получил звание генерала. Предположим также, что молодой офицер, когда похищал вражеский стандарт, помнил, как воровал яблоки в чужом саду, а когда получал звание генерала, еще помнил, как похищал флаг, но не помнил и не мог помнить воровства яблок в детстве. Если такая ситуация возможна, то согласно теории Локка человек, похитивший стандарт, та же личность, что ребенок, похитивший яблоки, а генерал — тот же человек, что похищал вражеский стандарт. Т.е. старый генерал — это храбрый офицер, а офицер — озорной ребенок. Но по критерию памяти старый генерал — не тот же человек, что озорной ребенок! Выходит, генерал одновременно и тот человек, и не тот человек, который воровал яблоки<sup>168</sup>.

Гипотетическая ситуация вскрывает важное отличие между отношением тождественности и памятью. Если  $x$  тождественен  $y$ , а  $y$

---

<sup>168</sup> Reid T. Of Mr. Locke's Account of our Personal Identity // Perry J. (ed.). Personal identity. Berkley: University of California press, 2008. P. 115.

тождественен  $z$ , то  $x$  должен быть тождественен  $z$ . Т.е. тождественность — это *транзитивное отношение*. Но если  $x$  помнит то, что происходило с  $y$ , а  $y$  помнит то, что происходило с  $z$ , это еще совсем не значит, что  $x$  помнит, что происходило с  $z$ . Т.е. память — не транзитивное отношение. И это явная проблема для критерия памяти, и еще не последняя проблема.

Память не позволяет идентифицировать человека с дефектами, а также здорового человека в состояниях, в которых он проводит почти половину жизни. Это состояния сна, бессознательные состояния и ситуации потери памяти. Интуитивно кажется, что тождество личности сохраняется в случае полной потери памяти. Даже в таком случае человек продолжает существование, оставаясь тем же человеком. Тем более личность остается «собой» во время сна, хотя состояний сна без сновидений человек не помнит. Однако, по локковскому критерию, человек, потерявший память, — не тот же человек, у которого были воспоминания. И во сне человек тоже становится другим, так как состояний сна он не помнит. Как можно парировать все эти возражения?

### С. Современные вариации психологического подхода

Среди современных аналитических философов сторонниками психологического критерия тождественности личности являются Э. Кинтон, П. Грайс, Дж. Перри и отчасти С. Шумейкер и Д. Парфит. Их задачей является усовершенствование идей Локка, разрешение трудностей, обнаруженных критиками. Сторонники психологической теории должны (1) преодолеть обвинения в цикличности критерия тождества, (2) решить проблему транзитивности и (3) найти способ сохранять тождество личности в случаях потери памяти и бессознательных состояний. Попытки преодолеть эти препятствия ведут к формулировке более сложных психологических критериев тождества.

Проблему (1) цикличности памяти как критерия тождества личности попытались решить С. Шумейкер и Д. Парфит<sup>169</sup>. Их предложение состояло в том, чтобы ввести новое, техническое понятие квази-памяти, или q-памяти, которое бы не предполагало тождественности помнящего и пережившего события в прошлом. Так, я могу иметь q-воспоминания о переживаниях в прошлом в том случае, если эти переживания были у *кого-то* в прошлом, и мои воспоминания каузально связаны правильным образом с этими переживаниями. Тогда обычная память будет просто подклассом q-памяти. А q-память, в свою очередь, можно использовать для теории тождества личности, избегая цикличности.

Не менее элегантное решение было найдено и для проблемы (2) транзитивности (в асинхронной ситуации). Оно состояло в объединении фаз личности с помощью пересекающихся воспоминаний или *преемственности* (continuity). В отличие от памяти, преемственность является транзитивной связью. Действительно, воспоминания о событиях далекого прошлого со временем угасают. Зато сравнительно недавние события помнятся отчетливо. Генерал помнит события, которые произошли с отважным офицером, офицер помнит события, которые произошли с озорным ребенком. И генерал может не помнить события, которые произошли с ребенком. Т.е. между генералом и ребенком *нет прямой связи через память* (*direct memory connections*). Но генерал связан с ребенком памятью через множество промежуточных фаз, между ними есть преемственность. Воспоминания генерала сейчас и в предыдущий день большей частью пересекаются, аналогичным образом они пересекались вчера и позавчера, и так — до самого детства. Например, в детстве ребенку были доступны воспоминания 1–5, во время службы офицером — с 3 до 8, во время отставки в звании генерала — с 6 до 10. Тогда между всеми фазами ребенка-генерала всегда было значительное количество пересекающихся воспоминаний (в нашем случае это 3). А это значит, что

---

<sup>169</sup> Shoemaker S. Persons and Their Pasts // American Philosophical Quarterly. 1970. Vol. 7. No. 4. P. 269-285.  
Parfit D. Reasons and Persons. Oxford, 1986.

между генералом и ребенком существует косвенная связь через память, или *преемственность памяти (continuity of memory)*. Эту преемственность и можно взять за основу тождественности. В таком случае проблема с транзитивностью и контрпримерами вроде случая с «храбрым офицером» будет решена.

Третью проблему можно попытаться обойти, введя дополнительные, более широкие критерии тождества. Помимо памяти, можно опираться на другие психологические свойства, такие как намерения, реализуемые в действиях, цели, убеждения, желания и характер. Тогда даже в условиях потери памяти, но при сохранении других психологических качеств, личность можно считать сохранившей идентичность. Суммировав все эти усовершенствования, можно получить новый *психологический критерий тождества личности*. Шумейкер предлагает такую формулировку: «X во время  $t_1$  будет той же личностью, что Y в момент  $t_2$ , если и только если X имеет уникальную непрерывную психологическую преемственность с Y, выражающуюся во множестве пересекающихся сильных психологических связей, таких как память, намерения, убеждения, желания и общности характера»<sup>170</sup>. Этот усовершенствованный критерий, кажется, выдерживает обозначенную выше критику. Но в конечном итоге даже этот новый психологический критерий оказывается неудовлетворительными.

Основной проблемой для современных психологических теорий является проблема редупликации. Мысленный эксперимент, раскрывающий эту проблему, был представлен американским философом Б. Вильямсом. Он является важным вкладом современной философии в исследование вопроса реидентификации. Можно даже утверждать, что после того, как проблема тождества личности была инициирована в философии Локка, Батлера и Юма, «контуры дискуссии вокруг проблемы были зафиксированы и оставались без

---

<sup>170</sup> Shoemaker D. Personal Identity and Ethics. // Stanford encyclopedia of philosophy; web: <http://plato.stanford.edu/entries/identity-ethics/>

изменений до конца 1956г. В этом году Бернард Вильямс опубликовал свою важнейшую статью, ...в которой он обнародовал Аргумент редупликации. Этот аргумент трансформировал последующую дискуссию проблемы и привел философов к формированию принципиально новых позиций»<sup>171</sup>.

#### **Д. Аргумент редупликации и Аргумент телепортации**

В статье «Тождество личности и индивидуализация» (1956г.) Вильямс критикует психологические теории тождества с помощью мысленного эксперимента. Автор предлагает вообразить ситуацию, в которой человек по имени Чарлз, живущий в XX веке, утверждает, что он — Гай Фокс, английский дворянин, участник заговора против английского короля в XVII в. Согласно мысленному эксперименту Вильямса, Чарлз знает в точности все детали жизни Фокса, даже неизвестные историкам. При этом все утверждения Чарлза о Фоксе находят удивительные подтверждения, которые могут быть объяснены только тем, что Чарлз действительно помнит жизнь Гая Фокса. Вильямс допускает, что кто-то мог бы интерпретировать этот случай как реинкарнацию, то есть утверждать, что Чарлз является той же личностью, что и Гай Фокс. Но автор видит, что такое суждение ведет к парадоксу.

«Если логически возможно, чтобы с Чарлзом произошли такие изменения [то есть чтобы он обрел воспоминания Гая Фокса — *Д.В.*], то логически возможно, чтобы другой человек одновременно испытал такие же изменения, т.е. например, чтобы и Чарлз, и его брат Роберт были в подобной ситуации»<sup>172</sup>. Тогда следует признать, что оба брата являются той же личностью, что и Гай Фокс. Но это утверждение, по мнению Вильямса, противоречиво. Ибо тогда получается, что Гай Фокс находится в двух местах одновременно и что оба брата как личности тождественны друг другу, а это невозможно.

---

<sup>171</sup> Noon H. Personal identity. London and New York: Routledge, 1991. P. 141.

<sup>172</sup> Williams B. Personal Identity and Individuation // Proceedings of the Aristotelian society, New series, vol. 57, 1956-1957. P. 238.

Из рассуждений Вильямса следует также, что даже когда Чарлз — единственный, кто помнит жизнь Гая Фокса, оснований считать его Гаем Фоксом тоже недостаточно. Ведь факт тождества Чарлза и Гая Фокса не может зависеть от существования или отсутствия других личностей с какими-то воспоминаниями. Тождественность является внутренним (intrinsic), а не контекстуально зависимым отношением. Таким образом, аргумент редупликации Вильямса подрывает убеждение, что воспоминания служат основанием тождественности личности. Аналогичным образом, аргумент дает основания считать любые психологические критерии недостаточными для тождества личности, ведь эти критерии также могут быть «редуплицированы» Чарлзом и Робертом.

Впрочем, защитник психологической теории мог бы возразить: Вильямс приводит любопытный сюжет, но этот сюжет не прояснен. Философ никак не объясняет, каким образом могут оба брата, Чарлз и Роберт, приобрести воспоминания Гая Фокса. Парировать это возражение можно с помощью другого мысленного эксперимента, который предлагает Д. Парфит в уже ставшей классической работе «Причины и личности» (1984). Эта фантастическая история о космическом путешествии с помощью телепортации называется «Телетранспортер». Парфит рассказывает её от первого лица:

Для путешествия с Земли на Марс мне предлагают воспользоваться новым аппаратом для телепортации. Аппарат значительно экономит время и расходы на космические путешествия. Аппарат состоит из двух частей: сканнера и репликатора. Сканнер на Земле записывает всю информацию о состоянии моего тела и мозга и передает ее на Марс. Затем все мои клетки на Земле мгновенно уничтожаются. Репликатор на Марсе, получив информацию с Земли, из нового материала воссоздает точную копию моего тела. Именно это тело продолжает мою жизненную траекторию на Марсе. Это тело обладает в точности такими же психическими характеристиками, что и уничтоженное тело на Земле. Таким образом, аппарат телепортации

обеспечивает психологическую преемственность. Обдумав доводы за и против, я соглашаюсь на телепортацию.

Я захожу в кабину для сканирования, нажимаю большую зеленую кнопку... но ничего не происходит. После короткого ожидания в кабине, я обращаюсь к диспетчеру: «У Вас сломался аппарат!» В недоумении диспетчер отвечает: «Одну минуточку, сейчас проверю». Через некоторое время он возвращается на связь: «Вы знаете, вся информация о составе вашего тела была скопирована и передана на Марс. Там уже собрана ваша точная копия. Правда... процесс уничтожения на Земле не закончился. Вместо полного уничтожения, у вас повредилась только сердечно-сосудистая система. Не беспокойтесь, через несколько дней у Вас наступит приступ и Ваша копия на Земле будет уничтожена. А на Марсе вы продолжите свою жизнь абсолютно здоровым. Если хотите, я сейчас соединю вас с вашей точной копией и, поговорив с ним, вы убедитесь, что это ваш полный психологический двойник»<sup>173</sup>.

Космическое путешествие представляет собой вариацию случая редупликации, описанного Вильямсом. Однако в отличие от исходного варианта, мысленный эксперимент Парфита обладает важным достоинством. Он хотя бы в общем виде описывает механизм, с помощью которого личность может быть раздвоена. Представляется, что такой механизм возможен хотя бы в принципе, а значит, логически возможна и описанная ситуация. При этом она, кажется, сводит психологическую теорию тождества к абсурду: у Парфита после телепортации есть двойник. Обе личности обладают психологической преемственностью, выражающейся во множестве пересекающихся сильных психологических связей, таких как память, намерения, убеждения, желания и общности характера, то есть, согласно психологическому критерию, они в равной степени тождественны Парфиту до телепортации. При этом Парфит на Земле и Парфит на Марсе — нумерически две разные личности, не тождественные друг другу. В

---

<sup>173</sup> Parfit D. Reasons and Persons. Oxford: Oxford University Press, 1984. P. 199.



формальном виде результат выглядит так:  $a$  тождественно  $b$ ,  $a$  тождественно  $v$ , при этом  $b$  не тождественно  $v$ . Это новый случай нарушения транзитивности связи (теперь в синхронной ситуации). Такое следствие является важнейшей проблемой для сторонников современных теорий тождества личности.

### **Е. Решение проблемы редупликации в современных психологических теориях**

Я могу выделить два способа преодоления проблемы редупликации, доступных для сторонников современных психологических теорий: 1) отрицание возможности мысленных экспериментов с редупликацией; 2) введение дополнительного условия тождества. Первый путь — это решение в духе Витгенштейна и Куайна. Этот путь заключается в отрицании конкретных мысленных экспериментов или метода в целом. Ситуации, которые описывают Вильямс и Парфит, являются вымышленными. Более того, в обоих случаях механизмы их реализаций могут оказаться физически неосуществимыми. Вполне возможно, что тогда на них нельзя полагаться вовсе и нельзя доверять интуициям, возникающим по отношению к ним.

Этот способ решения проблемы редупликации мне кажется наиболее слабым. Дело в том, что метод мысленных экспериментов активно используется самими сторонниками психологического подхода. Достаточно вспомнить пример с принцем и сапожником Локка. Без опоры на подобные гипотетические ситуации сложно было бы обосновать преимущества психологического подхода по отношению, например, к биологическому или субстанциональному. Кроме того, существуют более реалистичные сценарии редупликации личности, например, «расщепление» личности в результате пересадки одного полушария мозга в другое тело. В результате такой операции тоже могут появиться две личности с преемственными психологическими характеристиками, при этом они не будут тождественны друг другу. Фактическая возможность подобной операции подтверждается успешным опытом пересадки части мозга у мышей, проведенным учеными в

начале 1980-х<sup>174</sup>. Таким образом, есть основание считать, что первый путь не является перспективным для преодоления проблемы редупликации в контексте психологических теорий.

Второй способ заключается во введении дополнительного критерия тождества личности — отсутствия двойников. Согласно обновленной теории, личность Б является тождественной личности А в том случае, если Б обладает психологической преемственностью по отношению к А и нет другой личности, которая также обладала бы психологической преемственностью по отношению к А. Результатом такого нововведения, в частности, является отрицание тождества между Парфитом-героем до телепортации и двумя его преемниками. То есть, согласно новому критерию, Парфит-герой перестает существовать в случае редупликации. Это решение снимает проблему отсутствия транзитивности в психологических теориях. Но и у этого решения есть недостатки.

Введение дополнительного критерия на отсутствие «расщепления», вероятно, нарушает «Принцип только *x* и *y*»<sup>175</sup>. Согласно этому принципу тождество *x* и *y* зависит только от фактов об *x* и *y*. Как я уже говорил, тождественность является внутренним, а не контекстуально зависимым отношением. «Принцип только *x* и *y*» кажется достоверным, а его нарушение ведет к довольно любопытным следствиям. Их замечает Э. Кайнд. Она приводит остроумную вариацию сюжета о телепортации.

Представим, — пишет Кайнд, — что сломанный аппарат для телепортации создает несколько копий с небольшой задержкой. Парфит-герой заходит в камеру для телепортации на Земле, его тело уничтожается, и точная его копия возникает на Марсе. Но спустя несколько минут появляется еще одна копия. Учитывая дополнительный критерий, пока копия Парфита-героя существует в единичном экземпляре, у него есть тождественная личность. Но как только появляется второй двойник, тождественная

---

<sup>174</sup> Kind A. *Persons and Personal Identity*. Malden. 2015. P. 47.

<sup>175</sup> Kind A. *Persons and Personal Identity*. Malden. 2015. P. 61.

Парфиту-герою личность исчезает. Но как может двойной успех быть неудачей?

Мне представляется возможным еще больше усугубить абсурдность следствий второго подхода. Для этого стоит представить, что через некоторое время один из клонов погибает. Теперь опять возникает тождественная Парфиту-герою личность. Таким образом, Парфит-герой может существовать, исчезать и появляться вновь сколько угодно раз и с любыми паузами и промежуткам во времени. То есть Парфит будет «мерцающей личностью». При этом все это будет абсолютно незаметно как с позиции от первого лица, так и с позиции третьего лица. Кажется, что это действительно абсурдное следствие, то есть нельзя сказать, что второй подход полностью решает проблему редупликации. Таким образом, оба способа преодоления проблемы редупликации ошибочны. Остается искать другое решение проблемы тождества личности. Таким решением может служить биологический подход.

## 2. Биологический подход Э. Олсона

Некоторые философы отрицают, что психологические характеристики принципиально подходят для решения вопроса о тождестве личности. Это отрицание построено на альтернативном взгляде на сущность и основу идентичности человека. Среди критиков психологического подхода — Э. Олсон, Д. Деграция, Дж. Томпсон, М. Эйрс, П. Сноудон, П. ван Инваген. Многие из этих философов замечают, что жизнь человека не всегда сводится к наличию и сохранению психологических характеристик. Жизнь начинается с появления эмбриона, у которого нет развитых психических способностей, и заканчивается часто в ситуации, когда психика уже деградирует. В течение жизненного пути в результате травмы или болезни человек также может оказаться лишён каких-то когнитивных способностей: самосознания, рациональности, памяти. И не достаточно оснований вычеркивать все эти этапы жизни «Я» из автобиографии личности. Значит, есть что-то более фундаментальное, базовое в характеристиках человека, всегда

сопутствующее существованию этого «Я». Этой базой сторонники биологического подхода (или «анимализма») считают жизнь организма. Для укрепления такой позиции на основании вышеперечисленных интуиций один из самых ярких сторонников биологического подхода, Э. Олсон, разрабатывает «Аргумент вегетативного состояния» (АВС) и «Аргумент эмбриона» (АЭ).

#### **А. Аргументы Э. Олсона в пользу биологического подхода**

Представьте, — пишет Олсон<sup>176</sup>, — что в результате несчастного случая ваш организм оказывается в хроническом вегетативном состоянии (это неприятная гипотетическая ситуация, но она полезна для иллюстрации). В результате перебоя работы сердца, мозг десять минут не получает кислорода. Нейроны коры головного мозга гибнут от кислородного голодания, что приводит к перманентному отсутствию сознания. Речь, память и сложное поведение становятся принципиально невозможными. При этом сохраняются базовые функции: дыхание, пищеварение, кровообращение, метаболизм. Ведь субкортикальные части мозга, ответственные за эти функции (таламус, ствол мозга, мозжечок, базальные ганглии) более устойчивы к дефициту кислорода. Как определить такое состояние?

Описанное состояние не является смертью, так как по всем важным признакам этот человек жив. Он может умереть, если перестать его кормить. Описанное состояние не является комой или сном. Человек в вегетативном состоянии переживает периоды сна и условного бодрствования. В этом состоянии человек даже реагирует на внешние стимулы: свет, звук. У него работают рефлексy. Он моргает и кашляет. Он даже может глотать пищу. Это парадоксальное состояние бодрствования без осознанности. Как говорят медсестры: свет горит, но никого нет дома. Но кто находится в этом состоянии? У кого горит свет? Может ли такое случиться с вами? Или в

---

<sup>176</sup> Olson E. The Human Animal. Personal Identity Without Psychology. Oxford, 1997. P.7.

самом выражении «я буду находиться в вегетативном состоянии» уже есть противоречие? Будет ли такое состояние означать конец вашему существованию или, несмотря на чудовищное качество жизни, это будет все еще жизнь, продолжение жизни вашего «Я»?

Олсон склоняется к мнению, что даже такое существование — это продолжение жизни личности, хотя и чудовищного качества. Чтобы склонить к этому заключению и свою аудиторию, он предлагает сравнить случай с жизнью в вегетативном состоянии со случаем, когда после смерти друга делают памятник, точную копию человека из бронзы. Есть между этими случаями значимые отличия? — спрашивает он. Кажется, разница здесь принципиальная. В первом случае существование продолжается, а во втором — нет. Но на чем в таком случае основывается тождество личности человека в вегетативном состоянии и того, кем он был прежде? Возможны как минимум два варианта объяснений.

Во-первых, можно сказать, что между личностью до и после несчастного случая остается тождество в силу сохранения тела человека (большой его части), а во-вторых, что между ними существует тождество в силу сохранения преемственности биологических функций. Первый вариант ответа опирается на представление о том, что человек это — прежде всего тело, второй — что человек — это в первую очередь биологический организм, животное, *homo sapiens*. Первый завязан на телесный критерий, а второй — на биологический. Между этими критериями существует определенное различие, но для целей данного исследования можно объединить эти два критерия в общей биологической подход<sup>177</sup>. Исходя из этого, можно сформулировать новый общий критерий тождества личности: если *X* это личность в момент  $t_1$ , и *Y* существует в момент  $t_2$ , *X* тождественен *Y*, если и только если биологический организм *Y* является преемником

---

<sup>177</sup> Такое объединение отчасти оправдано. На это указывает Олсон в тексте «Человеческое животное». Olson E. *The Human Animal. Personal Identity Without Psychology*. Oxford, 1997. P.19.

биологического организма X. Именно такой критерий я буду использовать для характеристики биологического подхода.

Вторым важным доводом в пользу биологического подхода является Аргумент эмбриона. Он опирается на очевидный факт: существование каждого индивида, каждой личности начинается с существования эмбриона. Но как связаны эмбрион и личность? Олсон считает их нумерически тождественными и предлагает доказательство этого от противного. Он допускает, что эмбрион и личность нумерически различны, но из этого следуют одно из двух утверждений: (1) когда личность возникла, она заменила собой эмбрион в утробе матери; (2) с самого начала личность сосуществовала в одном пространстве с эмбрионом. Оба этих следствия Олсон считает абсурдными, и, следовательно, нужно сделать вывод о нумерическом тождестве эмбриона и личности. Это следствие, в свою очередь, доказывает, что психологическое тождество не необходимо для тождества личности, ведь у эмбриона нет психологических характеристик. Таким образом обосновывается анимализм.

Биологический подход, в частности подход Олсона, предполагает согласие с двумя утверждениями. Во-первых, что «мы являемся животными: точнее, видом рода Люди, или *Homo sapiens*»<sup>178</sup>. Во-вторых, что условием продолжения нашего существования является продолжение существования животного организма, а не продолжение существования личности или чего-то еще. Первое убеждение обычно не вызывает сомнения, в особенности потому, что оно не исключает возможности существования других типов личностей. «Из всего того, что я знаю, могут существовать разумные марсиане, боги, ангелы, демоны, тролли и даже рациональные, сознательные электронные компьютеры, сделанные из металла и кремниевых микросхем»<sup>179</sup>. Просто это не наш случай, считает философ, мы —

---

<sup>178</sup> Olson E. *The Human Animal. Personal Identity Without Psychology*. Oxford, 1997. P.17.

<sup>179</sup> Olson E. *The Human Animal. Personal Identity Without Psychology*. Oxford, 1997. P.17.

биологические существа. Но между этим очевидным утверждением и вторым, менее очевидным, Олсон прослеживает логическую связь.

Если мы — биологические существа, то именно в продолжении жизни биологического существа будет заключаться наше длительное существование. Из этого, в свою очередь, следует, что преемственность психологических характеристик не является необходимым условием тождества личности. «Очевидно, не для всех людей... требуется психологическая преемственность для продолжения существования. Люди, находящиеся в вегетативном состоянии, человеческие эмбрионы, анэнцефалитики — это, в частности, человеческие животные, которые выживают без каких-либо психологических свойств»<sup>180</sup>.

#### **В. Сравнение биологического, физиологического и психологического подходов**

По мнению Олсона, биологический подход следует отличать от физиологического подхода. Согласно физиологическому подходу, продолжение нашего существования основано на сохранении большей части мозга. Мозг выделяется как основной орган в первую очередь потому, что он является носителем психики. Именно поэтому физиологический подход правильней всего относить к разновидностям психологического. Это важное замечание, так как здесь термины способствуют путанице. Не менее важно проводить различие между биологическим подходом Олсона (как наиболее типичного его представителя) и теорией Д. Виггинса. Последнего тоже иногда называют теоретиком «биологического подхода», однако даже в поздних работах Виггинс утверждает важность некоторых психологических черт для сохранения тождественности личности. Различия психологического и биологического подходов лучше всего проследить на гипотетических примерах.

---

<sup>180</sup> Olson E. The Human Animal. Personal Identity Without Psychology. Oxford, 1997. P.18.

Первый из них — уже упомянутый выше «Телетранспортер» Д. Парфита. Напомню, в этом сценарии предлагается перемещение личности с помощью передачи данных о ее физическом носителе, организме. По условиям мысленного эксперимента Парфита, исходный организм уничтожается, а в другом месте «собирается» новый. Для сторонника психологического подхода «Телетранспортер» Парфита является способом путешествий. Но для сторонника биологического подхода это камера смерти. Анималисты уверены, что, поскольку биологическая жизнь исходной личности прерывается, уничтожается также и нумерическая личность. И это является концом для жизни индивида. Разницу в двух подходах можно продемонстрировать также на примере другой гипотетической ситуации.

В мысленном эксперименте «Трансплантация мозга» из тела одного человека забирается головной мозг и переносится в тело другого человека. Трансплантация проходит успешно, так, что новый человек полностью реабилитируется. У него восстанавливаются все когнитивные функции, в том числе воспоминания, характер и т.п. Хотя, конечно, некоторое время ему непривычно в его новом теле. Как интерпретировать данную ситуацию? Для сторонника психологического подхода это способ обретения нового тела, то есть трансплантация тела. А для сторонника биологического подхода это выживание без мозга, или донорство мозга. Как видно, подходы ведут к противоположным суждениям. Но достаточно ли обоснован биологический подход как критерий тождества личности?

### **С. Критика биологического подхода**

Биологический подход справляется с отождествлением личности в различных ситуациях: когда человек спит, когда находится в вегетативном состоянии или только появляется на свет. Помимо этого, биологический подход, кажется, наиболее близок бытовой практике идентификации. Ведь именно на внешние признаки, на продолжение существования организма мы чаще всего ориентируемся, когда устанавливаем идентичность личностей. И



тем не менее, биологический подход, как и психологический, обладает существенными недостатками. Основным недостатком анимализма я считаю саму постановку проблемы «тождества личности».

Выше проблема асинхронного тождества личности была сформулирована как проблема реидентификации личности, то есть как проблема установления критериев тождества личности в момент  $t_1$  с личностью в момент  $t_2$ . Олсон, очевидно, меняет эту формулировку. Он предлагает искать критерии тождества личности в момент  $t_1$  с чем-то еще в момент  $t_2$ , то есть реидентифицировать личность и объект. Но какой смысл в такой реидентификации? Как я уже говорил, задача реидентификации должна решаться в практическом плане, в частности для приписывания моральной ответственности и выявления условий выживания. Но анималистическая реидентификация не дает таких возможностей. Для иллюстрации предложу гипотетический случай «Лобстер».

Некоторое время назад жил человек по имени Грегор Замза. Грегор участвовал в спасательных работах на атомной электростанции, в результате чего подвергся сильному радиационному облучению. После облучения его клетки стали мутировать, и за несколько лет Грегор превратился в Лобстера. Все основные биологические функции его организма сохранились, однако он потерял большую часть своих психических способностей. После метаморфоз его психика стала полностью соответствовать психике ракообразного. Через некоторое время Лобстер еще раз подвергся сильному радиационному облучению и стал вновь меняться. Результатом метаморфоз стало появление нового существа по имени Кейти. Кейти — человек женского рода и обладает только смутными воспоминаниями о поздних стадиях метаморфоз. Все биологические функции Кейти преемственны с Лобстером, но ее психика полностью отличается как от его психики, так и от психики Грегора.

Согласно биологическому подходу, Грегор тождественен Кейти. Но разве это не абсурд? Можно ли утверждать, что Грегор выжил как Кейти, или

что Кейти заслуживает медаль за участие в спасательных работах? А если нет, то что означает здесь тождественность? Возможно, в каком-то аспекте эта тождественность существует, но практический смысл установления этого тождества отсутствует. Единственное, с моей точки зрения, зачем стоит искать биологическое тождество, так это затем, чтобы найти в нем основание тождества личности. Преемственность биологического организма — это лишь косвенный способ установить другую, более важную с практической точки зрения связь.

Нужно отметить, что даже сам Олсон не отрицает этого. Так, он пишет: «если практически значимые отношения тождества во времени обычно сопровождаются психологической преемственностью, ...то это любопытное этическое следствие для биологического подхода. В частности, это означает, что нумерическое тождество не необходимо для практического отношения тождества»<sup>181</sup>. Для Олсона это, однако, лишь любопытное следствие. Для нас в контексте данного исследования, это уже повод искать другой подход. Такой альтернативой является субстанциональная теория личности.

### 3. Субстанциональный подход и его сторонники

Согласно субстанциональному подходу, тождество личности заключается в тождестве субстанции. Иными словами, чтобы личность  $P_2$  в момент  $t_2$  была той же самой, что личность  $P_1$  в момент  $t_1$ , у них должна быть одна нумерическая субстанция. Под субстанцией обычно подразумевается принципиально единая и неделимая сущность вроде живой нематериальной души или монады, ментальной вещи, способной существовать автономно от тела и мозга. То есть сторонник этой теории должен занимать дуалистическую позицию в отношении проблемы сознание-тело. Субстанциональный подход является, по определению Д. Парфита, ведущего современного теоретика тождества личности, «простым решением» проблемы тождества в том смысле, что он предполагает одно единственное

---

<sup>181</sup> Olson E. The Human Animal. Personal Identity Without Psychology. Oxford, 1997. P.70.

обстоятельство, определяющее тождественность — неизменную духовную монаду. Простой подход не сводит тождество личности к тождеству или преемственности каких-то других свойств. В этом смысле он отличается от описанных выше «комплексных решений»: психологического и физиологического, которые по своей сути являются редукционистскими. Субстанциональный подход обеспечивает психологический комфорт, создавая личности как бы дополнительный контур безопасности. Возможно, в силу этого он разделяется большинством неспециалистов. В рамках этой теории можно непротиворечиво помыслить реинкарнацию, воскрешение в теле или бестелесную жизнь после смерти, переселение в другое тело или даже совместное использование двумя и более ментальными субстанциями одного тела. То есть эта теория согласуется с распространенными представлениями о возможностях существования и выживания личности. Субстанциональный подход позволяет также решить вопрос о единстве сознания, об уникальной природе «личности» (через нумерическую уникальность субстанции), установить тождество между бессознательными фазами жизни человека (в том числе в состоянии эмбриона и вегетативном состоянии) и зрелой сознательной личностью, объяснить особую заботу личности о своем будущем и моральную ответственность. Возможно, с учетом тезиса о неделимости ментальной субстанции, он также позволяет разрешить описанные выше парадоксы редупликации личности. В силу множества этих обстоятельств его разделяют и некоторые философы.

Самым знаменитым сторонником «простого» решения был Р. Декарт. Вслед за Декартом «простое» решение проблемы тождества личности принимали также Дж. Батлер и Т. Рид. Среди современников наиболее известными философами, отстаивающими эту позицию, считаются американец Р. Чизом<sup>182</sup> и британец Р. Суинбёрн. Здесь я остановлюсь подробнее на аргументах оксфордского философа. Его защита

---

<sup>182</sup> Chisolm R. On the simplicity of the soul // Philosophical perspectives, Vol. 5, Philosophy of religion, 1991.

субстанциональных критериев тождества построена на современной критике психологического и физиологического подходов и наиболее любопытна.

#### А. Субстанциональный поход Р. Суинбёрна

Суинбёрн различает эпистемологический и метафизический уровни решения проблемы тождества личности. Первый уровень, по его мнению, относится к практической задаче реидентификации, то есть заставляет исследователей искать наглядные свидетельства тождественности. Второй уровень относится к подлинному основанию тождественности. Философ считает, что эмпирицистские теории — так Суинбёрн называет психологический и физиологический подходы (отметим, что он не выделяет «биологический» подход отдельно, но его аргументы против физиологического подхода могут быть отчасти использованы и против биологического) — успешно справляются с первой задачей, но неспособны решить вторую. Память, а также психологическая и физиологическая преюмственность действительно являются надежными сопутствующими критериями, но эти факторы не являются основой тождественности личности. Ключевым основанием для этого вывода является неизбежное, по мнению Суинбёрна, следствие эмпирицистских теорий: в некоторых гипотетических случаях, согласно этим теориям, «не существует правильного ответа на вопрос, являются ли две личности тождественными»<sup>183</sup>. Он убежден, что все эмпирицистские теории ведут к абсурдным следствиям. Доказательство этого начинается с новой классификации, которую предлагает оксфордский философ и которая, по его замыслу, исчерпывает эмпирицистские альтернативы. Эта классификация предполагает три типа теорий тождества личности. Против каждого типа Суинбёрн использует отдельный аргумент. Результатом этих рассуждений должно стать опровержение всех эмпирицистских теорий.

---

<sup>183</sup> Shoemaker S., Swinburne R. Personal Identity. Oxford: Wiley-Blackwell, 1991. P. 235.

## В. Недостатки эмпирицистских теорий

Первый тип эмпирицистских теорий предполагает возможность редупликации. Такими теориями являются классические теории, основанные на психологических критериях: памяти и характере. Согласно таким теориям, личность  $P_2$  в момент  $t_2$  является тождественной личности  $P_1$  в момент  $t_1$ , если удовлетворяет какому-то критерию, при этом более одной личности может удовлетворять этому критерию. По мнению Суинбёрна, такие теории являются ложными потому, что неизбежным следствием их является противоречие. Одна личность в прошлом, согласно этим теориям, может оказаться двумя и более личностями в будущем, как в случае с Гаем Фоксом, Чарлзом и Робертом. Но две личности не могут быть тождественны одной, а также две тождественные личности не могут быть в разных местах. Следовательно, критерий тождества, выдвигаемый этими теориями, неверен.

Теории второго и третьего типа не допускают редупликации, но это достигается разными способами. Теории второго типа — это теории первого типа, к которым просто добавлено дополнительное условие, запрещающее редупликацию. Это условие может звучать так: личность в момент  $t_2$  не может быть той же личностью в момент  $t_1$ , если в момент  $t_2$  существует еще какая-то личность, удовлетворяющая тем же критериям тождества. Не только физиологические, но и психологические теории в целом могут относиться ко второму типу. Например, ко второму типу относятся теории, согласно которым утверждается, что  $P_2$  в момент  $t_2$  — это та же личность, что  $P_1$  в момент  $t_1$ , если и только если существует пространственно-временная преемственность между какой-то частью мозга  $P_1$  и  $P_2$  и не существует аналогичной преемственности между какой-то частью мозга  $P_1$  и другой личностью, кроме  $P_2$ . Теории второго типа, по мнению Суинбёрна, тоже ведут к абсурдному следствию. Этим следствием является то, что решение о тождестве и выживании личности в этих теориях зависит от судьбы каких-то других личностей. Примером может служить случай с трансплантацией двух полушарий разным донорам. Согласно теориям второго типа, выживание

личности будет возможно, только если один из доноров погибнет, а это крайне контринтуитивное следствие. Следовательно, заключает Суинбёрн, теории этого типа тоже ошибочны.

Теории третьего типа также не допускают редупликацию, но в них нет дополнительного условия. Просто в этих теориях критерии тождества определены так, чтобы им могла соответствовать только одна личность в будущем. Примером может служить теория, которая основывается на следующем физиологическом критерии тождества:  $P_2$  в момент  $t_2$  — это та же личность, что  $P_1$  в момент  $t_1$ , если и только если существует такая пространственно-временная преемственность между организмом  $P_1$  и  $P_2$ , что на каждой последующей временной стадии между  $t_1$  и  $t_2$  организм имел более 90% мозга тождественными соответствующей части мозга этого организма на предшествующей стадии. Такое условие не допускает «разделение» организма, ибо тогда не будет удовлетворено условие наличия 90% предшествующего мозга.

Теории такого типа тоже ведут к абсурдному следствию. Этим абсурдным следствием является то, что одна и та же личность может являться и не являться тождественной другой личности в зависимости от того, каким образом она претерпевала изменения. Так, личность, единомоментно потерявшая 50% мозга, не выживет (так как она сохранит менее 90% мозга), но если она потеряет 50% мозга в результате череды потерь, каждая из которых будет составлять менее 10%, то она продолжит существование.

Кроме того, критерий тождества, предлагаемый такой теорией, представляется произвольным. Почему 90%? Почему 70% или 80% мозга недостаточно для выживания личности? Как решить, где именно должна проходить грань, обеспечивающая тождество? Какое бы процентное отношение не было поставлено, оно будет выглядеть искусственным и произвольным. Кроме того, всегда остается возможность для пограничного случая. Предположим, теория допускает, что для выживания достаточно

сохранения 50% мозга. Тогда представим случай, что именно 50% мозга сохраняется у пациента после вынужденной операции. Прекратится ли его существование (даже если пациент выживет и сохранит когнитивные способности)? По мнению Суинбёрна, логично предположить, что до операции сам пациент будет переживать и оценивать операцию как рискованную, даже если все факты об исходе будут известны заранее. Это, по мнению оксфордского философа, свидетельствует о ложности теорий третьего типа.

Суинбёрн считает, что фундаментальная проблема этих теорий в том, что они вопрос о тождестве личности уподобляют вопросу о тождестве других физических объектов. «Нет ничего загадочного в отношении пограничного случая, когда новый автомобиль сконструирован из частей моей старой машины и новых частей. Поскольку у машин нет чувств и надежды на выживание... Но сознающий объект, которым является личность, желает оставаться в сознании, и кажется, что не может существовать промежуточной возможности между тем, чтобы это была та же личность или другая»<sup>184</sup>. По мнению оксфордского философа, личности (как объекты с сознанием) принципиально отличаются от других физических сущностей и имеют особые условия тождественности.

### **С. Особенности субстанционального подхода Суинбёрна**

Сторонники эмпирицистских теорий не проводят различия между тождеством физических объектов и личности, и Суинбёрн считает это основанием для сомнения в их истинности. Единственным решением этой метафизической головоломки, с его точки зрения, может быть субстанциональный подход. В основании его субстанционального подхода несколько утверждений, которые я обозначу как тезисы о: (1) скрытом основании, (2) определенности и (3) дискретности отношения тождества личности. Совокупность этих тезисов составляет содержательное ядро

---

<sup>184</sup> Shoemaker S., Swinburne R. Personal Identity. Oxford: Wiley-Blackwell, 1991. P. 240.

позиции Суибёрна. Согласно тезису о нередуцируемости, «тождество не заключается исключительно в преемственности одной или нескольких наблюдаемых характеристик... Единственной альтернативой является сказать, что тождество личности — это нечто фундаментальное»<sup>185</sup>. То есть, согласно теории Суинбёрна, единство личности несводимо к каким-то другим отношениям.

Фундаментальным, нередуцируемым отношением, по мнению философа, является тождество субстанции, нематериальной сущности. Это утверждение я называю *тезисом о скрытом основании*. Основными свойствами этой сущности является способность «иметь сознательные состояния (например, мысли и ощущения) и осуществлять намеренные действия»<sup>186</sup>. Согласно тезису о скрытом основании, отношения между наблюдаемыми характеристиками представляют лишь косвенные свидетельства тождества субстанции, но не гарантируют его.

Одна нематериальная субстанция всегда либо тождественна, либо не тождественна другой субстанции в другое время. Это — второй тезис, который я называю *тезисом об определенности*. В эпистемологическом плане даже наличия полной информации о наблюдаемых характеристиках, этих косвенных свидетельствах тождества, может быть недостаточно для того, чтобы сделать вывод о наличии или отсутствии тождества личности со всей определённойностью. Но это не означает, что одна личность может стоять в некотором неопределённо-тождественном отношении к другой личности в другой момент времени. Второй тезис утверждает, что это отношение всегда определённо.

Эмпирицистские теории склоняют нас к тому, чтобы считать тождество вопросом степени. То есть они допускают истинность высказываний «он в большей степени тот же человек, что и был». Но, по мнению Суинбёрна, эти

---

<sup>185</sup> Ibid., P. 240.

<sup>186</sup> Ibid., P. 27.



высказывания ложны. Высказывания о тождестве личности всегда дискретные: личность может быть либо тождественна другой, либо отличной от неё. И промежуточных случаев быть не может — личность не может быть в какой-то степени тождественной. Это утверждение мы будем называть *тезисом о дискретности*. Можно ввести понятие тождества личности со степенью тождества — по аналогии с тождеством физических объектов, которое может рассматриваться как недискретное. Но этот ход непродуктивный, так как он, с позиции Суинбёрна, не относится к понятию тождества личности, с которым мы имеем дело в обыденном языке. Все три тезиса, напротив, соответствуют обыденным интуициям в отношении тождества личности.

#### **Д. Критика субстанционального подхода с помощью верификационизма**

Субстанциональный подход имеет ряд описанных выше преимуществ: он согласуется с обыденными интуициями в отношении тождества и реидентификации личности, решает запутанные ситуации с редупликацией и разделением личностей и т.д. Но он сам является мишенью для критики. Против него можно использовать аргументы, построенные на принципе верификации. Этот принцип защищался логическими позитивистами в первой половине XX в. Однако в его основе лежат идеи, заложенные еще Д. Юмом. Как пишет историк философии, исследователь верификационизма Ш. Мисак: «Они [логические позитивисты — Д.В.] в качестве исходной точки взяли ключевую мысль Юма о том, что обоснованные убеждения должны быть обоснованы данными опыта»<sup>187</sup>. Согласно одной из формулировок принципа верификации, «осмысленные предложения — это те предложения, истинность или ложность которых на практике может быть подтверждена»<sup>188</sup>.

---

<sup>187</sup> Misak C. Verificationsim. Its history and prospects. London and New York: Routledge, 2005. P. 56.

<sup>188</sup> Ibid., P. 61.

Опираясь на принцип верификации в такой формулировке, можно попытаться доказать, что утверждение о субстанциональной основе тождества личности не является осмысленным. Для этого необходимо применить этот принцип к тезису о скрытом основании тождества из теории Суинбёрна. Согласно тезису о скрытом основании тождества, наблюдаемые характеристики могут лишь косвенным образом способствовать реидентификации. То есть любые внешние свойства, доступные в опыте, не являются достаточным основанием для реидентификации. То есть на практике невозможно окончательно подтвердить тождество личности в разное время. Но это противоречит принципу верификации в описанной выше формулировке. Значит ли это, что субстанциональный подход выдвигает бессмысленный критерий тождества личности?

Такой верификационистский аргумент мог бы быть использован против подхода Суинбёрна. Однако посылка в этом аргументе — не самая надежная. Принцип верификации в такой формулировке, в его сильной версии, сам уязвим для критики. На это указывает Суинбёрн: «Что касается меня, я не нахожу оснований принимать любые из этих видов верификационизма. До тех пор, пока мы понимаем грамматическую форму повествовательного предложения и слова, которые в нем употребляются, мы понимаем, что в нем утверждается, даже если нет мыслимых способов подтвердить его истинность или ложность»<sup>189</sup>. По мнению оксфордского философа, сильная версия верификационизма предъявляет слишком высокие требования, так что даже тривиальные утверждения попадают в категорию бессмысленных. Это сводит принцип к абсурду. В качестве примера он приводит выражение «в комнате стоит стол». Это утверждение нельзя абсолютно точно подтвердить или опровергнуть, так как восприятие стола может быть иллюзией или сном. Значит, согласно принципу верификации в

---

<sup>189</sup> Shoemaker S., Swinburne R. *Personal Identity*. Oxford: Wiley-Blackwell, 1991. P. 39-40.

сильной версии, это бессмысленное выражение. Но это абсурд, и, значит, принцип верификации в сильной версии ложен.

Суинбёрн считает, что для осмысленности выражения достаточно, чтобы был *хоть какой-нибудь способ вероятностной верификации или фальсификации высказывания*. Это утверждение можно считать слабой версией верификационизма. Оно вполне правдоподобно. Философ предлагает рассмотреть три предложения: «на нашей планете нет единорогов»; «в нашей галактике нет единорогов»; «в нашей Вселенной нет единорогов». Практическая возможность проверки всех трех утверждений разная, но представляется, что они в равной степени имеют смысл. Из этого, по мнению Суинбёрна, следует, что любой способ вероятностной верификации или фальсификации достаточен для осмысленности высказывания. «До тех пор, пока мы требуем только возможности подтверждения или опровержения, будь то возможность логическая, физическая или практическая, доступной хотя бы одной личности или кому-нибудь, сейчас или когда-нибудь, дуалистическая теория тождества личности не будет уязвима для верификационизма»<sup>190</sup>. Таким образом, Суинбёрн опровергает сильную версию и соглашается со слабой версией верификационизма. При этом он считает слабую версию верификационизма совместимой с утверждением о субстанциональных критериях тождества личности.

Действительно, психологические и физические характеристики не дают безусловной гарантии тождества. Но, согласно субстанциональному подходу Суинбёрна, наличие преимущества психологических, физиологических и биологических характеристик позволяет косвенным образом с некоторой вероятностью реидентифицировать личность. Так, сохранение большей части мозга и воспоминаний в рамках субстанционального подхода является подтверждением, что личность  $P_2$  — это та же личность, что и личностью  $P_1$ .

---

<sup>190</sup> Shoemaker S., Swinburne R. Personal Identity. Oxford: Wiley-Blackwell, 1991. P. 40-41.

А появление конкурента  $P^*_2$  с большей частью мозга и воспоминаний  $P_1$  — фальсификацией отождествления  $P_2$  и  $P_1$ . То есть, по мнению Суинбёрна, существуют вероятностные способы верификации субстанциональных критериев тождества личности, а значит, они осмысленны.

Ответ оксфордского философа на критику с помощью верификационизма я считаю не до конца удовлетворительным. Вопреки утверждениям Суинбёрна, я считаю, что даже слабая версия верификационизма несовместима с субстанциональным подходом к тождеству личности. Оксфордский философ утверждает, что биологические, физиологические и психологические характеристики могут служить косвенными основаниями для высказываний о тождестве. Но эти характеристики могли бы быть косвенными свидетельствами тождества субстанции, только если бы они были должным образом связаны с субстанцией. Эта особая связь могла бы быть отношением тождества, супервентности или каузальной зависимости. Но таких связей, как мне представляется, нет. Отношение тождества между нематериальной субстанцией и психологическими, физиологическими или биологическими характеристиками должно быть отвергнуто, так как, по определению Суинбёрна, нематериальная монада обладает совсем иными свойствами: она неделима и ее сущностные свойства не могут быть скопированы. Отношение супервентности предполагает прямую зависимость свойств: набор свойств  $B$  является супервентным по отношению к набору свойств  $A$  в том случае, если объект не может различаться в свойствах  $B$  в отсутствие различий в свойствах  $A$ . Из теории Суинбёрна следует, что субстанция может быть неизменной при изменении указанных характеристик. Следовательно, между свойствами субстанции и этими свойствами не существует отношения супервентности. И, наконец, тезис о каузальной связи между субстанцией и психологическими и телесными свойствами крайне проблематичен.

Очень мало философов допускают возможность участия субстанций в каузальных связях. Наиболее распространено мнение о том, что только события могут находиться в каузальной связи. Более того, каузальная эффективность нематериальной субстанции крайне экзотична и плохо согласуется с правдоподобным тезисом о каузальной замкнутости физического мира. Нужно опровергнуть много аргументов, чтобы доказать, что нематериальная субстанция может быть каузально связана с психическими, физиологическими и биологическими характеристиками. Таким образом, связь между монадой и этими свойствами, скорее всего, аналогична связи между изображениями единорогов, описаниями единорогов, продаваемыми «рогами» единорогов и единорогами.

Эти «косвенные доказательства» существования единорогов не подтверждают их существования потому, что между ними и единорогами отсутствует должная связь. Эти «доказательства» не тождественны единорогам, не супервентны и каузально не связаны с существованием единорогов. Описания этих удивительных животных не связаны каузально с восприятиями единорогов, а «рога» на самом деле являются рогами нарвалов. Как из обнаружения большего числа текстов или «псевдорогов» не следует увеличение вероятности существования единорогов, так из наличия большого числа преемственных психологических и прочих характеристик не следует большая вероятность тождества монад. Это можно проиллюстрировать гипотетическим случаем: души Чарлза и Роберта меняются местами. Физические характеристики этих двух людей не изменяются. Поскольку психологические характеристики супервентны на физических, психологические характеристики также остаются неизменными. Со стороны никак нельзя будет обнаружить, что произошла смена монад. Но и изнутри, в интроспекции Чарлз и Роберт не смогут обнаружить этой смены, ибо они сохраняют все ту же память и прочие психологические особенности. Выходит, никто и никак не сможет обнаружить смены монад, то есть

никаким образом и ни с какой вероятностью нельзя будет подтвердить или фальсифицировать высказывание об этой смене. Значит, констатация такого изменения является бессмысленным выражением, а следовательно, бессмысленным является и субстанциональный критерий тождества личности.

#### 4. Критика классических подходов с помощью Аргумента пяти трансплантаций

Похоже, мы оказались в патовой ситуации. Субстанциональный подход не проходит. Психологический и физиологический критерии противоречат одним гипотетическим ситуациям, биологический критерий — другим. Цепочку рассуждений, приведшую нас к этому пату, я резюмирую с помощью последовательности мысленных экспериментов. Для простоты все эти пять ситуаций предлагаю называть трансплантациями. Все пять трансплантаций будут ответами на гипотезы, антитезисами к выдвигаемым утверждениям.

*Гипотеза 1. Критерием тождества, необходимого для моральной ответственности, является тождество души.*

##### **А. Трансплантация души**

Предположим, Браун и Робинсон, как предполагают сторонники субстанционального критерия, обладают душами вдобавок к телам и их психологическим, физиологическим и биологическим характеристикам. У каждого из них есть своё уникальное содержание психики и различные психические способности. Демон решает сыграть с ними злую шутку. В какой-то момент он меняет их души местами, так что душа Брауна трансплантируется в тело Робинсона, и наоборот. При этом все психические свойства их тел остаются неизменными: каждое из тел продолжает иметь те же воспоминания, те же намерения и желания, те же психические качества и тот же характер. Для душ Брауна и Робинсона эта перемена должна остаться незаметной, ведь у них сохранились все те же воспоминания, что были. Для

внешних наблюдателей тоже ничего не изменилось. Браун и Робинсон не изменились. Стал ли Браун Робинсоном, а Робинсон — Брауном? Кажется, что этого не произошло. Должен ли теперь Робинсон нести ответственность за поступки, совершенные Брауном? Скорее всего, нет. Каждая личность осталась на своем месте. Из этого следует, что субстанциональный критерий тождества личности приводит к контринтуитивным выводам, и поэтому, скорее всего, ложен. Тогда вместо субстанционального подхода испытаем биологический подход.

*Гипотеза 2. Критерием тождества, необходимого для моральной ответственности, является тождество/преемственность тела (биологических функций).*

#### **В. Трансплантация мозга**

Злонамеренный нейрохирург решил провести эксперимент: он решил перенести мозг Брауна в тело Робинсона и наоборот. Пока оба человека спали, он ввел им наркоз и осуществил пересадку мозгов. Когда утром, после успешной операции, люди проснулись, мозг Брауна был в теле Робинсона и наоборот. Проснувшись, тело Брауна стало вести себя совсем как Робинсон. У него были характер Робинсона, воспоминания Робинсона, привычки Робинсона, которого теперь крайне удивлял собственный вид в зеркале. Аналогичное случилось и с телом Робинса. Оно стало вести себя как Браун. Где Браун, а где Робинсон?

Кажется, что в результате трансплантаций мозгов личности поменялись местами. Личности перешли вместе с перемещением мозгов, хотя органы, отвечающие за биологические функции, остались на своих местах, тела остались на своих местах. Должно ли тело Брауна нести ответственность (например, наказание) за поступки Брауна? Кажется, нет. Видимо, биологический критерий приводит к контринтуитивным выводам и поэтому, по-видимому, ложен. Тождественность, вероятно, связана с тождеством мозга.

*Гипотеза 3. Критерием тождества, необходимого для моральной ответственности, является тождество мозга.*

### **С. Трансплантация психики**

Ученые научились копировать информацию, которая хранится в нейронных сетях мозга. В этой информации заключено все содержание психики и все психические способности. Команда ученых проводит эксперимент. Они копируют все данные из мозга Робинсона в мозг Брауна и наоборот. Мозги Брауна и Робинсона остаются на месте, но все их психические особенности переносятся. Теперь Робинсон ведет себя точно так, как раньше вел себя Браун. Все воспоминания Брауна стали воспоминаниями Робинсона. Во много ситуация идентична той, что была описана выше. Где Браун, а где Робинсон?

Мне кажется, что интуиции склоняют к тому, что произошла смена тел у личностей. Биологические органы остались на месте, мозги остались на месте, но личности мигрировали. Но это противоречит Гипотезе 3. Возможно, это означает, что носителем личности является психика, а не физический орган, или субстрат. Это дает возможность надеяться на продолжение жизни после смерти. Если, конечно, ученые научатся копировать всю информацию о психике на новый носитель. Но для оптимизма пока недостаточно оснований. Стоит проверить и эту гипотезу.

*Гипотеза 4. Критерием тождества, необходимого для моральной ответственности, является тождество/преемственность психики.*

### **Д. Трансплантация психики с умножением личности**

Предположим, что Робинсон — идеальный сотрудник. Он безупречно выполняет все поручения и не требует повышений оплаты. Работодатель, впечатленный его исключительной результативностью, решает скопировать содержание психики и его психические способности всем сотрудникам. Они записываются на процедуру, и на следующий день все сотрудники



приобретают характер, привычки, намерения, желания и воспоминания Робинсона. Когда их спрашивают имя, они хором отвечают — Робинсон. Когда их спрашивают о биографии, они рассказывают биографию Робинсона. И работают точно так, как работает Робинсон. Это — полные психологические копии Робинсона, хотя они обладают другими телами (различия в которых, как мы видели, не так важны) и пространственным расположением. Нумерически это, по всей видимости, разные люди. Кто они? Несут ли они ответственность за поступки, совершенные Робинсоном в прошлом?

Кажется, что здесь интуиции начинают подводить. Этот случай настолько фантастический, что полагаться на опыт становится невозможно. Сходных ситуаций в жизни не встретишь. Что-то подсказывает, что эти люди все — Робинсоны. Но как много может быть Робинсонов? Нет ли здесь противоречия? Как может быть множество идентичных людей? Кто теперь женат на миссис Робинсон? Кто отец детей Робинсона? Кто имеет право получать пенсию Робинсона и отвечать за его поступки? Эти вопросы могут поставить в тупик. Судя по всему, возможность трансплантации психики сразу нескольким агентам ставит под сомнение и психологический подход к тождеству личности. Но если вариант умножения личности кажется малопонятным или недостаточно убедительным, вместо него против психологического подхода можно использовать вариант деления личности. Его эмпирическая возможность выше.

#### **Е. Трансплантация психики с делением личности (Fission case)**

Носителем высших психических способностей является мозг. Кажется, поэтому у нас есть интуиция, что при трансплантации мозга происходит трансплантация личности. Мозг, как известно, состоит из двух полушарий. При этом каждое полушарие автономно, может изолированно выполнять все высшие психические функции. Существует способ лечения обширной эпилепсии путем удаления одного из полушарий мозга — гемисферэктомия.

Если операция проводится в детском возрасте, благодаря нейропластичности все когнитивные функции восстанавливаются.

Предположим, что Вильямсон страдает редким заболеванием, которое связано с взаимодействием двух полушарий мозга. В качестве лечения ему предлагают операцию: пересадку обеих полушарий в организмы доноров Робинсона и Брауна. Мозги доноров полностью дисфункциональны, Робинсон и Браун находятся в хроническом вегетативном состоянии. Допустим, операция проходит удачно, и мозги Вильямсона приживаются в организмах Робинсона и Брауна. Тело Вильямсона остается в вегетативном состоянии. Где теперь Вильямсон? Тела Робинсона и Брауна теперь имеют психические способности Вильямсона, его убеждения и желания. Кто из них тождественен Вильямсону? Оба?

Предположим, Робинсон позавтракал, а Браун не успел, тогда Робинсон сытый, а Браун — голодный. Если обе эти личности — Вильямсон, то Вильямсон и сытый, и голодный. Это противоречие. Значит, этот пример ставит под сомнение психологический критерий. Дело в том, что *тождество* — это отношение один к одному, а психический подход допускает отношение один ко многим. Или Вильямсон уже изначально был множественен, и мы просто могли об этом не знать?

Случаи с умножением или делением личности являются проблемой для психологического подхода. Но в совокупности представленные гипотетические ситуации ставят под сомнение все описанные выше подходы идентификации личности. Совместно эти гипотетические ситуации можно назвать Аргументом пяти трансплантаций. Этот аргумент демонстрирует проблему Исчезающей личности и является серьезным препятствием к возможности свободы и моральной ответственности. Ключ к разрешению патовой ситуации можно попытаться отыскать в другом мысленном эксперименте, фантастической истории, которая подводит нас к новой

концепции личности. История эта так и называется: «Где Я?», а её автор — уже известный нам Д. Деннет.

#### IV. Теория личности Д. Деннета

##### 1. Мысленный эксперимент «Где Я?»

История «Где Я?» появилась впервые в сборнике «Мозговые штурмы» (1978). Она была включена в завершающую главу и значительно отличалась от всех предыдущих. Поэтому автор во введении сам назвал ее «десертом» к трудным, серьезным философским рассуждениям. Повторно эта история была опубликована в книге «Глаз разума. Фантазии и размышления о самосознании и о душе», вышедшей под совместной редакцией Д. Хофштадтера и Д. Деннета. По жанру она гораздо естественней подходила именно для этой публикации. В ней собрано, пожалуй, только «сладкое». В книге — эссе философов, писателей, ученых (Р. Нозика, Х. Борхеса, С. Лема, Р. Докинза, Дж. Сёрла), размышляющих над вопросами, «Что такое разум?», «Что такое я?», «Может ли материя думать или чувствовать?», дополненные комментариями редакторов. Но выводы, следующие из этой истории, сразу не очевидны. «Это... — попытка выявить возникающие... противоречия и описать их живо и образно. Мы хотели не столько ответить на вопросы, сколько взбудоражить всех, кто прочтет нашу книгу...»<sup>191</sup>, — признаются Деннет с Хофштадтером. История «Где Я?» настолько художественна, что даже стала одним из сюжетов одноименного фильма голландского режиссера Пьета Ходедреса, где в роли *Деннета* снимается сам Деннет. Вместе с ним героями фильма оказались его соредактор, философ-математик Дуглас Хофштадтер и когнитивист Марвин Мински.

Как и прочие эссе сборника «Глаз разума», фантастический рассказ Деннета оставляет читателя в задумчивости. Он посвящен вопросу тождества личности и самоидентичности. Но какой подход он представляет? В

---

<sup>191</sup> Хофштадтер Д., Деннет Д., Глаз разума. М.: Бахрах-М, 2003, С. 26.

некотором смысле история «Где Я?» похожа на уже упомянутый мысленный эксперимент с двумя черными ящиками, приведенный в первой главе. Здесь автор также не разъясняет свои выводы явным образом. Вместо этого в тексте мы находим разные, порой трудно соотносимые размышления героя истории и фрагментарные замечания автора. Какова позиция Деннета? Как и рассказ о хитроумной диковинной археологической находке, история «Где Я?» заставляет задуматься: «Как можно ее приспособить?», «Для чего бы она могла пригодиться?», «Каков был замысел создателя?», «Сделана ли она всерьёз или для забавы?». В надежде прояснить альтернативы решения проблемы Исчезающей личности мы проведем исследование этого фантастического сюжета.

Итак, *Деннет* выполняет секретную миссию Пентагона (конечно же, против «красных»). Ему придется обезвредить радиоактивную боевую головку где-то в недрах земли. Но для того, чтобы мозг не был поврежден смертоносным излучением (для других органов излучение относительно безопасно), он должен остаться в лаборатории в Хьюстоне. Оттуда мозг будет осуществлять контроль над телом с помощью системы радиосвязи. Каждое нервное окончание в пустом черепе будет соединено с микропередатчиками и микроприемниками, которые, в свою очередь, будут направлять сигнал непосредственно в мозг. «Представьте, что это всего лишь растяжение нервов... Мы всего лишь сделаем ваши нервы бесконечно растяжимыми, вставив в них радиоконтакты»<sup>192</sup>. *Деннета* кладут на операционный стол, анестезия... и вот, сразу после пробуждения он задает обычный для такого случая вопрос: «Где я?» Экстравагантная завязка нужна автору для того, чтобы обнаружить проблематичность этого вопроса. В самом деле, где *Деннет*, когда части его — в разных местах: мозг — в чане, а тело — на операционном столе?

---

<sup>192</sup> Хофштадтер Д., Деннет Д., Глаз разума. М.: Бахрах-М, 2003 С. 192.

### *Конфигурация 1. Гамлет (тело) и Йорик (мозг)*

Подобная дилемма уже встречалась раньше, когда обсуждались различные варианты трансплантаций. Эта первая конфигурация Деннета отчасти близка случаю с трансплантацией мозга. Но в той ситуации ответ казался очевидным: личность была там, где мозг. Когда производилась трансплантация мозга, вместе с ней трансплантировалась и психика личности. Это, скорее всего, предопределяло работу интуиции. В данном же случае всё не так однозначно. *Мозг Деннета трансплантирован, но тело сохранило функциональность и психологические характеристики.* Это запутанный вариант трансплантации. «Итак, я сижу на складном стуле и смотрю сквозь небьющееся стекло на свой собственный мозг. Но погодите... не должен ли я сказать, что плаваю в булькающей жидкости, озираемый собственными глазами?»<sup>193</sup>. Герой мысленного эксперимента, глядя на собственный мозг, философствует в шекспировском духе. И автор сюжета решает не ограничиваться намеком на знаменитую трагедию. «Совсем сбитый с толку, я попытался прибегнуть к любимому трюку философов. Я принялся раздавать вещам имена. "Йорик, — сказал я вслух своему мозгу, — ты мой мозг. Остальное тело, сидящее на этом стуле, я назову "Гамлетом"». Все мы сейчас здесь: мой мозг, Йорик, мое тело, Гамлет, и я сам, *Деннет*. Где же я *теперь?*»<sup>194</sup>

**Альтернатива 1. Где Йорик, там и Деннет.** Такое решение подходит некоторым сторонникам психологического подхода. В первую очередь — тем из них, кто ассоциируют тождество личности именно с мозгом<sup>195</sup>. В мозге запечатлены подробности жизни, ключевые психологические качества и способности личности. Но, как я отметил, случай трансплантации в истории Деннета особенный. Несмотря на то, что «управление» организмом

---

<sup>193</sup> Ibid. С. 193.

<sup>194</sup> Ibid., С. 194.

<sup>195</sup> О том, что сторонники тождества личности через тождество мозга являются фактически сторонниками психологического подхода, см. монографию Олсона «Человеческое животное: тождество личности без психологии». Olson E. The Human Animal. Personal Identity Without Psychology. NY, Oxford: Oxford University Press, 1999.

ведется из «чана» в научной лаборатории, перспектива героя — со стороны его тела. И основные жизненно важные события для *Деннета*, кажется, происходят не в чане, где Йорик, а там, где Гамлет. Несмотря на то, что мышление происходит «в чане», место личностных интересов — это местонахождение его тела.

**Альтернатива 2. Где Гамлет, там и *Деннет*.** Такое решение подходит сторонникам биологического подхода. И, кажется, эта альтернатива ближе субъективному восприятию *Деннета*, созерцающего собственный мозг. Шекспировский герой, глядя на череп, обнаруживает бренность человека, а *Деннет*, глядя на собственный мозг... что это не он! И правда, мозг, бездвижный, находится в чане, никуда, по-видимому, не собирается. А *Деннет* вот-вот отправится в Тулсу, где будет обезвреживать боеголовку. Или на секретное задание отправится не *Деннет*, а радиоуправляемый биологический робот? Слишком сильна интуиция, связывающая личность с телом, с действующим в мире и воспринимающим агентом. И разве нет противоречия в возможности воспринимать и быть воспринимаемым одновременно? Гамлетовские (Йориковские?) размышления наводят героя на любопытную мысль, примиряющую, кажется, обе позиции.

**Альтернатива 3. *Деннет* там, где он сам считает.** Эта неожиданная новая позиция любопытна тем, что не предполагает жесткой зависимости между местоположением личности и какой-либо из ее физических частей. Точка зрения, а с ней и субъект, может мгновенно перемещаться между различными физическими объектами. В нашем случае — между мозгом и телом. Более того, местоположение «Я» может вообще не соответствовать какому-либо физическому объекту. Допустим, мы наблюдаем за событиями с помощью видеокамеры, находящейся в другом помещении. Если ничто не отвлекает нас, складывается устойчивое ощущение присутствия в этом другом помещении. Присутствия прямо за объективом видеокамеры. Значит,

точка зрения может вообще не совпадать с положением частей человека. К этой позиции, по-моему, склоняется герой мысленного эксперимента.

Но нас, читателей, она может насторожить. Совпадает ли на этот счет позиция *Деннета-героя* и *Деннета-автора*? В конечном итоге, Деннет — один из самых последовательных сторонников материализма. Может ли он допустить, чтобы личность, подобно призраку, перескакивала между объектами, и даже существовала в пустоте? Дальнейшие злоключения героя мысленного эксперимента усиливают сомнения в полном совпадении взглядов героя и автора.

Итак, Йорик остается в безопасности, в стенах тexasской лаборатории, а Гамлет отправляется к радиоактивному объекту. Но, не успев завершить задание, он попадает в беду: «Я совершенно оглох. Сначала я подумал, что сломались наушники моей рации, но когда я постучал по шлему, то ничего не услышал. По-видимому, сломались слуховые приемопередатчики... Я начал описывать то, что со мной происходит. В середине предложения я понял, что случилось еще что-то. Мой голосовой аппарат внезапно прекратил работать...». Следом отказала правая рука, потом левая и *Деннет* совсем ослеп. Проклиная ученых, впутавших его в эту авантюру, он сидит, оглохший, ослепший и неподвижный в радиоактивной шахте, но вот... с отключением последнего передатчика он мигом переносится... в Хьюстон. «Если минуту назад я был заживо погребен в Оклахоме, то теперь я был лишен тела в Хьюстоне»<sup>196</sup>.

Моментальный перенос точки зрения становится поводом к «озарению» *Деннета*. Он столкнулся с подтверждением нематериальности души! «Ведь когда последняя радиосвязь между Тулсой и Хьюстоном прекратилась, я поменял местоположение со скоростью света! И моя масса при этом не увеличилась... Я не мог себе представить, как философ-

---

<sup>196</sup> Ibid. С. 198.

физикалист мог бы это опровергнуть, не прибегнув к крайней, противоречащей здравому смыслу мере — запрещению всякого упоминания о личностях»<sup>197</sup>. Убеждение героя в субстанциональной нематериальной сущности «Я» вряд ли разделяет Деннет-автор. Но есть ли у последнего другой способ объяснить мгновенные перемещения личности героя? Пока мы размышляем над этим вопросом, герой радуется своей интеллектуальной находке: «я радовался своему философскому открытию и ломал голову... над тем, как сообщить о нем в журналы»<sup>198</sup>. К счастью, возможность установить связь с внешним миром и поделиться впечатлениями у него тоже вскоре появилась.

*Конфигурация 2. Йорик (старый мозг) и Фортинбрас (новое тело)*

Ученые погружают мозг *Деннета* в сон, а после пробуждения он обнаруживает, что обрел новое тело. «...Посмотрев в зеркало, я удивился, увидав там незнакомца. Он был с бородой и немного толще меня». Да, Гамлет остался в подземной шахте, а у *Деннета* — новое тело. Его наш герой именует, конечно, Фортинбрасом. Оно также оснащено радиоконтактами. Поэтому Фортинбрас, как и Гамлет ранее, управляется Йориком дистанционно. Вскоре *Деннет* свыкается с новым обликом, привыкает к изменившемуся голосу, мускулатуре. Отражение в зеркале более не удивляет его — лицо несет все тот же оттенок светлого ума и решительного характера. Ведь со сменой тела личность не меняется. И вот, в один из дней он решает навестить Йорика, старого беднягу Йорика, который, как и прежде, находится в чане в Хьюстонской лаборатории.

«Я решил, что добрый старый Йорик заслуживает, чтобы его навестили. Я и мое новое тело... вошли в знакомую лаборатории под аплодисменты сотрудников». Перед *Деннетом* — прозрачный чан, в нем — мозг. А рядом — выключатель, который управляет системой радиопередатчиков между мозгом и телом. «Повинуясь случайному капризу,

---

<sup>197</sup> Ibid. С. 198.

<sup>198</sup> Хофштадтер Д., Деннет Д., Глаз разума. М.: Бахрах-М, 2003 г. С. 198.



я протянул руку и перебросил переключатель в положение "Выкл". Представьте себе мое удивление, когда ничего особенного не случилось. Я не зашатался, не упал без сознания... — я не почувствовал ничего!»<sup>199</sup>. *Деннет* продолжает стоять напротив чана с собственным мозгом и размышлять, в то время как связь между ними, между Йориком и Фортинбрасом, оказалась нарушена. Что случилось?

*Конфигурация 3. Йорик (старый мозг) и Фортинбрас / Губерт (новый мозг) и Фортинбрас, Йорик = Губерт*

Еще до того, как *Деннета* оперировали в первый раз, был изготовлен компьютерный дубликат его мозга. Была написана программа, которая работала в точности как независимая функциональная копия его мозга. Если на входе она получала аналогичные данные, сигналы на выходе ничем не отличались от тех, что произвёл бы мозг *Деннета* до первой операции. Программу подключили к радиопередатчикам Фортинбраса, так что теперь она получала всю ту же информацию, что и Йорик. Подобно мозгу, она обрабатывала данные и выдавала соответствующую реакцию. Только эта реакция не передавалась Фортинбрасу, а записывалась и сравнивалась. Наблюдение за синхронностью работы Йорика и Губерта — так назвали программу ученые, не озабоченные Шекспировскими параллелями — длилось несколько месяцев, так что ученые смогли убедиться в абсолютной идентичности и независимости их работы. Тогда они решили совершить переключение. Когда *Деннет* прибыл в лабораторию, контроль над его телом уже осуществлял Губерт. Поэтому выключение радиопередатчиков Йорика никак не повлияло на поведение Фортинбраса.

В ходе мысленного эксперимента автор представил несколько возможных конфигураций субъекта: 1) *Деннет*, 2) Йорик — Гамлет, мозг и тело, соединенные дистанционно радиосигналами 3) Йорик — Фортинбрас, мозг и новое тело, соединенные дистанционно с помощью радиосигналов. И вот теперь 4) Йорик, Губерт и Фортинбрас: одно новое тело и два мозга —

---

<sup>199</sup> Ibid. С. 199.

искусственный и органический — работающие синхронно, одновременно получающие сигналы от органов чувств и попеременно выполняющие функцию управления организмом. Мыслительных центров у личности стало два, но, кажется, это не повлияло на количество личностей.

Изменения положения тумблера (Йорик→Губерт, Губерт→Йорик) проходят незаметно и для окружающих, и для самого субъекта. Ведь точка зрения не смещается, и восприятия — одни и те же. «Я обнаружил, что, не считая щелчка, я не ощущал совершенно никакой разницы. Я мог повернуть тумблер в середине предложения, и фраза, которую я начал под контролем Йорика, завершалась без малейшей паузы под контролем Губерта»<sup>200</sup>. При изменении положения тумблера сохраняется тождество личности. Находясь и под управлением Йорика, и под управлением Губерта, *Деннет* не перестает быть одним *Деннетом*. У него остаются те же воспоминания, такой же характер, те же цели и желания. Значит, теперь у *Деннета* есть полноценный запасной мозг. *Деннет* — одна личность с двумя независимыми мозгами. «Я оказался обладателем искусственного мозга, который мог оказаться мне весьма полезен, если бы в будущем с Йориком что-нибудь случилось. Я также мог держать Йорика про запас и использовать Губерта. По-видимому, было совершенно безразлично, которого из двух я выбирал»<sup>201</sup>. Сюжетная линия мысленного эксперимента заставляет нас вообразить человека с двумя мыслительными центрами, работающими синхронно. И поверить, что они формируют одну личность.

Эта конфигурация напоминает трансплантацию психики. Но специфична тем, что вместо биологической модели реализации психика реализуется в программе компьютера. И к тому же новый компьютерный мозг продолжает сохранять связь с Фортинбрасом. Это вполне экзотичный

---

<sup>200</sup> Ibid. С. 199.

<sup>201</sup> Ibid. С. 200.

вариант. Но это — не последняя метаморфоза в злключениях главного героя.

*Конфигурация 4. Йорик и Фортинбрас / Губерт и Фортинбрас, Йорик ≠ Губерт; New fission case*

Вполне ожидаемым продолжением сюжета могло быть наделение Губерта отдельным телом, скажем, Розенкранцом или Гильденштерном. Это было бы гуманно. Но автор избирает более экзотическую развязку. Это распад личности *Деннета* при сохранении одного тела. «Слава Богу! Я думал, что до этого дело так и не дойдет! Ты не можешь себе представить, насколько ужасны были эти две последних недели — но теперь ты испытаешь это на собственной шкуре, поскольку наступила твоя очередь отправляться в чистилище», — такие слова вдруг произнес *Деннет* после очередного переключения тумблера. Дело в том, что около двух недель назад мозги перестали работать одинаково. Может, в код программы закралась какая-то ошибка или сбой произошел в радиопередатчиках, но состояния Губерта стали отличаться от состояний Йорика. «В мгновение ока иллюзия того, что я контролирую мое тело — наше тело — оказалась полностью нарушенной... Я чувствовал себя так, словно меня носили в клетке, или, точнее, словно я был одержим. Я слышал, как мой собственный голос произносил слова, которых я не хотел говорить, я в отчаянии наблюдал, как мои руки делали то, чего я не хотел делать...»<sup>202</sup>. Пятая, последняя представленная нам конфигурация: Йорик, Губерт и Фортинбрас. При этом Йорик и Губерт функционально различаются. В этой пятой конфигурации *Деннет-герой*, в представлении автора, распадается на две личности, два различных мыслительных потока, две независимые и несовпадающие воли, вынужденные делить одно тело. Хотя от органов чувств Фортинбраса на оба мыслительных центра все еще поступают идентичные сигналы, реакции Йорика и Губерта различны. И один из них вынужден быть узником, пока тумблер не перейдет в новое положение или пока ему не подберут новое

---

<sup>202</sup> Ibid. С. 202.

тело. Рассказ заканчивается обещаниями одного *Деннета* найти подходящее тело другому. Ну, пока он (они) заняты поисками, мы попробуем добыть философские выводы.

История «Где Я?» — это художественное изложение проблемы Исчезающей личности. И это мысленный эксперимент, в котором слышен голос самого агента, с которым происходят метаморфозы. История проблематизирует то, что кажется простым и очевидным — местоположение личности. Деннету удастся представить это в виде проблемы потому, что он «разбирает» личность на модули и разносит их дистанционно. Но проблема, как было описано выше, в скрытом виде присутствует и в стандартной ситуации. Просто компактное расположение модулей прикрывает проблему локализации личности. Когда на скамейке сидят три человека, можно с большой вероятностью утверждать, что на скамейке находятся три личности. Но где эти личности находятся, с точностью сказать невозможно. Стоит направить на человека увеличительные приборы, личность вдруг окажется невидимой и неуловимой, как и в деннетовской истории. Ведь даже в обычной жизни функции агента распределены, просто распределённость их не так масштабна. Дистанцирование мозга от тела заставляет героя и читателей задуматься: с каким из модулей правильней ассоциировать личность? Локализация когнитивных функций склоняет интуиции в пользу ассоциации с мозгом, а расположение органов чувств — в сторону ассоциации с телом. Но обе альтернативы, скорее всего, ложные. Это становится понятно, когда в конфигурацию распределенного агента добавляется запасной электронный мозг. Теперь альтернатив становится три, а в перспективе их бесконечное множество, ведь копий может быть сколько угодно. К тому же личность удивительным образом может мгновенно перемещаться между всеми этими модулями. Как объяснить эти перемещения, не покидая пределов физикализма?

Еще более запутанной ситуация становится, когда мозги *Деннета* рассинхронизовываются. Вдобавок к локализации, теперь очевидной

становится и проблема идентификации. Теперь не ясно не только где локализуется личность, но и кто из личностей — *Деннет*. Объяснить мгновенные перемещения героя можно, если считать личность не конкретной, физической, а абстрактной сущностью. Именно к этой идее подталкивает читателя автор. Абстрактные сущности нереперенциальны, т.е. не отсылают к каким-либо конкретным физическим объектам. И потому как их поведение как в мире, так и их описание сильно отличается. Меры длины, метры и километры, к примеру, необходимы для характеристики расстояния. Их использование вполне законно и не противоречит физикализму. Но было бы ошибочно искать их где-то на дороге, считать количество таких характеристик или заключать об их тождестве между собой.

Атомы и пустота — не единственное, что признает наука. В ее онтологии есть место и для более странных вещей. Например, центр тяжести — понятие, которое составляет важную часть ньютоновской физики. Центр тяжести — не какая-то конкретная частица, не атом. Это условная точка, пространственная координата. У нее нет ни запаха, ни цвета, ни плотности, ни массы, нет никаких физических свойств. И тем не менее, она необходима для того, чтобы точно предсказать движение объекта. Она даже фигурирует в каузальных объяснениях. В этом смысле она является реальным объектом, реальной объективной причиной событий. Центр тяжести может совпадать с геометрическим центром предмета, а может быть вне предмета. Центр тяжести Земли — где-то во внутреннем ядре, а центр тяжести бублика — в дырке. Перемещаться центр тяжести может не только постепенно, но и мгновенно, минуя промежуточные этапы. Такова экзотическая судьба абстрактных сущностей. Если следовать аналогии и предположить, что личность — это что-то вроде центра тяжести, можно без труда и без дуализма объяснить феномен, с которым столкнулся герой деннетовского мысленного эксперимента — мгновенное перенесение личности. Впрочем, если личность сходна с абстрактным понятием, как могло возникнуть это

абстрактное представление? В книге «Объясненное сознание» Деннет предлагает свою концептуальную схему возникновения личности.

## 2. Личность как результат эволюции

### А. Появление границ, предпочтений и агентов

Некоторое время назад не было ни агентов, и личностей. В природе были причины, но отсутствовали мотивы. Было разнообразие, но не было лучшего и худшего. Не было целей, так как некому их было иметь. Существенные изменения произошли с появлением химических структур, способных к самокопированию, — начинает свою историю Деннет<sup>203</sup>. Это были простейшие живые организмы. Самокопирование было их ключевым свойством. Воспроизведение собственных копий стало их первым «интересом». Естественно, этот интерес у них был не осознан, но он объективно проявлялся в их поведении. Ведь выживать стали те, у кого была лучше способность к копированию. Они копировались чаще, и те копии, что производились с еще большей предрасположенностью копироваться, воспроизводились быстрее и в большем масштабе. То, что мешало воспроизводству и препятствовало работе механизмов самокопирования, стало «отрицательным», а то, что способствовало, — «положительным». Так появились предпочтения и «точка зрения». Конечно, эта точка зрения была внутренне пустой, «необитаемой». Но она уже существовала как некое место, с которого могла бы быть рассмотрена ситуация или... рассказана история<sup>204</sup>.

Примитивные системы, какими были первые репликаторы, не интериоризировали, не осознавали собственные предпочтения. Но то, что могло бы быть описано как их предпочтения, явно присутствовало в их поведении. Оно стало выражать рациональную схему. Поведение репликаторов возможно было интерпретировать как интенциональное, преднамеренное. Рациональность появилась как объективная схема. Это

---

<sup>203</sup> Dennett D. *Consciousness Explained*. Boston, Massachusetts: Little Brown & Co, 1991. P. 172.

<sup>204</sup> Подобную формулировку можно найти в Размышлениях Деннета к истории Р. Ракера «Программы» в сборнике «Глаз разума». Хофштадтер Д., Деннет Д. Глаз разума. М.: Бахрах-М, 2003. С. 238.

была независимая от агента, свободная рациональность. И эта рациональность не могла существовать без определения границ между организмом и окружающей его средой.

Если главной задачей является производство как можно большего числа копий, сначала нужно определиться, что копировать, а что — нет. Невозможно же скопировать весь мир. К тому же для создания копий необходимо обеспечить стабильность, безопасность процессу копирования. Так возникла потребность в границах, разделяющих внутреннее и внешнее. «Как только твоим основным делом становится самосохранение, границы становятся необходимыми, т.к. если ты хочешь сохранить себя, ты не должен растратить все силы, сохраняя весь мир»<sup>205</sup>.

В мире неживой природы границ нет. Безразлично, где заканчивается один камень и начинается другой. Разлом или трещина может проходить здесь или там — никто не будет вынужден защищаться или отступать. Совсем другое дело с живыми существами. Они вынуждены следить за собственными границами. Иначе бы разрушилось их функциональное единство. Так что были созданы сложные механизмы защиты: шерсть, чешуя, панцири, зубы и клешни. Иммунная система — это тоже инструмент защиты границ, целая армия из миллионов различных антител, чьей основной задачей является защита организма от незваных и вредных для здоровья гостей. Правда, все механизмы защиты должны были быть направлены вовне, а не против себя. Как биологический организм мог определить, где его границы? Скорее всего, эта способность жестко запрограммирована эволюционным процессом. И она присутствовала у самых простых организмов. У всех организмов должно существовать рудиментарное представление о себе, о своей структуре. Иначе они не смогли бы адекватно реагировать на изменения во внешней среде. Эта рудиментарная схема и есть их собственное Я, простейшее биологическое Я.

---

<sup>205</sup> Dennett D. *Consciousness Explained*. Boston, Massachusetts: Little Brown & Co, 1991. P. 173.

### *В. Личность как расширенный фенотип*

Существование границ и механизмов защиты у организмов не означало полной непроницаемости между внешним и внутренним. В организмах всегда были и есть пришельцы: это вирусы и бактерии. Эти пришельцы, сами по себе являются автономными системами, у них есть собственные границы. Часть из них паразитируют на организме хозяина. А некоторые не только безвредны, но и необходимы для нормального функционирования последнего. Некоторые бактерии, например, населяющие пищеварительную систему, позволяют усваивать пищу, а другие — вырабатывают полезные вещества. Даже митохондрии, основной источник энергии в наших клетках, скорее всего, некогда являлись «посторонними» — бактериями, которые были поглощены, но нашли вариант взаимовыгодного симбиоза с организмом-хозяином. Границы организма были проницаемы и в другую сторону — вовне. Ведь пространство агента может распространяться за пределы его физического тела<sup>206</sup>.

Сами живые существа проводили экспансию во внешний мир. Многие виды приспособились осваивать окружающее пространство так, что оно как бы стало их естественным продолжением. Улитка накапливает кальций, чтобы потом выделять его вовне и формировать твердую ракушку, а рак-отшельник присваивает уже готовую ракушку, пауки плетут паутины, а бобры строят плотины. Эти артефакты нужны им для того, чтобы повысить свои шансы на выживание и, как показал уже упоминавшийся биолог-эволюционист Ричард Докинз, они также являются объектами естественного отбора. Они представляют так называемый «расширенный фенотип».

«Расширенный фенотип» может продлевать территориальные границы организма на несколько сантиметров (как в случае с улиткой), а может — на километры (как в случае с границами заводи после строительства плотины бобром); он может быть относительно простым, а может быть

---

<sup>206</sup> Dennett D. *Consciousness Explained*. Boston, Massachusetts: Little Brown & Co, 1991. P. 413.



необыкновенно искусно сделанным, как, например, термитник или гнезда австралийского шалашника. Он может включать в себя неорганические предметы, а может включать даже особей того же вида, как в случае с колонией муравьев. Но самые удивительные и сложные конструкции создаются человеком.

«Мозг человека создает уникальную по своей сложности паутину из поступков и текстов»<sup>207</sup>. Как и искусственные артефакты животных, эта сеть — личная история — выполняет защитную функцию и увеличивает шансы организма на репродукцию. И как большинство артефактов животных, она собрана из элементов (и событий), уже существующих в окружающей среде. Это универсальная тактика самозащиты, самоконтроля и самоопределения. Язык и образованный на его основе биографический текст являются основой для появления личности — наиболее удивительного вида «расширенного фенотипа» среди живых существ. Он создает личность как тренд, как геометрический центр биографических историй, как условную точку, куда указывает стрелка компаса поступков. Личность, согласно емкому определению Деннета, это *центр нарративной гравитации*. «Наши истории закручиваются, но в большей степени не мы закручиваем их, а они — нас. Наше человеческое сознание и наша нарративная личность — это продукты, а не источники»<sup>208</sup>, — пишет американский философ.

Как и любое биологическое существо, человек и его мозг — это система из множества модулей. Ни один из этих модулей не обладает всеми уникальными качествами личности. Ни одному из них не могут быть приписаны желания, убеждения, действия агента и ответственность за них. Но в результате совместной работы всех этих модулей организм демонстрирует сложное целенаправленное поведение, которое и можно назвать поведением личности. Ее заметно с высоты птичьего полета, с

---

<sup>207</sup> Dennett D. *Consciousness Explained*. Boston, Massachusetts: Little Brown & Co, 1991. P. 415.

<sup>208</sup> Dennett D.C. *Consciousness Explained*. Boston, 1991. P. 418.

позиции автобиографической истории. Текст происходит как будто из одного источника, так что создается ощущение единого агента, который создает эту историю.

История происхождения агентов, собственных *Я*, личности, в том виде, как эта история была представлена выше, конечно же, — набросок, модель, воспроизводящая логику эволюционного процесса. В этом сюжете нет подробностей и исторических дат. Тем не менее, эта история позволяет наглядно проследить, как из неживой природы с помощью одного только слепого алгоритма могли возникнуть системы с собственной точкой зрения, границами, саморепрезентацией и даже автобиографией. Эта история показывает, что «личность» может быть продуктом биологического организма, результатом адаптации, лишь по уровню сложности, а не по принципу, отличающимся от других искусственных артефактов, созданных живыми существами. И возникновение такой личности естественно. Оно не требует осознанного плана по ее созданию, как плетение паутины не требует от паука интеллектуального замысла, как строительство термитника не требует архитектурного проекта у термитов, как работа иммунной системы не предполагает медицинских знаний у антител. Но из всего этого не следует, что личность — это что-то обыкновенное. Да, в её основе лежит рудиментарное представление о себе, саморепрезентация, которая присутствует почти у всех живых организмов. И она является одним из вариантов расширенного фенотипа, который встречается даже у муравьев, бобров и пауков. Но при этом кажется, что личности все-таки принципиально отличаются от расширенных фенотипов других живых организмов. В чем заключаются эти отличия?

### 3. Критерии личности

«Можно предположить, что такое существенное понятие, применяемое нами с уверенностью, имеет ясно сформулированные необходимые и достаточные критерии, но если эти критерии и существуют, они пока не

выявлены», — пишет Деннет в статье «Условия личности»<sup>209</sup>. Более того, философ считает, что их полный список в принципе не может быть составлен. Для того, чтобы показать это, он идет «от противного». Деннет составляет список, который должен исчерпать достаточные и необходимые критерии личности, а потом критикует его, показывая, что не только этот список, но и любой другой не сможет подойти для такой цели. Свой аргумент он начинает с общих наблюдений.

Класс личностей, — замечает Деннет, — по всей видимости, примерно совпадает с классом людей. Но связь этих классов не основана на определении. Личность — это понятие, относящееся не к биологическому виду, а, скорее, к функциональному. Другие биологические виды (и даже механические существа) теоретически могли бы тоже обладать этой характеристикой. Однако требования к наличию личности, кажется, специфичны и высоки, так что другие существа, известные до настоящего времени, даже высшие млекопитающие, такие как собаки, обезьяны и дельфины, скорее всего, не соответствуют этим требованиям. Более того, даже некоторые люди не проходят по этим критериям, например, младенцы и люди, страдающие психическими расстройствами. Только к части людей мы относим понятие личностей.

Кажется, критерии личности бывают двух типов: этические и метафизические. Уже в трактатах прошлого философы обнаружили это. Так, Локк писал, что личность — «это юридический термин, касающийся действий и их ценности и относящийся поэтому только к разумным существам, знающим, что такое закон, счастье и несчастье. Эта личность простирает себя за пределы настоящего существования, к прошлому только силой сознания; вследствие этого она беспокоится о прошлых действиях, становится ответственной за них, признавая за свои и приписывая их себе

---

<sup>209</sup> Dennett D. Conditions of Personhood // Rorty A. (ed.). Identities of Persons. Berkley: University of California Press, 1976. P. 178.

совершенно на том же самом основании и по той же причине, что и настоящие действия»<sup>210</sup>. Этический критерий личности — принадлежность к обществу морально ответственных агентов<sup>211</sup>. Согласно этому критерию, если агент принадлежит к обществу морально ответственных агентов, то он — личность. С этим критерием нельзя не согласиться. Но он неинформативен. Получается круг: моральная ответственность предполагает существование личности, а личность понимается как то, что является основанием моральной ответственности. Для нетавтологического определения того, что такое личность, нужно использовать метафизические критерии.

Метафизические характеристики могут прояснить специфику личности. К метафизическим чертами личности Локк причисляет разумность, наличие сознания и памяти, знание добра и зла. Последнее можно называть моральной компетентностью. Эти условия кажутся вполне актуальными и сейчас, в особенности если им придать современные значения. В своей статье, посвященной понятию личности, «Условия личностности», Деннет как раз это и делает. Американский философ выделяет шесть основных условий: (1) рациональность (2) сознание (3) возможность быть объектом личностной установки (4) способность занимать личностную установку (5) язык (6) самосознание<sup>212</sup>.

(1) Разумность или рациональность — это закономерность, телеологическая схема поведения, обусловленная эволюционным процессом. Она присуща множеству живых существ. Она предполагает наличие интересов и предпочтений у организма, а также минимальной способности эти потребности удовлетворять.

---

<sup>210</sup> Локк Дж. Сочинения в 3-х т. Т. 1. Опыт о человеческом разумении. М.: Мысль, 1985. С. 400.

<sup>211</sup> Т.е. чтобы агент нес моральную ответственность, он должен быть личностью. Но «быть личностью» не означает быть морально ответственным за все поступки. Могут существовать условия, которые снимают с личности ответственность.

<sup>212</sup> Dennett D. Conditions of Personhood // Rorty A. (ed.) Identities of Persons. Berkley: University of California Press, 1976. P. 175.

- (2) Другая важная особенность — сознание. Это тоже общая черта для некоторых животных. Под сознанием Деннет в данном случае понимает не качественные состояния, а только возможность приписывать организму интенциональные содержания: ментальные или психологические характеристики, желания и убеждения. Это функциональная характеристика, определяемая по поведению.
- (3) Следующее условие — способность быть личностью в глазах других. Деннет предлагает любопытную инверсию. Он считает: не столько объективный факт того, что агент является личностью, заставляет других относиться к нему так, а наоборот: относясь к какому-то как к личности, мы устанавливаем некоторое положение дел в отношении этого агента. «Некто может считаться личностью или нет в зависимости от отношения, которое к этому некто имеют другие, в зависимости от установки, которая к нему применима»<sup>213</sup>.

Три первые характеристики, по мнению Деннета выделяют класс, который он обозначает как класс интенциональных систем. «Интенциональные системы — это системы, чье поведение может быть (как минимум иногда) объяснено и предсказано с помощью приписывания убеждений и желаний...»<sup>214</sup>. Класс интенциональных систем, по мнению Деннета, гораздо шире, чем класс личностей: в него попадают и животные, и даже некоторые машины. А вот следующие характеристики специфичны только для личностей.

- (4) Способность обходиться с другими как с личностями — это еще одна важная характеристика. Это способность признавать собственные интересы у другой личности, учитывать эти интересы при выборе своих действий. Чаще всего речь идет об уважительном отношении, поведении с учетом интересов и достоинства другой личности. Но не

---

<sup>213</sup> Dennett D. Conditions of Personhood // Rorty A. (ed.) Identities of Persons. Berkley: University of California Press, 1976. P. 177.

<sup>214</sup> Ibid. P. 179.

только. К личности также применимы и особые требования, в частности моральная ответственность. Даже некоторые формы преднамеренного крайне уничижительного отношения будут формой межличностных отношений. Личности по отношению к другим личностям могут проявлять особенную жестокость. Это тоже будет выражать специфическое личностное отношение.

- (5) Одним из важнейших свойств личности является язык. Он дает возможность интериоризировать рациональность. Тогда рациональность не только объективно прослеживается в поведении, в том, что существо имеет определенные цели и способно в поведении их добиваться, но и выражается в способности личность анализировать и планировать действия. Личности способны осуществлять предварительный отбор вариантов поведения внутри, в процессе рассуждений, и позволяют «гипотезам умирать вместо них самих»<sup>215</sup>. В отличие от других существ, выживающих только благодаря тому, что удачно выбрали вариант действия, личности не играют в слепую рулетку — их выбор надежнее, чем надежда на случай. Если живые существа в целом — просто удачливы, имея готовые адаптированные схемы, личности удачливы настолько, чтобы с помощью языковых выражений моделировать будущее.

(6) И, наконец, последнее, по мнению Деннета, условие — способность к самосознанию. Под самосознанием подразумевается представление о себе как об авторе собственных действий, способность описывать эти действия, наличие образа себя как длящейся во времени сущности. Эта способность напрямую связывает условие личности с условиями ответственности. В этом утверждении Деннет опирается на позицию Э. Энском, которая считала, что «для того, чтобы быть ответственным за действие... я должен осознавать это действие в этом

---

<sup>215</sup> Цит. по Dennett D. *Darwin's Dangerous Idea: Evolution and the Meaning of Life*. New York, 1995. P. 375.

описании»<sup>216</sup>, т.е. как своё собственное. Узнавание действия в описании обвиняющего или поощряющего как своего действия, в свою очередь, необходимо для того, чтобы иметь возможность объяснить его с собственной позиции, что, в свою очередь, необходимо для ответственности. Ведь если человек не может описать свое действие или опознать его в некоторой формулировке, значит, он не имеет особого привилегированного доступа к мотивам этого действия, не может признать это действие своим и, соответственно, нет оснований считать именно его ответственным за это действие, в частности требовать извинений.

Приведенный выше список в целом характеризует способности обычного здорового взрослого человека. И представляется, что он походит для критериев личности. Все перечисленные условия необходимы для того, чтобы определить личность. Но, по мнению философа, они все равно не работают как достаточные критерии. Ни этот, ни какой-либо другой сходный список работать не будет. Это связано с тем, что все перечисленные (и другие возможные) условия личности нормативны по своей сути. Они представляют абстрактный идеал, направление движения. Но не позволяют провести четкую объективную разграничительную линию. «Люди или другие сущности могут только стремиться приближаться к идеалу, и не может быть "зачетного балла", который не был бы произвольным»<sup>217</sup>. Рациональность, предсказуемость с позиции интенциональной установки, самосознание — все это идеальные модели. Ни одна личность не соответствует всегда этим идеальным схемам, и поэтому наше суждение о принадлежности кого-то к классу личностей, к классу ответственных агентов все равно будет достаточно произвольным. «Личности в абсолютном смысле» не существует также, как не существует «ответственности в конечном счете». Для того,

---

<sup>216</sup> Dennett D. Conditions of Personhood // Rorty A. (ed.) Identities of Persons. Berkley: University of California Press, 1976. P. 191.

<sup>217</sup> Dennett D. Conditions of Personhood // Rorty A. (ed.) Identities of Persons. Berkley: University of California Press, 1976. P. 193.

чтобы убедиться, насколько тонка грань между личностями и теми, кто находится за пределами этой категории, можно обратиться к случаям умеренных отклонений от норм: взрослых с признаками аутизма, умственной отсталости и психопатии.

*А. Психопатия как иллюстрация субъективности границ класса личностей*

Особенно полезен в качестве иллюстрации феномен психопатии<sup>218</sup>.

Именно в отношении этого феномена проявляется коллизия юридической и моральной трактовки ответственности. Множество современных аналитических философов и психологов (Г. Ватсон, Д. Шумейкер, Н. Леви, С. Морс) считают, что психопаты не являются полностью ответственными моральными агентами. При этом в юридическом поле ответственность психопатов почти не подвергается сомнению. К психопатам применяются даже самые жесткие меры наказания, вплоть до смертной казни и пожизненного заключения. Значит ли это, что соответствующие законы сильно устарели и необходимо смягчить меры наказаний для психопатов? Или, наоборот, что критерии для «общества моральных агентов» философами применяются слишком строго? Наконец, может, допустима коллизия юридической и моральной трактовки ответственности?

Мне кажется, что этой коллизии быть не должно. Моральная и юридическая ответственность строятся на общих основаниях. В системе юридической ответственности наказание выполняет четыре основные функции. Эти функции: устрашение, ограничение дальнейшего возможного ущерба, реабилитация и воздаяние. Первые три функции имеют утилитаристское, прагматическое, направленное в будущее значение. Эти функции тривиальны и могут быть реализованы не только в отношении личностей, но любых агентов.

---

<sup>218</sup> Психопатия как термин имеет широкое значение. Так, в справочниках по психиатрии иногда насчитывается до 11 типов психопатии. Здесь в качестве примера я анализирую наиболее часто используемый в литературе (в частности в работах наиболее известного канадского психолога Р. Хаэра) тип с характерными чертами. Этот тип психопатии лучше всего подходит для иллюстрации работы критериев личности Д. Деннета.



А вот последняя, четвертая функция наказания специфична. Кажется, чтобы ее реализовать, агент, в отношении которого она реализуется, должен обладать особыми свойствами. Он должен обладать компетенциями морального характера. Функция воздаяния имеет деонтологическое значение. Она направлена на установление справедливости как таковой, она направлена в прошлое. Но невозможно вершить справедливость в отношении того, кто абсолютно невосприимчив к категории справедливости. Аналогичное мы видим и в отношении моральной ответственности. Моральная ответственность предполагает наличие у агента некоторой моральной компетентности. Соответственно, юридическая и моральная ответственность применимы в отношении одной особой группы агентов. Агенты, не подходящие для воздаяния, не могут быть ни морально, ни юридически ответственными. А значит, существующая коллизия должна быть разрешена. Но в чью пользу? Оправдано ли освободить психопатов от моральной и юридической ответственности, отказав им в статусе личности? Или нужно санкционировать применение этих видов ответственности в отношении психопатов?

#### *В. Психопатия как расстройство личности*

Психопатия — это разновидность «диссоциального расстройства личности». Это не столько болезнь, сколько тип характера. Диагностика психопатии осуществляется с помощью вопросника, который направлен на выявление определенных стереотипов поведения. Среди этих стереотипов — ненадежность, лживость, отсутствие переживаний вины и стыда, крайняя эгоцентричность и убежденность в собственной значимости, отсутствие склонности к глубоким и долгосрочным межличностным отношениям, отсутствие долгосрочного планирования, склонность манипулировать людьми, отсутствие сопереживания, импульсивность и проч. У психопатов отсутствует эмпатия — способность понимать чувства других и сопереживать другим людям. Психопаты, судя по всему, не переживают

«сложных эмоций»: чувства вины, стыда, сожаления. Вследствие этого психопаты часто совершают аморальные поступки и преступления.

При этом основные когнитивные функции у этих преступников соответствуют норме. У них полностью развито мышление, память, существует система ценностей. Только система ценностей у них эгоцентрическая. Преступления психопаты совершают не под воздействием бредовых состояний, а полностью отдавая себе в них отчёт, преследуя собственные цели. Мотивом антисоциального поведения психопатов является обогащение, повышение статуса или удовлетворение сексуальных желаний<sup>219</sup>, т.е. вполне «рациональные» соображения.

Причины психопатии изучаются. Этими причинами могут быть особенности воспитания и физиология. Среди нейрофизиологических причин наиболее вероятной является дисфункция миндалин и глазнично-лобных участков коры головного мозга<sup>220</sup>. Эти участки отвечают за условно-рефлекторное отвращение и инструментальное обучение. И обе функции нужны для адаптации в обществе. Например, чтобы ребенок научился тому, что бить других — плохо, нужно, чтобы драка ассоциировалась у него с негативными реакциями: слезами, плачем, криками и страданием на лице жертвы. Так должно происходить отучение и от «аморальных действий» — через отрицательный стимул — отвращение к страданиям жертвы. Но у психопатов вид чужого страдания не вызывает отвращения. У них — дисфункция условно-рефлекторного отвращения. Поскольку они не испытывают отвращения к страданиям жертвы, они не научаются избегать подобных ситуаций. Они абсолютно глухи в отношении чужого горя и невосприимчивы к моральному закону. Поэтому с легкостью его переступают — конечно, когда считают, что это сойдёт им с рук.

---

<sup>219</sup> Cornell D., Warren J., Hawk G. Psychopathy in Instrumental and Reactive Violent Offenders // Journal of Consulting and Clinical Psychology. 1996. № 64. P. 783-790.

<sup>220</sup> James R., Blair R. Neurobiological Basis of Psychopathy // The British Journal of Psychiatry. 2003. № 182. P. 5-7.

*Отсутствие сопереживания.* «Переживаю ли я за других людей? Это трудный вопрос. Ну, да, конечно, я переживаю... но не даю чувствам быть помехой... Я, конечно, такой же сердечный и заботливый, как любой другой, но давайте говорить на чистоту, все хотят тебя кинуть... Я должен позаботиться о себе сам, припарковать свои чувства, так сказать. К примеру, тебе что-то нужно или кто-то с тобой плохо обходится, хочет содрать с тебя двойную цену... приходится позаботиться о себе... ну, сделать то, что должно быть сделано. Чувствую ли я себя плохо, если причиняю боль? Да, иногда. Ну, это... как... [смешок]... как ты чувствовал себя в последний раз, когда раздавливал клопа»<sup>221</sup>. Это фрагмент из интервью, взятый у Джо, психопата, отбывающего срок за похищение людей, изнасилование и вымогательство. В его собственных глазах он –самый обычный человек. И все его поступки — самые обычные и самые логичные. По его мнению, на его месте так поступал бы каждый. Просто ему не повезло в этот раз. Возможно, повезет в следующий.

У психопатов отсутствует беспокойство, угрызение совести и стыд за причиненный ими вред. Их не волнуют последствия их действий для других. Участник ограбления, Крис, психопат, ранивший ножом случайную жертву, так комментирует свое преступление: «Да вы с Луны что ли? Он проваляется несколько месяцев в больнице, а я гнию здесь. Я его немного порезал, но если бы я хотел его прикончить, я бы перерезал ему глотку. Я такой человек. Я его просто пожалел». Как видно, этот человек ни о чем не жалеет. Для него что сделано, то сделано, и на то были причины. И это типично для психопатов. А вот само откровенное признание — это, скорее, редкость. Об отсутствии чувства вины и сопереживания правильной всего заключать из повторяющихся паттернов поведения, бесконечных рецидивов преступлений. Ведь психопаты часто умело симулируют раскаяние. Ложность раскаяния обнаруживается только потом, в поступках и других высказываниях.

---

<sup>221</sup> Hare R. Without Conscience: The Disturbing World of the Psychopaths Among Us. N.Y., 1993. P. 34.

Притворство, хамелеонство — одно из наиболее развитых качеств психопатов.

*Хамелеонство.* Р. Хаэр, американский психиатр, разработавший тот самый вопросник для диагностики психопатии, в своей книге «Лишенные совести. Пугающий мир психопатов» приводит такой случай. В начале практики он устроился работать тюремным психологом. Его первым посетителем был Рэй. Рэй начал разговор с того, что попросил помощи: «Док, как дела? Слушай, у меня есть проблема. Мне нужна помощь. Я бы хотел поговорить на этот счёт». После этого Рэй вытащил нож. Проблема Рэя была якобы в том, что к его «другу» приставал другой заключенный, и Рэй хотел тому отомстить. По этому поводу он, по его же словам, хотел посоветоваться с психологом. Но Рэй лукавил. В действительности он тестировал нового сотрудника Колумбийской тюрьмы. Заключенным запрещалось иметь оружие, и перед Хаэром Рэй поставил непростой выбор: доложить начальству и разрушить свою репутацию перед всеми заключенными, или сохранить тайну и убедить Рэя не применять нож. Психолог выбрал последнее, и... нарушив тюремный режим, сам стал жертвой манипуляций Рэя. Сначала тот убедил доверчивого психиатра, что его призвание — быть поваром. По ходатайству Хаэра Рэй был переведен для работы на кухню. Но поваром Рэй становиться не торопился. Получив доступ к запасам сахара, картошки и фруктов, он изготовил самогонный аппарат и стал производить алкоголь. Когда аппарат через некоторое время обнаружили, Рэй, как будто ничего не случилось, вновь появился в кабинете у Хаэра, на этот раз с просьбой о переводе в авторемонтный цех. Он, по его словам, чувствовал в себе способности мастера «и должен был готовиться к жизни за стенами тюрьмы». В конце концов Рэю опять удалось уговорить Хаэра. За месяц до ухода Хаэра с позиции тюремного психолога, от Рэя поступила очередная просьба — рекомендовать его на работу к своему отцу. Рэй врал настолько искусно, что мог убедить любого, даже самого бывалого собеседника. К моменту знакомства с Хаэром за ним числилось множество

преступлений, в том числе с применением насилия. Он половину жизни провел в тюремной камере, и тем не менее, почти уговорил Хаэра. Когда тот, все-таки отказал, Рэй разразился ругательствами. По дороге с работы у Хаэра в машине отказали тормоза — оказался надрезанным провод в тормозной системе. Накануне автомобиль ремонтировался в той самой мастерской, куда перевели работать Рэя.

Но спустя несколько лет им предстояла еще одна встреча, когда Хаэр уже преподавал в университете Британской Колумбии. В один из приемных дней он вдруг услышал уже почти забытый голос: «Да, я был помощником Доктора Хаэра, когда он работал в тюрьме Британской Колумбии. Я помогал ему с документами, он обсуждал со мной все трудные случаи». Это был Рэй. Он пытался записаться на какой-то из курсов. Увидев Хаэра, он ничуть не смутился: «А, привет, док!» И тут же продолжил, как ни в чем не бывало. Его не волновало, что ложь вскрывалась. Он просто менял тему или искал другую жертву. Так поступают психопаты. Хамелеонство и лживость наряду с обаянием — их типичные черты. Они согласуются с их представлением о других людях как о средствах, инструментах в достижении корыстных целей. Возможно ли в таком случае относить психопатов к категории личностей? Как это соотносится с критериями, предложенными Деннетом?

По всей видимости, психопаты удовлетворяют большей части критериев, обозначенных американским философом. Они рациональны, сознательны, они свободно владеют языком (и владеют им иногда даже лучше нормальных людей), к ним относятся как к личностям. Вызывать сомнения могут только два пункта: (4) способность обходиться с другими людьми как с личностями, что является неотъемлемой характеристикой личностей, и (6) самосознание. Сторонник лишения психопатов ответственности и статуса личностей мог бы сослаться на два этих пункта. Но мне кажется, что его позиция все же не имеет достаточно надежных оснований: психопаты в принципе понимают, как нужно обходиться с людьми, и осознают свои поступки в этическом плане.

В одном из примеров Джо явно показал, что достаточно остро переживает, когда с ним самим обходятся несправедливо, «хотят содрать с него двойную цену». И этим обосновывал своё жестокое ответное поведение. Судя по другим примерам, психопаты также способны понимать интересы и намерения других людей и даже иногда учитывают их в корыстных целях. Это видно, в частности, из ситуаций с Реем. Так, он, манипулируя Хаэром, понимал, как, скорее всего, поведет себя новый тюремный психолог. Он догадывался, что Хаэр заинтересован в установлении доверительных отношений с заключенными даже больше, чем в соблюдении тюремных правил. И именно на этом строил свой план действий. Исходя из этого, можно заключить, что психопаты в целом способны понимать, как следует обращаться с другими личностями. Ведь для этого достаточно в принципе иметь способность понимать интересы других и знать, как к ним следует относиться. Способности же понимать, как следует обращаться с другими личностями, с моей точки зрения достаточно, чтобы иметь способность относиться к другим как к личностям.

На это сторонник лишения психопатов статус моральных агентов мог бы возразить: пусть так, пусть иногда психопаты способны относиться к другим людям как к личностям; но они совершают преступления, и именно в этих конкретных ситуациях они не способны видеть в других людях личностей. А значит, они не способны относиться к ним как к личностям и вести себя по отношению к ним иначе в момент преступления. Значит, в момент совершения преступления они сами не являются личностями, и это освобождает их от ответственности. Как личность может нести ответственность за поступки, которые были совершены не личностью, а более простым агентом?

Проблема с этим возражением в том, что «способность поступать иначе» в нем определена слишком узко. Сторонник аргумента пытается показать, что именно в данном конкретном случае в момент совершения преступления способности поступить иначе у преступника нет. Но мне

кажется, что в лучшем случае он показывает, что в момент совершения преступления у преступника нет «возможности» поступать иначе, именно возможности, а не способности. По фактическим обстоятельствам мы не можем определять наличие или отсутствие способности. Способность определяется не фактическими обстоятельствами, а гипотетическими. К примеру, я сейчас сижу, но мог бы ходить. У меня есть способность как сидеть, так и ходить. Конечно, фиксируя все обстоятельства текущего момента, я не могу ходить, но на основании этого нельзя делать вывод о том, что у меня в данный момент такой способности нет. Таким образом, я не вижу достаточных оснований считать, что психопаты не способны относиться к другим как к личностям. Тогда как обстоит дело с самосознанием?

Думаю, можно выделить две основные части этой компетенции: когнитивную и эмоциональную. Мне кажется, что приведённые выше примеры свидетельствуют о наличии у психопатов когнитивной, эпистемической моральной компетенции. Джо в принципе понимает, что значит быть сердечным и заботливым. Крис имеет представление, хотя и вполне ущербное, о жалости, а Рэй — о справедливости и воздаянии и соответствующем поведении. Психопаты знают о моральных нормах, ссылаются на эти нормы в речи и способны соотносить их со своими поступками и поступками других людей. В этом смысле они различают добро и зло, истину и ложь. Но соображения о добре и истине сами по себе не являются для них достаточным основанием поступать сообразно им. Поэтому, скорее, не когнитивный, а эмоциональный аспект моральной компетенции отсутствует или функционирует со сбоями у психопатов. «Психопаты мало склонны испытывать эмоциональные реакции — страх и тревогу — являющиеся ключевым механизмом в работе совести»<sup>222</sup>. У них как бы приглушен внутренний голос, останавливающий нормального человека от совершения антисоциальных поступков. Точнее, этот

---

<sup>222</sup> Hare R. Without Conscience: The Disturbing World of the Psychopaths Among Us. N.Y., 1993. P. 34.

внутренний голос лишен убедительности, которую придаёт ему эмоциональная окраска. К примеру, когда Крис намеревается ударить ножом свою жертву, он, вполне вероятно, осознает: «если я его ударю, он будет страдать и истекать кровью». Но это осознание в эмоциональном плане для него почти равнозначно мысли: «если я возьму такси, это будет несколько дороже, чем ехать на метро». Поэтому Крис не воздерживается от преступления.

Убеждение, что у психопатов нарушен именно эмоциональный, мотивационный аспект осознания, разделяет много исследователей. Среди них — Д. Шумейкер, Дж. Мёрфи и Э. Дафф<sup>223</sup>. В силу этого порока они считают психопатов неприемлемой мишенью для моральных оценок и неподходящим объектом для межличностных отношений. Шумейкер так аргументирует свою позицию. Для того, чтобы обладать моральной компетенцией, быть объектом для действия моральных законов, агент должен быть способным действовать исходя из моральных соображений. Долженствование предполагает возможность. Психопаты способны совершать поступки, которые только выглядят как моральные, но их мотивация в любом случае эгоистична. Действия из моральных соображений предполагают способность к эмпатии, а она отсутствует у психопатов. Психопат не может искренне заботиться о другом человеке. Его интерес к другим исключительно «инструментальный». Ценность других людей для него определяется их полезностью. Поэтому психопаты неспособны действовать из моральных соображений. Они не обладают моральной компетенцией. В этом смысле они не участвуют в межличностных отношениях и, возможно, отстоят даже дальше от нормальных людей, чем некоторые животные. Шумейкер иллюстрирует этот аргумент воображаемой ситуацией.

---

<sup>223</sup> Shoemaker D. Moral Address, Moral Responsibility and the Boundaries of The Moral Community // Ethics. 2007. Vol. 118. No. 1. P. 70-108; Murphy J. Moral Death: A Kantian Essay on Psychopathy // Ethics. 1972. № 82. P. 284-298; Duff A. Psychopathy and Moral Understanding // American Philosophical Quarterly. 1977. № 14. P. 189-200.



### *С. Моральное чувство инопланетян*

Предположим, среди людей поселяются инопланетяне. Их моральные принципы в целом соответствуют человеческим, но распространяются на более широкий класс объектов. В частности, инопланетяне считают аморальным ходить по траве, т.к. трава от этого получает ущерб. Нанесение вреда траве само по себе плохо, — считают инопланетяне и мотивируют это тем, что знают, «что значит быть травой», которой нанесен физический ущерб. Когда люди им возражают, обращая внимание на то, что трава не обладает сознанием, инопланетяне отвечают, что это не важно. Для того, чтобы знать, что значит быть травой, нет необходимости переносить на себя её ментальные состояния. Достаточно «войти в её пространственное существование». Для людей эти основания, безусловно, представляют бессмыслицу.

Предположим теперь, что какой-то человек прогуливается по парку и видит вдали любопытную формацию горных пород. Из любопытства он решает подойти к ней поближе, чтобы лучше её разглядеть. Приближаясь к формации, человек наступает на траву. Он не может испытывать эмпатии к траве, поэтому не чувствует аморальности своего поступка. Но тут к нему в ярости приближается инопланетянин: «Как плохо вредить траве!!! Как ты мог совершить такой аморальный поступок?!!» Есть ли основания у инопланетянина считать человека морально ответственным за нанесение вреда траве? Нет, — считает Шумейкер, — ведь люди просто не способны понять, какой ценностью обладает неприкосновенность травы. Сколько бы инопланетяне не растолковывали людям о «вхождении в пространственное существование травы», это никак не убедит людей в их «моральном долге» перед травой. И психопаты — в таком же положении. Они неспособны воспринять мотивирующую силу моральных норм в отношении других. Кажется, что психопат в обществе других личностей должен чувствовать себя, как человек среди инопланетян со сверхчувственной моральной компетенцией: он должен понимать требования окружающих, но не видеть у

них оснований. И при первом возможном случае нарушить правила, если это останется безнаказанным. Психопат не обладает моральной компетенцией в отношении других людей и потому не является полноценной морально ответственной личностью.

Я не согласен с такими выводами из вышеописанного случая. Интуиции об отсутствии моральной ответственности в данной истории основаны на иррациональности требований инопланетян Шумейкера. «Пространственное существование» растений не удовлетворяет критериям рациональности оснований морального долга. Если под аналогичные требования подвести рациональные основания, даже если субъект действий не будет обладать соответствующим чувством эмпатии, мысленный эксперимент вызовет интуиции о наличии у него моральной ответственности. Для того, чтобы это подтвердить я предлагаю сравнить историю Шумейкера с новой историей. В ней требования рациональны, но субъект не испытывает эмпатии. Это — новая история про общину джайнов.

#### *D.    Община джайнов и ахимса*

Молодой человек присоединяется к общине джайнов. Философия и правила жизни джайнов направлены на совершенствование души и освобождение от круговорота рождений и смертей. Путь к совершенству лежит в соблюдении пяти принципов: быть искренним и благочестивым, не красть, не прелюбодействовать, не стяжать и не причинять вреда живому (ахимса). Из них пятый принцип — самый важный. Он запрещает джайнам наносить любой вред любому живому организму, даже насекомым и растениям. В основе этого принципа — убеждение в том, что любое существо наделено индивидуальной и вечной душой (дживой), способной к страданию и обладающей абсолютной ценностью. Соблюдая этот принцип, джайны трепетно относятся ко всему окружающему. Они избегают передвигаться ночью или даже пить непроцеженную воду, т.к. боятся случайно нанести вред крохотным организмам. Молодой человек хорошо знаком с порядками в общине джайнов. Но, возвращаясь в одиночестве

домой, он решает пренебречь обетами общины. Вместо того, чтобы следовать по окружной дороге, он решает пойти напрямик и, не глядя под ноги, пересекает луг, протаптывая себе тропинку. Он не чувствует мотивации соблюдать обеты джайнов, не испытывает эмпатии к насекомым и растениям, которые могут оказаться под его ногами. Поэтому идёт без всякой осторожности. В этот момент его окликает учитель. Он укоряет юношу за нарушение заповедей. Он считает, что юноша виноват в несоблюдении правил общины. Оправдан ли в данном случае укор?

Мне кажется, укор и осуждение вполне здесь уместны. Уместно говорить о моральной ответственности юноши. Она уместна, даже если молодой человек не испытывает мотивации и не чувствует эмпатии ко всем живым существам. Присоединяясь к общине джайнов, он принимает их установки в качестве минимальных моральных требований. К тому же он рационально вполне может понять природу их требований. И то, что молодого человека больше мотивирует желание сэкономить время, чем обеты джайнов, не снимает с него ответственности. Этот пример является доводом в пользу моральной ответственности психопатов. Психопаты среди нормальных людей — как молодой человек среди джайнов, а не как люди среди инопланетян Шумейкера. Но чтобы убедиться в наличии моральной ответственности психопатов, мне кажется необходимым привести еще один аргумент. Этот аргумент построен на аналогии между психопатами и закоренелыми преступниками.

#### *Е. Психопаты и закоренелые преступники*

Сторонник освобождения психопатов от моральной ответственности рассуждет так: психопаты постоянно совершают жестокие преступления, значит, они лишены способности обходиться с другими как с личностями и/или не обладают моральной компетенцией. Поэтому они не могут быть морально ответственными. Но этот аргумент верен и в отношении закоренелых преступников. Ведь они тоже постоянно совершают жестокие преступления. Значит ли это, что они тоже должны быть освобождены от

моральной ответственности, а также — от... уголовной? Если моральная компетенция снижается пропорционально частоте и тяжести аморальных поступков, самые закоренелые и жестокие преступники оказывались бы за рамками морального осуждения. Но тогда сама практика моральной критики потеряла бы всякий смысл.

Это сведение к абсурду первого аргумента. Очевидно, закоренелые преступники должны подвергаться осуждению и воздаянию. И должны быть уголовно наказуемы. Следовательно, и психопаты должны подвергаться такому осуждению. И дело здесь даже не столько в наличии способностей или их отсутствии, сколько в чертах характера и качестве воли преступников. И психопаты, и «психически нормальные» закоренелые преступники склонны к регулярному совершению преступлений. И те, и другие обладают скверным характером и являются аморальными личностями. И те, и другие обладают «злой волей», т.е. намеренно причиняют зло другим. И способность к эмпатии или её отсутствие здесь не так важны.

Что даёт наличие самой возможности эмпатии? Сравним аморальные действия двух человек. Предположим, они совершают одинаковые действия с одинаковыми мотивами — мошенничеством крадут деньги у пожилой женщины. Один человек, более жестокий, никогда не испытывает эмпатии, а другой — переживает её в редких случаях. Является ли особенностью характера первого основанием освобождения его ответственности? Мне кажется, что нет. Во-первых, у преступников — одинаковые мотивы. Эти мотивы основаны на личной выгоде. И действия совершены не бессознательно, а именно на основе этих мотивов. Во-вторых, оба преступника продемонстрировали «злую волю» — в поведении их проявилось суждение «мои интересы — превыше всего». Тогда почему только второй должен нести ответственность? Почему нереализовавшаяся альтернативная возможность должна повлиять на оценку актуализованного осознанного действия? На мой взгляд, если и есть разница в оценке поступков двух этих людей, она — не в пользу первого преступника. Первый

является жестокосердным по натуре. И это — лишь дополнительное основание для морального осуждения. Представляется, что основаниями ответственности служат не столько альтернативные возможности, сколько актуальные мотивы и ценности. Следовательно, и психопаты должны нести моральную ответственность. Возможно, даже в больше степени, чем преступники с нормальной психикой. Психопаты и закоренелые преступники различают добро и зло как минимум в том смысле, что могут определять, что интерпретируется обществом как зло, а что — как добро, они способны сопоставлять эти нормы со своими планируемыми действиями и вступают в обществе в межличностные связи, поэтому они являются личностями и несут моральную ответственность.

Впрочем, этот спор показывает нам, что границы класса личностей размыты. Можно оспорить соответствие психопатов двум необходимым требованиям, предъявляемых к личности, как это было рассмотрено выше. Но можно также оспорить соответствие каждого человека, совершающего антисоциальные поступки, этим и другим требованиям. И тогда, если признать этот ход корректным, все злоумышленники окажутся освобожденными от ответственности. Это явно абсурдная ситуация. Следовательно, к каждому критерию нужно применять оператор «отчасти»: личность — это человек «отчасти разумный», «отчасти сознательный», «отчасти способный относиться к другим как к личностям», «отчасти обладающий самосознанием» и «моральной компетенцией». Без этого оператора категория личностей останется пустой. Как и моральная компетенция, остальные критерии являются идеальными конструкциями, которым нельзя подобрать точные соответствия в мире. Рациональный, самосознательный, морально компетентный и даже просто сознательный — это абстрактные характеристики, ориентиры, а не реальные свойства людей. Агент в различное время оказывается дальше или ближе в отношении этих ориентиров, но никогда не достигает их в полном объеме, тем более во всех своих действиях. Таким образом, личности «в конечном счёте» не

существует и существовать не может, кроме как в метафизических фантазиях. Но это не лишает смысла рассуждения о личностях.

Пусть понятие личности — нормативная модель и представляет собой слишком высокую планку для биологических организмов. Но это не отменяет практическую пользу этого понятия. Если бы нормативные понятия вообще не имели смысла, не имели бы смысла и все этические построения. Зато теперь мы знаем, какого рода сущность следует искать. Это знание поможет нам в решении вопроса о реидентификации и атрибуции ответственности. Это поможет, в частности, принять некоторые контринтуитивные положения нарративного подхода — наиболее перспективной теории личности.

## V. Нарративный подход в теориях Д. Деннета и М. Шехтман

Нарративный подход отличается от подходов, которые мы исследовали раньше. Основное отличие — в том, что его сторонники меняют фокус, они корректируют постановку вопроса. Сторонники психологического, субстанционального и биологического подходов заняты вопросом *реидентификации* личности в разное время. Они пытаются определить, что делает одну личность в момент  $t_2$  той же самой личностью, что существовала в момент  $t_1$  (или, точнее, определяют, что делает одну временную стадию личности тождественной другой временной стадии личности). То есть они с целью обеспечения длительности существования личности устанавливают *отношения нумерической тождественности* между временными стадиями личности (person time-slices или person-stages). Их оппоненты, сторонники неклассического, нарративного подхода, в частности М. Шехтман, Д. Деннет, К. Эткинз<sup>224</sup>, считают это отношение нерелевантным для практических целей. Для сохранения длительного существования личности и обеспечения

---

<sup>224</sup> Другими сторонниками нарративного подхода являются Тейлор, Веллеман, Макинтайр.

условий моральной ответственности (как, впрочем, и для критериев выживания и справедливой компенсации), по их мнению, нужно решать не вопрос реидентификации, а вопрос *характеризации личности*<sup>225</sup>. Для того, чтобы определить, насколько личность несет ответственность, не нужно сравнивать стадии личности, а нужно выявить, какие психологические характеристики, ментальные события и физические действия могут быть действительно приписаны личности. А для этого следует исследовать моральную сущность личности, определить, «какова эта существующая личность», какие черты для нее являются ключевыми, а какие — случайными. Таким образом, сторонники нарративного подхода заняты отношением принадлежности личностям соответствующих свойств, состояний и действий, а не отношением нумерического тождества между временными стадиями личности.

Принадлежность психологических характеристик, ментальных событий и физических действий в прошлом личности в нарративном подходе определяется возможностью подходящего и непротиворечивого включения их в биографическую историю личности как принадлежащих самой личности. Под биографической историей подразумевается обычный последовательный рассказ о развитии личности, преимущественно от первого лица. Именно возможность непротиворечивого включения в последовательную реалистичную автобиографию делает эти события и характеристики частью психологической сущности и подлинной идентичности личности, а также является основанием её моральной ответственности за соответствующие поступки. Нарративный подход кажется наиболее близким обыденной практике вынесения моральных суждений. Он даже ближе, чем биологический подход. Допустим, вы пытаетесь разобраться, несет ли ответственность человек за какое-то

---

<sup>225</sup> Такая постановка вопроса особенно характерна для М. Шехтман. Schechtman M. *The Constitution of Selves*. Ithaca: Cornell University Press. 2007.

происшествие. Конечно, можно опираться на одни только улики, вещественные доказательства присутствия человека на месте происшествия. Но логичней всего задать ему и свидетелям событий вопрос об этом происшествии. В ответ, скорее всего, последует разъяснение в виде истории. Агент расскажет о том, что произошло, о чем он думал, на что надеялся, что получилось. По рассказу можно будет понять, какое отношение имеет этот человек к происшествию, какие мотивы им двигали, какую выгоду он приобрел и какие потери понес. Только по истории, в которой объединятся перспективы разных участников и свидетелей, можно будет определить вину или невиновность того или иного лица. Без истории нельзя определить, случилось ли убийство или это была самозащита, что было случайным стечением обстоятельств, а что — реализацией хорошо продуманного плана. Без истории нельзя ни разобраться, что произошло, ни справедливо оценить происшествие.

### *1. Конструкция характеризующего биографического нарратива*

Но как определить, что может, а что не может быть непротиворечиво включено в историю, в биографический контур нарратива? Что может объединять историю и служить фильтром для нерелевантных элементов? Во-первых, историю может объединять *перспектива*. Истории обычно рассказываются или от лица какого-то персонажа, или поочередно от лиц нескольких героев, или с позиции автора. Но в каждый момент эта перспектива определена. Перспектива — это взгляд с точки с четкими временными и географическими координатами. Перемещение этих точек в развивающейся истории в основном последовательно и непрерывно как в пространстве, так и во времени. Даже тогда, когда это перемещение скачкообразно (как в истории «Где Я?»), скачки имеют объяснения. Конкретное расположение точки перспективы обуславливает избирательное отношение к обстоятельствам и выражается в описаниях под каким-то специфическим углом зрения. Даже когда это перспектива с высоты «птичьего полета», это — конкретная перспектива. Она действует как



прожектор, выхватывая из бесконечного множества допустимых описаний основное. Исходя из некоторой перспективы, часть обстоятельств выделяется как значимая, важная, основная, а другая часть, собственно, большая часть, попросту игнорируется. Значимые события изображаются подробней, их временные рамки как бы раздвигаются. А незначительные обстоятельства проходят мимолетно или вообще не упоминаются. Отношение между различными планами по важности систематично. Выделяются главные и побочные персонажи, ключевые, образующие историю события и случайные обстоятельства. Ключевые события образуют канву сюжета или вехи биографии. Таким образом, законы перспективы позволяют отобрать и правильно расположить элементы рассказа.

Второе возможное объединяющее начало — это *траектория психической жизни персонажа*. История обычно состоит из объективных событий и действий героев. Действия, как мы уже знаем, увязываются с ментальными состояниями, а эти ментальные состояния, сочетаясь между собой, согласуются с сущностными характеристиками личностей. Ментальные состояния объясняют поступки, а характер и склонности личности — ментальные состояния. Из одних ментальных состояний проистекают другие. Таким образом, вся жизненная траектория должна быть в истории интеллигибельна. Это не означает, конечно, неизменности характера. Обстоятельства истории влияют на героя, они могут значительно изменить его состояния, характеристики и даже сущность. Но все эти изменения происходят не без причин. Причины включаются в историю и объясняют эти изменения. Таким образом, вся история становится понятной, последовательной, осмысленной, а траектория психической жизни персонажа не перестает быть непрерывной. В этой интеллигибельности — второе основание единства и фильтр для элементов нарратива.

Третий скрепляющий элемент биографической истории и одновременно её фильтр — *телеологическая направленность*. История

может объединяться движением личности к каким-то долгосрочным целям. Иерархия целей может увенчиваться каким-то основным ориентиром, главным маяком в путешествии личности. Именно этот главный ориентир, конечный смысл в масштабе всей жизни наделяет значением отдельные поступки и состояния. Кажется, главные ориентиры могут быть разные. Это может быть выживание, личное благополучие, доминирование, навязывание другим своей воли или получение максимального удовольствия. Но некоторые сторонники нарративного подхода (в первую очередь, герменевтического направления) настаивают, что подлинное «личное благо» должно совпадать с добром как всеобщей ценностью: желанием жить хорошо вместе с другими и для других в справедливом обществе. Только наличие этого главного этического ориентира, по их мнению, делает осмысленной жизнь в целом и каждое действие в отдельности.

Перечисленные выше характеристики позволяют изнутри скрепить нарратив, сложить жизненную историю в одно целое. Они являются *внутренними* требованиями к содержанию (логике и структуре) биографической истории. Но есть еще два необходимых *внешних* условия, предъявляемых к биографическим нарративам для атрибуции характеристик агентам. Это условия *возможности изложения* и *правдоподобия нарратива*. Биография может быть последовательна в отношении перспективы, представлять ясный психологический портрет героя, быть осмысленной, обладать телеологичностью, но при этом не давать достаточных оснований для атрибуции действий и моральной ответственности за них. Это — случаи, когда сам агент не может сформулировать и выразить собственный биографический нарратив или его часть, или когда нарратив целиком представляет собой фикцию. В первом случае агент может быть неспособен в достаточной мере анализировать и оценивать свое поведение, а во втором его оценки могут быть далеки от реальности.

Способность сформулировать *полную* жизненную историю в рассказе от первого лица, удовлетворяющем внутренним требованиям, не является абсолютной необходимостью для атрибуции действий и ответственности. В конце концов, агент может не понимать мотивов своего поведения, не всегда различать конечные цели, даже не иметь возможности последовательно и подробно изложить события всей своей жизни. Но важно, чтобы агент имел способность организовывать свой опыт и представления в жизненную историю. Свидетельством её наличия является способность агента выражать в истории хотя бы часть жизненных событий и объяснять их. Я имею в виду способность описывать свои поступки от первого лица, ссылаться на их ментальные причины и обосновывать свои суждения об опыте. Если агент способен к такому описанию фрагментов жизни, значит, он способен структурировать свой опыт в виде нарратива и, следовательно, то, что вписывается в этот нарратив и отвечает минимальным требованиям реализма, может служить его характеристикой.

Правдоподобие этой биографической истории — второй, но не менее важный внешний критерий. Требование правдоподобности, как и требования способности описать опыт, также не может быть абсолютным. Повествующий агент может допускать и даже всегда допускает какие-то ошибки. Но хотя бы часть суждений о событиях в нарративе должна совпадать с суждениями других личностей. Иными словами, описания событий от первого лица должны хотя бы частично совпадать с описаниями от третьего лица других участников и свидетелей событий. Иначе это будет фантастический нарратив, и его нельзя будет использовать в качестве основания атрибуции действий и ответственности. Некоторые причины тому были обсуждены еще во Введении. Выше в тексте этой главы также было показано, что людей, представления которых полностью оторваны от реальности, вообще невозможно относить к категории полноценных личностей. Даже если они способны понимать этические модели поведения

(в том числе способны представлять, как можно относиться к другим как к личностям), они неспособны применять эти модели на практике. А без этого они не будут удовлетворять требованию моральной компетенции. И фантастическая автобиография такого человека не будет удовлетворять требованиям нарративного подхода. Работу внутренних и внешних критериев можно продемонстрировать на примере следующей ситуации.

Предположим, кто-то ошибочно называет себя Наполеоном и утверждает, что был командующим в знаменитой битве у Ватерлоо. Участие в этой битве он обосновывает стремлением восстановить величие французской империи, а разгром считает личной ответственностью. Теперь его целью является мировое признание и получение должного императору уважения. Как соотносится такая история с требованиями к нарративу? История артикулирована. В ней есть некоторая телеологическая целостность — стремление к мировому признанию. Но остальным критериям она не удовлетворяет. При анализе этого нарратива возникнут несоответствия. Нарушена будет непрерывность точки перспективы. Например, заблуждающемуся человеку сложно будет объяснить, что он делал в период между 1821г. и 2015г. и как сочетается его биография и ментальная жизнь с фактами жизни Наполеона. Скорее всего, нарушится требование к интеллигентности. Скажем, этому человеку будет трудно объяснить, зачем в XX веке он поступал в начальную школу, если в конце XVIII в. уже заканчивал кадетский коллеж. Но основной будет проблема несоответствия реальности. Так, этому человеку будет трудно объяснить физические различия между Наполеоном и им самим. Таким образом, его история не будет соответствовать критериям интеллигентности, единства перспективы и правдоподобия. Следовательно, она не сможет быть характеристикой этого человека и основанием для атрибуции ему описываемых действий. Совсем другая характеристика будет более

подходящей для этого человека: что он — вдохновленный историей фантазер или психически больной.

## 2. *Преимущества нарративного подхода*

Нарративный подход предполагает, что личность формируется из рассказа и, соответственно, обладает характеристиками, расположенными в разных временных координатах. Личность реализуется в жизненной истории, автобиографии. В этом и есть сущность личности. Поэтому для личности особенно важно, чтобы нарратив развивался максимально благополучно. Личность имеет основания ожидать справедливой компенсации в рамках этого нарратива и несет ответственность за действия, включенные в нарратив. Особую роль в формировании этого нарратива играет перспектива первого лица. То, что рассказ из этой перспективы имеет важное значение, даёт дополнительные автономии, контроль и свободу личности<sup>226</sup>. При этом автономия личности не абсолютная. Автономия ограничивается внутренними и внешними критериями характеризующего автобиографического нарратива. Имеет ли этот подход преимущества в сравнении с другими подходами? И решает ли он проблему, обозначенную в Аргументе пяти трансплантаций?

Аргумент пяти трансплантаций показывает, что субстанциональный, биологический и психологический подходы не объясняют *особой* заботы личности о будущем. Забота о «своем» будущем, в соответствии с этими подходами, в целом не должна отличаться от заботы о будущем других личностей. Парфит называет это утверждение «Крайним утверждением» (*Extreme claim*)<sup>227</sup>. Крайнее утверждение следует из Аргумента пяти трансплантаций. Обратной стороной Крайнего утверждения является то, что я назову «Исключающим утверждением». Исключающее утверждение состоит в отрицании возможности атрибуции моральной ответственности

---

<sup>226</sup> Многие философы думали, что человеческая автономия включает или, «возможно, даже состоит из способности самосозидания — способности определять, изобретать и создавать себя». Velleman. *Self to self* // *The Philosophical Review*, vol. 105, No.1. P. 203.

<sup>227</sup> Parfit D. *Reasons and persons*. Oxford: Oxford University Press. 1984. P. 307

ввиду отсутствия асинхронного тождества личности, исчезновения агента во времени. Нарративный подход, определяя личность не как последовательность временных фаз, а как распределенный во времени рассказ, решает эту проблему и, следовательно, обходит как Крайнее, так и Исключающее утверждение. Этот подход опирается на идею фундаментального единства нарратива и его частей. И благодаря этому способен объяснить особенную заботу личности о своем будущем и ей ответственность за свое прошлое.

Субстанциональный, биологический и психологический подходы оказываются уязвимыми для гипотетических случаев с делением и умножением личности. Следствием применения этих подходов к подобным ситуациям является противоречие или эпистемологический пат. У сторонников этих подходов есть только две альтернативы. Либо им необходимо признать, что определить наследника существовавшей до деления или умножения личности невозможно. Либо что одна личность вдруг становится тождественной нескольким разным личностям. Нарративный подход обходит эту дилемму. Это оказывается возможным потому, что связь через нарратив фиксирует отношение принадлежности, а не отношение асинхронного тождества. Предположим, произошло деление или умножение личности. Согласно нарративному подходу, наследниками действий и моральной ответственности одной личности будут несколько человек. У этих людей будет одинаковое прошлое. Но в данном случае противоречия не возникнет. Как стена может быть частью и гостиной, и спальни одновременно, так и часть нарратива может входить в разные биографии. Следовательно, гипотетические случаи с делением и умножением личностей не являются проблематичными для нарративного подхода. И этот подход обладает преимуществами по отношению к другим подходам. Но насколько он уязвим для критики? Кажется, некоторые варианты возражений, уже опробованные выше, могут быть применены и к

нарративному подходу, в частности возражение, основанное на разнообразии модусов человеческой жизни.

### 3. Возражения

#### Возражение 1. Разнообразие модусов жизни

Нарративный подход предполагает, что сущностью личности является последовательная жизненная история. И что для реализации этой сущности необходима способность структурировать свой опыт, а также способность хотя бы отчасти выражать его: описывать поступки от первого лица и связывать их со своими ментальными состояниями. Но такие способности отсутствуют у людей в младенчестве, а иногда и в старости или в состояниях психических заболеваний (при слабоумии, в хроническом вегетативном состоянии и т.п.). Таким образом, кажется, что нарративный подход вынуждает нас вычеркивать эти этапы из жизни личности. А это противоречит интуициям. И порождает новую проблему — *«Проблему кого-то еще»*. Если человек в хроническом вегетативном состоянии (ХВС) — не та же личность, что был до этого состояния, то кто этот человек в ХВС? Кто этот «кто-то еще»?

Действительно, исключение подобных этапов из истории личности может быть следствием применения нарративного подхода к характеристике личности. Но это не обязательное следствие. Так, эти этапы личности можно включить в историю личности, сформулированную не самой личностью, а другими. Когда речь шла о требовании реализма нарратива, мы уже говорили о необходимости согласования нарратива от первого лица с нарративами, выстроенными из других перспектив. Таким образом, уже было очевидно, что другие личности имеют некоторое отношение к биографическому нарративу. Для того, чтобы особые этапы жизни личности можно было включить в ее историю, стоит добавить возможность того, чтобы даже в условиях ее полного «молчания», другие могли дополнять, расширять и уточнять ее автобиографический нарратив. Т.е. автобиографический

нарратив — это история, написанная не только от первого лица, но и дополненная другими участниками, свидетелями событий из жизни личности. Если дети и слабоумные не могут «выразить свой нарратив», он может быть выражен в сотрудничестве с другими. Такой вариант нарративного подхода обходит Проблему кого-то еще.

При этом стоит сделать оговорку: участие других в создании нарратива приемлемо только тогда, когда сама личность либо имеет перспективу в дальнейшем выразить предыдущий невыраженный ею опыт, либо когда-то имела такую возможность. Эта оговорка дает основания исключить совсем неподходящие случаи, вроде личностей плюшевых мишек или деревьев. Следствием такой оговорки будет проведение принципиальных различий между здоровыми младенцами и младенцами с врожденными тяжелыми когнитивными дефектами, а также взрослыми, находящимися в состоянии ХВС после травмы и от рождения. По моему представлению, последние (младенцы и взрослые с врожденными недостатками) не имеют и не могут иметь автобиографического нарратива, и посторонние личности не могут им помочь в его создании.

## Возражение 2. «Эпизодические личности»

Нарративный подход можно критиковать также с помощью контрпримеров. Такую тактику предлагает американский философ Г. Стросон. По его мнению, существует тип людей, которые, кажется, не подойдут под критерии личностей, представленные сторонниками нарративного подхода. Таких людей Г. Стросон называет *«эпизодическими личностями»*. «Эпизодические личности» — это те, кто не считает себя кем-то, кто существовал в прошлом и будет существовать в будущем. При этом «эпизодические личности», естественно, понимают, что биологическое существо, в котором реализуется их «я», обладает временной длительностью, но само «я» они не считают длящимся<sup>228</sup>. «Эпизодические личности» не

---

<sup>228</sup> Strawson G. *Against Narrativity* // Strawson G. (ed.). *The Self?* Oxford, 2005. P. 65.



видят свою жизнь как разворачивающийся нарратив. В этом смысле они живут только в настоящем, полностью в настоящем. Их Стросон противопоставляет «*диахроническим* личностям». Диахронические личности осознают себя как того, кто существовал в прошлом и будет существовать в будущем. Возможно, диахронических личностей большинство. Но, по мнению Стросона, «эпизодические» тоже существуют. К «эпизодическим личностям» он относит и себя. Как бы то ни было, помыслить существование «эпизодических личностей» можно. Значит, в наших интуициях обнаруживается противоречие. Из этого, по мнению Стросона, следует, что, согласно интуициям, нарратив не является неотъемлемым и сущностным элементом личности. А значит, нарративный подход ложен.

Стросон предвидит соображения, которые мог бы выдвинуть сторонник нарративного подхода в свою защиту — аргумент, доказывающий, нарративную структуру даже «эпизодических личностей». Аргумент состоит в следующем. Жизнь «эпизодических личностей» — это *развитие*, как и жизнь диахронических личностей. В этом смысле жизнь «эпизодических личностей» тоже представляет последовательность. Эту последовательность могут описывать как сами эти люди, так и их окружение. Следовательно, «эпизодические личности» также реализуются в нарративах. Но Стросон считает, что этот аргумент неудачен, так как он сводит нарративный подход к тривиальности. Если жизнь «эпизодической личности», можно, тем не менее, считать реализующейся в нарративе, то и жизнь любого человека будет в этом смысле нарративной. Более того, даже жизнь лошади или собаки можно таким образом считать реализующейся в нарративе. В конце концов, и в жизни лошади или собаки будут развитие и последовательность. Таким образом, Стросон предлагает своеобразную стратегию: либо нарративный подход тривиален, либо он опровергается контрпримером. Так ли это?

Мне кажется, что критика Стросона может быть эффективна для некоторых нарративных концепций личности. Но она бьет мимо тезиса, который я пытаюсь защищать в этой работе. Для удобства свой тезис я назову «*Прагматическим тезисом*» о нарративной природе личности. Суть его в том, что именно моральная ответственность за действия личности в прошлом (а также другие прагматические вопросы, такие как забота о выживании или справедливой компенсации) базируются на нарративной структуре личности. Прагматический тезис следует отличать от «*Психологического тезиса*» о нарративной природе личности<sup>229</sup>. Согласно Психологическому тезису, все люди переживают свою жизнь как историю, нарратив. Именно Психологический тезис уязвим для критики с помощью контрпримеров с «эпизодическими личностями», а не Прагматический. Прагматический тезис предполагает лишь то, что основанием применения моральной ответственности будет нарративная структура и единство личности. Но человек может существовать и без такой структуры. В таком случае, однако, он просто не будет личностью в том понимании этого слова, как оно используется в данной работе, и применение моральной ответственности к нему будет неуместным, неоправданным, безосновательным. Моральная ответственность, как плод на ветке нарратива: падает ветка, падает и плод. Значит ли это, что «эпизодические личности» не являются, собственно, личностями и не несут ответственности? В частности, означает ли это, что к Стросону нельзя применять эту категорию?

В строгом смысле — да. Действительно, «эпизодические личности» не должны нести ответственности за действия. Но является ли личность «эпизодической» или диахронической — вопрос не такой однозначный. В решении этого вопроса не только сама личность, но и представления окружающих об этой личности имеют значение. Человек может

---

<sup>229</sup> Г. Стросон также выделяет «*Этический тезис*» о нарративной природе личности. Согласно этому тезису, переживание и осознание жизни как нарратива является необходимым для благополучной жизни. Strawson G. *Against Narrativity* // Strawson G. (ed.). *The self?* Oxford, 2005. P. 63.

декларировать отсутствие особой привязанности к своему прошлому и особой заботы о своем будущем, как это делает Г. Стросон. «Я имею прошлое, как любой другой человек, и я прекрасно понимаю, что имею прошлое. У меня достаточно фактических знаний о нем и я также помню некоторые из моих переживаний в прошлом "изнутри", как говорят философы. И при этом я ни в коей мере не переживаю свою жизнь в нарративной форме или как нарратив без структуры. Совсем не переживаю. У меня нет особого интереса к своему прошлому и у меня нет особой заботы о своем будущем»<sup>230</sup>. И при этом из его поведения может следовать обратное, по крайней мере, иногда. В таком случае этого человека нельзя считать «эпизодической личностью». К примеру, если человек все-таки иногда сожалеет или гордится событиями своего прошлого, если он иногда с большей тревогой относится к переживаниям в будущем собственной боли, чем будущим переживаниям боли других, значит, он — не «эпизодическая», а диахроническая личность. И он несет моральную ответственность за действия, которые следует относить к его биографической истории, даже если сам он их явным образом к себе не относит. Каким человеком является Гален Стросон? Мне сложно сказать, но мне сложно представить, чтобы человек, живущий вполне обыкновенной жизнью, был «эпизодической личностью».

С другой стороны, я, конечно же, могу представить себе человека, проживающего экстраординарную жизнь, который был бы «эпизодической личностью». Таким человеком мог бы быть отшельник, проводящий время в медитации и реализовавший буддийские идеалы: человек, лишенный страдания и освободившийся от желаний. Нет ничего противоречивого в том, чтобы помыслить человека с такими психологическими характеристиками. Но его поведение, по моему мнению, радикально отличалось бы от поведения обычных людей, и это было бы очевидно со стороны. Такого

---

<sup>230</sup> Strawson, G. *Against Narrativity* // Strawson G. (ed.). *The self?* Oxford, 2005. P. 67.

человека вряд ли можно было бы считать обычной личностью или вообще личностью. Более того, думаю, ему вообще трудно было бы приписывать наличие «Я».

Это отчасти подтверждают описания переживаний существования «эпизодических личностей» самого Г. Стросона. «Я не имею осмысленного ощущения, что "Я" — "Я", сейчас размышляющий над этим вопросом, — существовал когда-то в прошлом. И для меня это не ошибка восприятия. А скорее регистрация факта обо мне — о том, кто сейчас размышляет над этой проблемой»<sup>231</sup>. Стросон ссылается на то, что у него нет ощущения, что «Я» присутствовало в прошлом. Но он также не разъясняет, в чем заключается осмысленное ощущение существования «Я» в настоящем. Если «переживания перспективы изнутри» не являются достаточным для переживания «Я», то ни в прошлом, ни в настоящем с помощью интроспекции нельзя обнаружить «Я». Достаточно вспомнить приведенные в начале главы рассуждения Юма на этот счёт. Таким образом, в ответ на стратегему Стросона, я предлагаю другую: либо большинство людей, считающих себя «эпизодическими личностями», фактически являются диахроническими личностями, т.е. просто личностями — в том смысле, как этот термин понимается в данной работе, либо такие люди действительно «эпизодические личности», но тогда они не только не личности в собственном смысле слова, но и не обладают «Я». Эти специальные случаи «эпизодических личностей» возможны, но они не являются угрозой для прагматического тезиса о нарративной структуре личности. Таким образом, я считаю критику Стросона не опасной.

Не будучи опасной, она, тем не менее, бесполезна. Замечания Стросона подтверждают, что свойство «быть личностью» не является фундаментальным и простым свойством объектов. Оно основано на паттерне событий. Как и ментальная каузальность, это — сложное высокоуровневое

---

<sup>231</sup> Strawson, G. *Against Narrativity* // Strawson G. (ed.). *The self?* Oxford, 2005 p. 68.

свойство, различимое только из перспективы моральной ответственности и рождающее иллюзию метафизической простоты, иллюзию точки сборки и источника действий.

Анализ преимуществ нарративного подхода и замечаний в его адрес позволяет сделать вывод о том, что это — наиболее удачное решение проблемы связи событий и атрибуции их личности. Нарративный подход позволяет объединить атрибуты личности как единомоментно, так и в темпоральной перспективе. Атрибуция событий, в свою очередь, обеспечивает основания приписывания моральной ответственности за действия, произведенные личностью в прошлом. Однако остается не до конца проясненным соотношение нарратива и личности. Ведь именно личность является субъектом ответственности. Нельзя же приписывать ответственность тексту!

Итак, нарратив позволяет охарактеризовать личность. Но кажется, что личность — не то же самое, что биографический нарратив. Представления о нарративе и интуиции о личности, об эго, о «Я» значительно расходятся. Выше, в начале главы, я уже перечислял традиционные характеристики личности. Среди них — представление, что личность — это простая, элементарная, неделимая сущность. Что она является единым командным пунктом, где принимаются волевые решения. Что эго — «главное место», и что личность сохраняет нумерическое тождество во времени и является абсолютно уникальной, что её невозможно копировать или размножить. Если нарратив и связан с личностью, то, согласно этим первичным интуициям, он каузально зависим от «эго», «Я». «Я» является автором нарратива, разве не так?

## VI. Центр нарративной гравитации

## 1. Полезные аналогии, проясняющие, что такое личность

С тем, что субъект, или «эго», является подходящим объектом для атрибуции ответственности, сложно не согласиться. Именно «эго» мыслится рациональным источником воли и «точкой сборки» всех ментальных содержаний. Это — элементарная, неделимая сущность, сохраняющая нумерическую тождественность во времени, уникальная и защищенная от копирования. Но это «эго», судя по всему, не является реальным объектом. Это не вещь, не что-то конкретное. Скорее, это теоретическая конструкция, абстрактная сущность, и в некотором смысле — вымысел. Такую позицию занимают Д. Деннет, Д. Вегнер, М. Газзанига.

Как всякая фикция, «эго» не может быть каузально эффективным. Не абстрактное «эго» является причиной появления автобиографического текста, а автобиографический текст является источником убеждения в существовании «эго» с описанными выше характеристиками. В данном случае, «Я» — это не автор, а персонаж текста. Но не текст создается персонажем, а персонаж — текстом. Представление об этом персонаже и его характеристиках рождается в результате интерпретации поведения организма и произведенного мозгом текста. Интерпретацию осуществляют как другие, так и сам агент. При этом представление об «эго» — не случайный вымысел. Это очень полезная конструкция, так как она позволяет делать успешные предсказания, влиять на поведение организма и взаимодействовать с ним. Ее онтологический статус не многим хуже, чем статус научных понятий, таких как крутящий момент, электрическая мощность или, например, центр тяжести в физике. В качестве иллюстрации такого представления Деннет предлагает несколько аналогий<sup>232</sup>.

*Аналогия 1. Центр тяжести.* Центр тяжести играет важную роль в определении поведения объекта. Возьмем, к примеру, неваляшку, некогда

---

<sup>232</sup> Dennett, D. The Self as a Center of Narrative Gravity // Kessel F., Cole P. and Johnson D. (eds.) Self and Consciousness: Multiple Perspectives. Hillsdale, NJ: Erlbaum, 1992. P. 275-288.

одну из самых распространенных детских игрушек. Как ни наклонять эту фигурку в разные стороны, она в итоге примет исходное вертикальное положение. Это происходит потому, что ее центр тяжести находится в нижней части игрушки и стремится занять самое нижнее положение. Наклоняя фигурку, мы смещаем эту точку вверх. Но стоит только ее отпустить, как центр тяжести вновь займет самое близкое к опоре положение.

Зная законы физики, можно найти точку, в которой находится центр тяжести любого объекта. Теоретически в некоторых случаях можно указать даже конкретный атом, совпадающий с центром тяжести. Но пространственное совпадение не будет означать тождественности. Недопустимо идентифицировать центр тяжести объекта с конкретным атомом. Это было бы категориальной ошибкой. Атом обладает физическими свойствами: массой, размером, электрическим зарядом, а центр тяжести — нет. Это просто геометрическая точка в пространстве. И вести себя она может иначе, чем частицы материи. Так, если прикрепить к одной доске другую другой формы, центр тяжести объединенного предмета окажется в другом месте. Но перемещение центра тяжести будет не последовательное, как у всех материальных объектов, а дискретное, скачкообразное. Некоторые физические ограничения не действуют на абстрактные теоретические сущности.

Физик, изучая движение вещей, допускает существование центра тяжести. А антрополог-любитель, каковыми мы все являемся, постулирует существование «эго» как в себе, так и в других людях. Оба случая принципиально схожи, так как пополняют онтологию теоретическими сущностями, полезными в реализации практических целей. И эти теоретические сущности имеют любопытные особенности. Эти особенности можно исследовать на примере другой фикции — вымышленных персонажей.

*Аналогия 2. Персонаж литературного произведения.* Откроем роман «Моби Дик». На первой странице читаем: «Зовите меня Измаил». Кого называть Измаилом? Автора этих строк, Мелвилла? Нет, мы зовем Измаилом Измаила. Но кто он такой? Это вымышленный персонаж Мелвилла. Довольно много о нём мы узнаем из романа. Что он китобой и что он ищет приключения на судне «Пекод». Кое-что оказывается понятным между строк. Например, что у него две руки и две ноги. Ведь кроме особенностей, описанных автором, мы полагаем, что его герой в остальном был похож на прочих людей (в отличие от капитана Ахава, например). Но что-то об Измаиле останется принципиально неопределённым. К примеру, было ли у него родимое пятно на плече? Этот вопрос не может быть выяснен, т.к. Мелвилл об этом ничего не говорит. Существование такого рода неопределённостей и есть основное отличие вымышленных персонажей от исторических. В отношении первых не всегда действует принцип бивалентности. Утверждения о них могут носить не только истинный или ложный характер, но и **неопределённый**. Ведь их существование определено только границами текста и культуры, в которых они создаются. То же самое применимо к личностям. Как, впрочем, и к центрам тяжести. Они имеют только те свойства, каким наделила их теория. Что, однако, не делает эти понятия бессмысленными или неприемлемыми для серьёзной дискуссии и даже науки. В конце концов, наука не смогла бы продвинуться без подобных понятий.

История с Измаилом — хорошая иллюстрация появления вымышленных личностей. Но она, кроме пользы, может отбросить нас назад, вернув к исходным интуициям. Она может напомнить о связи вымышленных героев с реальными авторами. Для нейтрализации этой интуиции стоит использовать ещё одну ситуацию, придуманную уже упомянутым Д. Деннетом специально для этой цели — ситуацию с машиной, которая пишет романы. Это будет третья аналогия, которую мы проведем для прояснения сущности личности.



*Аналогия 3. Созданный алгоритмом образ.* Представьте машину-писателя. В ней есть набор различных текстовых фрагментов, правил и каких-то алгоритмов перемешивания слов. Эта машина создана не для написания какого-то конкретного романа, а просто чтобы генерировать осмысленные тексты. Создана так, что создатели сами не знают, какой сюжет и каких героев она в этот раз выпишет. Вот мы включаем машину, и из ее принтера появляется заглавие, а за ним — первые строки романа: «Описание жизни Гилберта». «Зовите меня Гилберт». За этим появляются страницы биографического описания жизни и приключений Гилберта. Машина, не имеющая представлений ни о мире, ни о романе, вполне способна создать вымышленного и героя, и его биографию. Теперь усложним ситуацию: поместим машину в небольшого робота на колесах. Представьте, как он вместе с машиной-писателем, попадает в чулан и там случайно закрывается. И тут же из его принтера появляется продолжение истории: «Меня заперли в чулане. Помогите!» Помочь кому? Гилберту, наверное. Но кто этот Гилберт? Робот? Или вымышленный персонаж, созданный роботом? Мы спасаем машину из чулана, и она тут же выдает новый текст: «Благодарю, Ваш Гилберт». Теперь уже нельзя игнорировать любопытное соответствие между историями вымышленного персонажа Гилберта и робота. Но сам робот — не персонаж, он может даже не знать, что создает персонажа. Он фактически не обладает качествами, приписываемыми персонажу, в частности той целостностью, рациональностью и сентиментальностью Гилберта. Робот — только машина с различными механическими частями. И тем не менее, эта машина способна породить образ рационального, чувствительного, целостного «эго». Скорее всего, машина человеческого мозга делает что-то подобное.

Доводы в пользу такого суждения и релевантности приведенных выше трех аналогий можно почерпнуть, в частности, из психиатрии и нейронауки. В качестве основы для одного из доводов может служить *феномен*

*множественной личности* (иначе это психическое нарушение называется диссоциативным расстройством личности). Другой довод можно построить на наблюдениях за *пациентами с расщепленными полушариями мозга*.

## 2. Эмпирические доводы в пользу концепции Центра нарративной гравитации

### А. Диссоциативное расстройство личности.

Синдром множественной личности (СМЛ) — это психическое расстройство, при котором один человек, один организм обладает двумя или более идентичностями или личностными состояниями. Каждая «личность» (или альтер-эго) имеет собственный характер, мировоззрение и воспоминания (а иногда и возраст, расу, пол), отличающиеся от таковых характеристик других «личностей». Эти «личности» попеременно контролируют поведение человека. Поступки, совершенные под контролем одной «личности», не оставляют в памяти другой «личности» следов. Поэтому как минимум одна «личность» не знает о существовании другой. Свидетельства о существовании подобного синдрома существуют с XVII века, но именно на конец XX в. приходится основная масса исследований этого заболевания. К 1995 г. было зарегистрировано почти 40 000 подобных пациентов. Большая часть из них диагностирована в Северной Америке.

Обычно у расщепленной личности есть «хозяин», доминирующая по времени личность. Эта личность имеет довольно слабую эмоциональность. В отличие от нее, альтер-эго демонстрируют доминирование какой-то одной, но сильной эмоции или качества: раздражительность, заботливость, сексуальность, инфантильность. Жизнь альтер-эго состоит из тех эпизодов, когда оно выходит на сцену. Эти эпизоды оно запоминает и из них составляет историю о том, что оно такое. Чаще всего альтер-эго имеют собственные имена. Количество таких личностей может быть от двух до нескольких десятков. Замечено, что когда «семья» становится большой, начинают появляться личности противоположного пола. Своеобразие феномена можно почувствовать, только если рассмотреть какой-нибудь

конкретный пример. Возьмем историю болезни, воспроизведенную Деннетом в статье «Говоря за нас». Это история о Мэри. История эта вымышленная, но она не так далека от реально зафиксированных историй пациентов с подобными симптомами.

Тридцатилетняя Мэри страдает депрессией, резкой переменой настроений и потерей памяти. Последние несколько лет она часто попадает в клиники. Ей ставят разные диагнозы: шизофрения, маниакально-депрессивный психоз, пограничное расстройство личности. Ее случай, наконец, привлекает внимание доктора Р, занимающегося диссоциативными расстройствами. Мэри рассказывает врачу свою историю.

Ее отец умер, когда ей было два года. Поэтому Мэри воспитывал отчим. В целом он был добр к Мэри, но иногда «это заходило слишком далеко». В подростковом возрасте для нее характерны были резкие изменения настроений, плохая дисциплина в школе. Ее воспоминания о детстве, о школе и о занятиях очень отрывочны. К примеру, она хорошо играет на гитаре, но не помнит, как этому научилась. Рассказывая о себе, она иногда начинает говорить в третьем лице. Сейчас она также обнаруживает провалы в памяти: находит в гардеробе одежду, которую не покупала, отправляет родственникам несколько повторных поздравительных открыток. Она утверждает, что у нее высокие моральные ценности, но признает, что у других о ней может складываться противоположное мнение.

На одном из сеансов врач решает идти ва-банк. С согласия Мэри он вводит ее в гипнотическое состояние и просит появиться у той части Мэри, что не выходила раньше. И вдруг Мэри, или та, что недавно была Мэри, неожиданно бросает доктору кокетливый взгляд: «Привет, док! Я Салли... Эта Мэри — просто тряпка. Она думает, будто все знает, но я могу рассказать тебе гораздо больше». Внезапно появившись, Салли также внезапно исчезает. На следующих сеансах на сцене появляются новые лица — «Хейти» и «Пегги». Мэри об их существовании, конечно, ничего не знает. Ее истории диссоциированы. Чтобы вылечить пациентку, врачу нужно

познакомить различные «личности» Мэри друг с другом и интегрировать их в одну, а также выяснить причину возникшего расстройства. Только в таком случае Мэри сможет вернуться к нормальной жизни. По крупице врачу удается восстановить историю заболевания.

С четырех лет отчим стал брать Мэри с собой в кровать. Он ласково называл ее Сандрой, гладил и ласкал ее. Он говорил, что это будет их секретом. Иногда Мэри старалась доставить радость отчиму, иногда притворялась бесчувственной. Когда же, наконец, Мэри сказала ему, что больше не может терпеть и что расскажет обо всем другим — он ударил ее и предупредил, что тогда они оба отправятся в тюрьму. Переживания оказались невыносимыми. Поэтому Мэри просто перестала ассоциировать себя с Сандрой. Через какое-то время и Сандра «расщепилась», уступив место Салли — той части, что старалась доставить отчиму удовольствие, Хейти — той, что испытывала злобу и гнев, и Пегги — той, что притворялась бесчувственной. Во взрослой жизни Мэри эти личности нашли себе подходящие роли. Когда Мэри испытывала злобу, ее место занимала Хейти, когда ее целовал мужчина, на сцене возникала Салли. В юности все эти перемены объяснялись окружением Мэри как особенности «трудного возраста». К тридцати они уже доставляли ей серьезные проблемы.

История Мэри вполне типична для пациентов с СМЛ: детская травма и внутренний конфликт, затем подсознательная выработка внутреннего решения и изоляция травмирующего опыта, структурные изменения личности и терапия, вследствие которой травмирующий опыт восстанавливается и нейтрализуется. Эта история демонстрирует работу одного из защитных механизмов психики — диссоциации. Суть этого механизма заключается в том, что человек перестает ассоциировать действующего персонажа собственных воспоминаний с самим собой и отдаёт эту роль «кому-то другому», перенося на него и связанные с этими воспоминаниями переживания. Такая способность может реализовываться и у здоровых людей в измененных состояниях сознания или в критические,

особенно стрессовые моменты. Она представляет собой любопытный феномен. Но история Мэри в частности и феномен СМЛ в целом для нашего исследования полезны не столько как иллюстрация диссоциации, сколько как модель обратного, более базового, стандартного процесса — *процесса ассоциации или индивидуации личности*, выработки собственного «Я». Люди, страдающие СМЛ, ассоциируют ментальные содержания во множество кластеров, а здоровые — преимущественно в один. Если нарратив последователен, то в процессе индивидуации появляется один агент. А если нарратив не может быть выстроен последовательно ввиду его травматичности — то много. Но суть процесса от этого не меняется. Как пишет гарвардский психолог Д. Вегнер, субъективное эго, которое есть у нас в таком контексте — это только один виртуальный агент из множества возможных<sup>233</sup>. И судя по описанным случаям, процесс индивидуации личности, появления этого виртуального агента, больше похож на выработку писателем образа персонажа или постулирование теоретической сущности, чем на обнаружение нового природного объекта.

Для взаимодействия с внешним миром и предсказания даже собственных действий, удобно постулирование единой, рациональной, уникальной, непрерывно существующей личности. Но если сделать это не получается, один организм, один мозг может создать два образа или даже целую когорту. Личность способна интегрироваться или дезинтегрироваться как центр тяжести: стоит разделить физический объект, и у его частей тут же возникнут два собственных центра. Но если мы объединим части вновь, центр тяжести вновь будет один. Какой центр тяжести был новым, какой — старым? Существовали ли эти центры одновременно в одном объекте? Будет ли вновь объединенный центр тяжести тем же самым, что был раньше? Такие вопросы можно задавать. Но если ответы на них и будут различаться, это будут, скорее всего, лишь терминологические различия. То же самое верно и в отношении вопроса о количестве и тождестве личностей Мэри. Можно

---

<sup>233</sup> Wegner D. The Illusion of Conscious Will. Oxford, 2002. P. 255.

сказать, что в действительности у нее одна личность, но она многогранна, а можно сказать, что у нее 4 личности. Оба ответа будут в какой-то степени верны и будут различаться лишь терминологически. Если личность, как следует из наших размышлений, это абстрактная сущность, данные вопросы не имеют принципиального значения.

#### В. Пациенты с разделенными полушариями

Представление о том, что личность как единый рациональный субъект представляет собой удобную фикцию, поддерживается также наблюдениями за людьми с расщепленными полушариями мозга (split-brain patients). Эти люди в лабораторных условиях демонстрируют удивительный феномен — отдельную, автономную интеллектуальную работу полушарий мозга. Исходя из этих наблюдений, можно допустить у них два отдельных потока сознания с различными воспоминаниями, желаниями, убеждениями. При этом в обычных условиях такие пациенты практически неотличимы от здоровых людей. Ни люди со стороны, ни даже сам субъект не обнаруживает аномалию. Кажется, что даже в патологических случаях психика находит способы ассоциации переживаний и ретроспективного объяснения поведения, как будто оно происходит из единого волевого центра. Организм домысливает существование интегрированного агента даже тогда, когда он, очевидно, отсутствует.

Первооткрывателями феномена автономной работы полушарий были нейропсихологи М. Газзанига, Р. Сперри и Дж. Боген. С начала 60-х они наблюдали поведение пациентов, которые прошли операцию по рассечению мозолистого тела, сплетения нервных волокон, образующих прямую связь между правым и левым полушариями. Эта операция называется каллозотомией. Каллозотомия использовалась в качестве средства борьбы с тяжелыми формами эпилепсии. Она позволяла локализовать эпилептические симптомы в одной части мозга. Таким образом, в результате нее припадки эпилепсии обычно пропадали, а пациент после нескольких недель восстанавливался. Но восстановление, как выяснили Газзанига, Сперри и

Боген, происходило только «с виду». Пациенты никогда не возвращались к исходному психологическому состоянию. Эксперименты Газзаниги и его коллег показали, что после операции полушария мозга начинали работать как два изолированных, автономных рациональных модуля. Оба полушария оказываются способны отдельно воспринимать и анализировать информацию, самостоятельно инициировать поведение. Между ними возможны разногласия и конфликты. Как это ни удивительно, сам пациент при этом как будто ничего не замечает. Не замечали раньше эти внутренние конфликты и исследователи<sup>234</sup>. Как такая серьезная дезинтеграция внутренних процессов могла остаться незамеченной?

Избежать очевидных нестыковок и конфликтов помогала ориентация на косвенные внешние сигналы. Разделенные полушарии мозга, не имея прямого канала сообщения, использовали косвенные сигналы, чтобы «общаться» друг с другом. Они ориентировались на внешние проявления процессов в соседнем полушарии, такие как движения противоположной части тела или речь. Полушарии вели себя как два различных, но близких человека, договорившихся о полном согласии во всем: когда один что-то говорил, другой внимательно слушал, когда один смеялся, другой делал вид, что тоже обнаруживал что-то смешное. При этом субъект вовсе не обращал на имевшуюся дезинтеграцию внимания и вел себя так, как будто после операции в его внутренней жизни ничего не изменилось. Обнаружить патологию удалось только с помощью хитроумных экспериментов, которые и провели нейропсихологи. Одни из экспериментов был построен следующим образом.

Пациент видит две картинки, по одной каждым глазом: к примеру, лапу цыпленка слева и заснеженный пейзаж справа. Каждое полушарие получает эксклюзивно один образ от одного глаза (с противоположной стороны).

---

<sup>234</sup> Верификацию полного восстановления функциональности мозга провел на 26 пациентах с рассеченным мозолистым телом Э. Акелайтис. Ему не удалось найти никаких особенностей в постоперационном поведении этих пациентов. Akelaitis A. Studies on the corpus callosum: IV Diagnostic syprazia in epileptics following partial and compete section of the corpus callosum// American Journal of Psychiatry.1945. No 101.

После этого пациенту предлагают выбрать из набора картинок те, которые ассоциативно связаны с начальным образом. Пациент правой рукой выбирает цыпленка, а левой — лопату. (Лопата в данном случае — правильная ассоциация со снежным пейзажем.) Руководитель эксперимента просит пациента обосновать свой выбор. Пациент отвечает: «Все очень просто. Цыпленок ассоциируется с лапой цыпленка. А лопата нужна, чтобы чистить курятник». Интересно, что левое полушарие, обнаруживая поведение левой руки (управляемой правым полушарием), пытается интерпретировать его, используя ту информацию, которая в ходе эксперимента была доступна именно ему, левому полушарию, не используя информацию, которая была доступна правому, и даже не предполагая, что реакция левой руки могла быть обусловлена чем-то другим. Учитывая, что пациент осведомлен о прошедшей операции и имеет достаточное представление о ее последствиях, гораздо логичней был бы другой ответ: «Я не знаю причины выбора лопаты. Возможно, какая-то другая информация была передана моему второму, изолированному полушарию». Удивительно, но никто из пациентов не приходит к такому выводу. Они даже не обнаруживают аномалию. Этот феномен стабильно проявляется и в других экспериментах с пациентами с разделенными полушариями.

В одном из них какую-нибудь команду, например, команду «улыбайся», предъявляют через левый глаз правому полушарию пациента. Пациент улыбается. Руководитель эксперимента просит объяснить причину улыбки. Пациент объясняет: «Вы, ребята, приходите и тестируете нас каждый месяц. Вот классная работка!» Если показать правому полушарию пациента команду «встань», пациент встанет и пойдет. Когда его спросят, почему, он объяснит: «Хочу пойти в соседнюю комнату, взять колы». Подобные эксперименты всегда приводят к аналогичным результатам. Что это означает? Как это соотносится с приведенными выше рассуждениями о личности?



Пациенты с разделенными полушариями демонстрируют удивительную черту — способность присваивать и домысливать ментальные содержания. Они делают это абсолютно искреннее, незаметно для себя. Они интерпретируют свое поведение как поведение единого рационального агента, при этом таковым, очевидно, не являясь. Они присваивают источник и причину поведения, которая абсолютно не прозрачна для них. Значит, у пациентов встроены соответствующие психические механизмы. Но если такие механизмы есть у пациентов с разделенными полушариями, эти механизмы должны присутствовать и у нормальных, здоровых людей. Тогда, вполне возможно, что и здоровые люди точно так же ассоциируют ментальные содержания и домысливают существование единого длящегося рационального агента. Эти механизмы интерпретируют работу разделенных, автономных когнитивных модулей здорового человека как работу одного целостного рационального субъекта и создают удобную, но фиктивную сущность — личность, главного героя нарратива. Таким образом люди решают практическую задачу идентификации для прогнозирования собственного и чужого поведения и кооперации с другими людьми.

Исследования пациентов с разделенными полушариями мозга и пациентов с диссоциативным расстройством личности позволяют прояснить механизмы интеграции личности, которые имеются и у здоровых людей. Это механизмы предполагают интерпретацию поведения, домысливание ментальных содержаний и ассоциацию их в кластеры, а также формирование из этих кластеров представления об идентичности личности. Ассоциация ментальных содержаний и формирование личности имеет сходства с постулированием теоретических ненаблюдаемых научных сущностей и созданием персонажей художественных произведений. Но отличается от них в том, что изобретателем личности служит не рациональный агент, а система из связанных между собой когнитивных модулей мозга. Представить такое полифоническое решение поможет четвертая аналогия.

*Аналогия 4. Голос Америки.* Допустим, существует радиостанция, которая вещает новости на разных языках. Множество журналистов собирают информацию по всему миру. Сотни внештатных и штатных корреспондентов на местах интервьюируют экспертов, аналитиков, очевидцев событий. Собранную информацию они отправляют в офисы компании. Там десятки редакторов обрабатывают полученные сведения. Они сверяют полученные данные так, чтобы информация из одних источников подтверждалась информацией из других, уточняют детали, компонуют информацию по категориям, готовят тексты. Десятки радиоведущих озвучивают новостные сообщения, подавая их с соответствующим тоном и эмоциональной окраской. Звукорежиссеры и другие технические специалисты записывают аудиосообщения и транслируют их слушателям. Менеджеры обеспечивают работу всей инфраструктуры и задают основные направления политики радиостанции. Весь этот комплекс называется «Голосом Америки».

У этого «Голоса» есть многие характеристики личности. Можно сказать, что у радиостанции есть память, чувства, предпочтения, надежды, таланты. «Голос Америки» тяготится воспоминаниями о Вьетнамской войне, не любит коммунизм, но любит яблочный пирог, немного сентиментален. Это правда. Но справедливость всех этих характеристик не означает, что они персонифицированы в ком-то одном из членов команды «Голоса». Эти характеристики сближают всех, кто создает радиопрограмму. Но ни один человек не знает всего того, что «знает» «Голос Америки», и ни один человек, наверное, не согласен со всем тем, что говорит «Голос Америки». Это просто условная линия консенсуса как для внешнего мира, так и для составных частей системы. Нечто подобное верно и в отношении личности. Личность представляет собой условную линию консенсуса для связанных, но автономных когнитивных процессов и модулей мозга. Ей приписываются

собираемые общие характеристики, которыми не обладает ни один из процессов и модулей в отдельности<sup>235</sup>.

### С. Вымышленные мотивы и вымышленные агенты

Выше я показал, что психические механизмы ассоциации позволяют вполне произвольно объединять ментальные содержания в кластеры. Тем самым я стремился доказать, что определение личности — это не *fait accompli*, а непрерывающийся творческий процесс. Выше я также указал на существование механизмов ретроспективной атрибуции мотивов действий, домысливание ментальных содержаний. Так я хотел продемонстрировать, что многие атрибуты личности являются фантастическими. Но все эти выводы были основаны на аномальных случаях, на наблюдениях за пациентами с психическими дефектами. Можно ли делать экстраполяцию этих выводов на здоровых людей? Можно ли делать выводы о работе нормальной психики на основе наблюдений за нарушениями?

По моему мнению, нет оснований считать, что обнаруженные психические механизмы являются следствиями нарушений. Просто дефекты позволяют явным образом обнаружить механизмы функционирования нормальной психики. И тем не менее, я согласен, что справедливость этих выводов для здоровой психики стоило бы подкрепить отдельно. И такая возможность существует. Это можно сделать с помощью примеров, которые приводит уже упомянутый выше профессор психологии Д. Вегнер в книге «Иллюзия свободной воли». Его примеры позволяют подтвердить, что ретроспективная атрибуция ментальных содержаний и их произвольная кластеризация являются стандартными механизмами здоровой психики.

---

<sup>235</sup> Эта аналогия в том числе оправдывает вывод С. Левина о том, что согласно нарративному подходу не только люди могут быть личностями, но и корпорации. В статье «Моральная ответственность и личность корпораций: нарративный подход» он пишет: «Приведенные рассуждения призваны показать, как нарративный подход к тождеству личности может снять одно из самых существенных возражений против корпоративной моральной ответственности. Возражение строится на том, что только личности могут быть морально ответственны, а у корпораций личности нет, и поэтому они не могут быть моральными субъектами». Левин С.М. Моральная ответственность и личность корпораций: нарративный подход // Мысль: Журнал Петербургского философского общества. 2016. № 21. С. 74.

#### а) *Ретроспективная атрибуция ментальных содержаний*

«Голодная Лиса пробралась в сад и на высокой ветке увидела сочную гроздь винограда. "Этого-то мне и надобно!" — воскликнула она, разбежалась и прыгнула один раз, другой, третий... но всё бесполезно — до винограда никак не добраться. "Ах, так я и знала, зелен он ещё!" — фыркнула Лиса себе в оправдание и заспешила прочь». Эта басня Эзопа отражает вполне типичную ситуацию: ретроспективную подмену намерений. Исходным мотивом персонажа басни была добыча лакомства, но когда виноград оказался недоступным, лиса сменила свое отношение. Мотив достать виноград сменило решение игнорировать несъедобный фрукт. Подобную подмену совершают и многие люди в повседневной жизни. Часто действие оказывается противоречащим исходным интенциям, а результат — ожиданиям. Но вместо того, чтобы признать расхождение, обнаружить противоречие и удивиться, человек просто меняет исходную диспозицию: «Что? Да я вовсе и не собирался этого делать! Как вы могли подумать?» Известно, что актеры, сыгравшие роль литературного персонажа, склонны в дальнейшем разделять его позицию, даже если она противоречила их собственным изначальным установкам. А авторы эссе на заданную тему имеют склонность защищать основной тезис, даже если этот тезис им вначале казался ложным. Выходит, суждения личности о собственных предпочтениях и мотивах опираются больше на объективные наблюдения за собственным поведением, чем на интроспекцию. Самыми яркими примерами таких ретроспективных интерпретаций собственных мотивов являются примеры с *постгипнотическим внушением*.

Люди, которые получили инструкции во время гипноза, после пробуждения часто выполняют эти инструкции. При этом они не помнят внушения, не знают настоящей причины собственных действий и придумывают самые невообразимые мотивы, чтобы объяснить свое неожиданное поведение. Во время одного из сеансов гипнотизер внушил

женщине, что она должна взять книгу со стола и положить ее на полку. Пробудившись, она так и сделала. Когда женщину попросили объяснить ее действия, она сказала: «Мне не нравится, чтобы вещи были в беспорядке: место для книг — на полке. Поэтому я положила её туда». Можно предположить, что это стечение обстоятельств и что у женщины действительно были подобные желания еще до гипноза. Что ж, другой случай не оставляет сомнений в творческой природе ретроспективных объяснений. Гипнотизер внушает человеку абсурдное действие: он должен взять с окна цветочный горшок, обернуть его в ткань, положить на диван и поклониться горшку три раза. После пробуждения человек выполняет данную инструкцию. Какое же объяснение он может дать столь странному поведению? «Вы знаете, когда я пробудился, я увидел цветочный горшок там и подумал, что здесь достаточно холодно и лучше этот горшок согреть, иначе цветок погибнет. Поэтому я обернул его тканью. Далее я заметил, что диван стоит как раз около огня, поэтому положил цветок греться туда. И я поклонился потому, то остался доволен тем, насколько хорошая идея пришла мне в голову». К удивлению окружающих, сам человек не считал свои действия безумными, а определил их как вполне рациональные, ведь он смог дать им объяснение<sup>236</sup>. Мне кажется, эти случаи дополняют аргумент в пользу творческого ретроспективного редактирования мотивов. Аналогичным образом, пожалуй, можно дополнить аргумент в пользу произвольной кластеризации ментальных содержаний.

#### *b) Атрибуция ментальных содержаний виртуальным агентам*

Американский актер и радиоведущий Эдвард Берген стал популярным в 30-40ых гг. в первую очередь как чревовещатель с куклой Чарли МакКарти. Случай из его жизни позволяет предположить, что он настолько сжился с образом куклы, что искренне приписывал ей часть своих ментальных содержаний. Однажды гость зашел в комнату к Бергену и обнаружил, что тот

---

<sup>236</sup> Цитаты — по книге Wegner D. The Illusion of Conscious Will. Oxford, 2002.

не репетирует, а самым настоящим образом разговаривает со своей куклой. Берген задавал Чарли философские вопросы, а Чарли отвечал с сократической мудростью. Увидев гостя, Берген смутился, покраснел и сказал, что у него сейчас состоялась удивительно интересная беседа. На что гость заметил, что Берген вел эту беседу сам с собой. «В каком-то смысле это и так, но я задаю вопросы, а Чарли отвечает. У меня нет ни малейшей догадки, что он скажет в следующий момент, и я поражен его пронизательностью»<sup>237</sup>. Возможно, этот случай — голливудский миф. Но мне кажется, что он вполне реалистичен, учитывая творческий талант и развитое воображение актёра. Но если читателю трудно поверить в его правдоподобие или если этот пример кажется ему признаком душевной болезни, стоит просто представить на месте Э. Бергена ребёнка. Тогда такое поведение должно перестать быть аномалией. Дети воображают подобных собеседников часто. И в современной психологии это считается признаком становления здоровой психики.

Исследования показывают, что примерно 65% детей имеют *воображаемых друзей*<sup>238</sup>. Воображаемый друг — это существо, придуманное ребёнком, с которым ребёнок регулярно общается. Такими друзьями могут быть невидимые существа, одушевленные в воображении ребёнка предметы (в том числе игрушки), а также вымышленные агенты, роль которых разыгрывает сам ребёнок. Систематические исследования этого феномена психологами начались в конце XX в. И теперь есть даже основания говорить о появлении нового поля научных исследований<sup>239</sup>.

Изначально наличие вымышленных друзей исследователи интерпретировали как отклонение от нормы. Одни психологи считали, что их присутствие свидетельствует о патологической неспособности ребёнка отличать вымысел от реальности, иногда — о психической болезни или

---

<sup>237</sup> Случай пересказан по Wegner D. *The Illusion of Conscious Will*. Oxford, 2002 P. 221.

<sup>238</sup> Klaussman E., Passman R. *Pretend Companions (Imaginary playmates): The Emergence of a Field* // *The Journal of Genetic Psychology*.

<sup>239</sup> Klaussman E., Passman R. *Pretend Companions (Imaginary playmates): The Emergence of a Field* // *The Journal of Genetic Psychology*.

неблагоприятных обстоятельствах воспитания. Другие, напротив, полагали, что этот феномен свидетельствует о необычайной одаренности детей, показателем особых творческих способностей. Одним из первых, кто счел наличие вымышленных друзей этапом нормального психического развития, был швейцарский психолог Ж. Пиаже. Он сделал вывод, что общение с вымышленными друзьями служит транзитной стадией развития языковых способностей (от внутренней речи — self-talk — к социальным коммуникациям). Впоследствии мнение о том, что наличие в раннем возрасте вымышленных друзей является нормой развития личности, закрепилось. В частности, ее придерживается М. Тейлор, профессор психологии из университета Орегона, посвятившая феномену вымышленных друзей монографию. Она приводит следующие случаи в качестве примера работы здоровой детской психики.

Одна 4-х летняя девочка, участвовавшая в исследовании, рассказала о том, что дружит с двумя невидимыми птичками по имени Натси и Натси, которые живут на дереве за окном её спальни. По словам девочки, у двух Натси цветные перья, они 30 см. ростом и постоянно болтают. О двух птичках знали даже родители. Более того, они были фактически домашними питомцами. Когда семья выезжала куда-то, птички ехали на крыше машины, когда все садились за стол —птичкам оставляли место. И все вместе смеялись их глупым выходкам. Спустя два года девочка все еще помнила про своих питомцев, и её мама рассказывала, что они иногда садились с дочкой и вспоминали птичек Натси и Натси.

Это описание является наиболее характерным для случаев с вымышленными друзьями. Однако существует множество вариаций. Вымышленные персонажи могут принимать самые различные формы и воплощения. У некоторых детей есть вымышленные копии их реальных друзей, у некоторых — это копия персонажа из фильма, другие одушевляют мягкие игрушки и даже обычные предметы. Так, у одного мальчика была длительная и бурная история отношений с комодом. Другие же

«разыгрывают» вымышленных персонажей сами. Мама одной девочки сообщила, что ее дочка дружит с вымышленным существом по имени Applejack (Яблочный Пирожок). Когда психологи стали расспрашивать девочку, как часто она с ним играет, та стала недовольно возражать: «Это я, я и есть Applejack». Оказывается, мама не разобралась, что девочка не столько дружила с вымышленным героем, сколько иногда им была.

Вымышленные персонажи в детстве — это распространенное явление и признак нормального развития психики. Ни большое количество таких воображаемых друзей, ни многообразие их форм не является симптомом отклонений. Но даже существование таких «друзей» во взрослом возрасте не обязательно является аномалией. Это подтверждает другое исследование М. Тейлор, посвященное отношениям между авторами художественных произведений и их персонажами. По мнению Тейлор, их отношения во многом схожи с отношениями между детьми и их воображаемыми друзьями.

Ключевое сходство в том, что воображаемые персонажи представляются в обоих случаях как автономные самостоятельные агенты с собственными мыслями, чувствами и действиями. Это выражается в непредсказуемости и неподконтрольности человеку их поведения. Дети часто жалуются, что их воображаемые друзья не ведут себя так, как они хотели бы: не делятся тем, что у них есть, не всегда появляются, когда они этого хотели бы, или, наоборот, надоедают им своим присутствием, разговаривают слишком громко, мешают. На подобную автономию вымышленных персонажей указывают также писатели. В интервью национальному общественному радио Джоан Роулинг, автор мирового бестселлера про Гарри Поттера, на вопрос о том, почему главный герой её произведения мальчик, а не девочка, отвечала: «Спустя шесть месяцев после начала работы над романом я подумала, что я женщина, а он мальчик. Но было уже слишком поздно, чтобы сделать его девочкой Гэри Гариет. Он был для меня реальным как мальчик, и надеть на него платье означало бы совершить



насилие над ним. Я никогда не пишу так: "Ок, мне нужен такой-то герой". Мои герои являются ко мне в результате мистического процесса, который никто не понимает. Они просто появляются».

Некоторые авторы «живут» со своими персонажами, беседуют, спорят. Иногда авторы даже вынуждены вести переговоры с героями об их судьбе в романе.

Сара Парецки в интервью радио сообщала, что ей пришлось идти на уступки, когда по сюжету ее герой должен был попасть в тюрьму. Герой отказывался идти на это до тех пор, пока она не пообещала ему взамен настоящую любовь. Подобные комментарии, комментарии о независимой и «свободной» воле персонажей, встречаются у множества писателей и сценаристов. Среди них — Достоевский, Сартр, Пруст, Джеймс, Тарантино. Английский романист Э.М. Форстер так описывает процесс создания романа: «Герои появляются тогда, когда их пробуждаешь, но полные бунтовского духа. Из-за того, что у них столь много общего с людьми, они пытаются, как мы, прожить собственную жизнь и часто вступают в заговор против основного сюжета книги. Они сбегают, выходят из-под контроля: они — творения внутри творения, и часто не помещаются в него. Если им дать полную свободу, они запинаят книгу ногами, пока она не разлетится по страницам, но если слишком сильно держать их в узде, они отомстят своей смертью и отравят ее своим гниением»<sup>240</sup>.

Описание Форстера показывает устойчивость иллюзии независимого существования персонажей. И является хорошим подтверждением нашей гипотезы, гипотезы о том, что здоровая психика может компоновать, кластеризовать ментальные содержания так, чтобы они образовывали нарративы. Если этот нарратив один, в результате его интерпретации появится единственный агент, если множество — агентов тоже будет множество. В любом случае, за процессом индивидуации личности стоит

---

<sup>240</sup> Taylor M. The Illusion of Independent Agency // *Imagination, Cognition and Personality*. 2003. No. 22. June, P. 364

творческий процесс. Таким образом, приведенные примеры должны показать, что аналогии Деннета, приведенные выше, релевантны по отношению к личности. Они укрепляют гипотезу, что личность — это абстрактная сущность, некоторая удобная фикция, рождающаяся в ходе интерпретации нарратива. Эта гипотеза, в свою очередь, позволяет нам разрешить проблему, с которой мы начинали эту главу — Проблему исчезающей личности.

Впрочем, особо внимательный читатель, мог бы на это возразить: если личность это — фикция, и, как показано в диссертации, личность необходима для реализации свободы воли и моральной ответственности, значит, свобода воли и моральная ответственность — тоже фикция? Как это следствие соотносится с компатибилистической позицией автора? Это очень важное замечание, и я должен в заключение представить на него ответ. Да, я действительно считаю свободу воли и моральную ответственность в некотором смысле фикциями, но только в том смысле, в каком их обычно отстаивают сторонники субстанционального дуализма или агентского либертарианства. Свобода воли и моральная ответственность, необходимые для рационализации существующих межличностных отношений, напротив, существуют. Референтами этих понятий служат объективно существующие абстрактные паттерны в мире, которые, в свою очередь, являются неотъемлемой частью описания реальности. Свобода воли, моральная ответственность и личность — это фикции, но *полезные* фикции, место которым найдется в науках так же, как для понятий центр тяжести, вектор движения или ген. Такой онтологии придерживается и Деннет, называя её «умеренным реализмом»<sup>241</sup>. Однако наиболее ёмкое и запоминающееся название этой онтологической позиции дал профессор философии

---

<sup>241</sup> Dennett D. Real Patterns // The Journal of Philosophy. 1991. Vol. 88, No. 1. P. 30.

Кейптаунского университета Дон Росс. Он назвал её «реализмом тропических лесов»<sup>242</sup>.

## VII. Итоги третьей главы

Как было показано в тексте третьей главы, событийный каузализм создаёт опасность для существования и тождества личности. Её можно назвать Проблемой исчезающей личности. По мнению множества философов, существование и тождество личности являются условиями моральной ответственности, но если событийный каузализм верен, то эти условия не соблюдаются, личность как бы «исчезает» и некому оказывается приписывать ответственность. Препятствием для признания наличия свободы воли и ответственности также является противоречие между обыденными интуициями о свойствах личности и эмпирическими данными о том, как устроен человек и его мозг. Это можно обозначить, вслед за У. Селларсом, проблемой согласования «данности и научных взглядов»<sup>243</sup> в отношении личности и ее действий. Эта проблема была обнаружена уже философами Нового времени, на нее указывали Г. Лейбниц и Д. Юм. Современная разработка этой проблемы сделана, в частности, Д. Деннетом. Он делает выводы о том, что рациональность, сознательность, самосознание и другие классические условия существования личности не могут быть полностью удовлетворены в принципе, а личности как конкретной, единой сущности не существует.

В поисках решений этих проблем мы обратились к современным теориям личности. Были проанализированы психологический подход Дж. Перри и С. Шумейкера, биологический подход Э. Олсона и субстанциональный подход Р. Суинбёрна. В ходе исследования было показано, что эти подходы оказываются неприменимы к некоторым гипотетическим ситуациям, таким «Телетранспортер» Д. Парфита и «Где Я?»

---

<sup>242</sup> Ross D. Rainforest Realism: A Dennettian Theory of Existence // Dennett's Philosophy: A Comprehensive Assessment. Cambridge: MIT Press, 2000. P. 147.

<sup>243</sup> Sellars W. Science, Perception and Reality. N.Y.: Ridgeview Pub. Co., 1991.

Д. Деннета. Эти гипотетические ситуации еще нереализуемы, однако они выявляют проблемы концептуальных построений современных аналитических философов. Возражения к описанным подходам были резюмированы в Аргументе пяти трансплантаций. В контексте этого аргумента и представленных мысленных экспериментов вновь встал вопрос о наличии свободы и основаниях приписывания ответственности.

В качестве наиболее перспективного решения был представлен нарративный подход в том его виде, как его предлагает Д. Деннет. Этот подход не является классическим, так как он заменяет вопрос о реидентификации вопросом о характеристике личности. Но именно он позволяет решить проблему рациональных оснований ответственности. Согласно разработанной Деннетом концепции, каркасом личности является биологический организм. Одна из функций организма — создание расширенного фенотипа. Расширенным фенотипом человека служат в том числе созданные им тексты. Именно эти тексты обеспечивают связь атрибутов человека в прошлом и настоящем, его ментальных содержаний и действий. Таким образом, сама структура текста позволяет избежать вопроса о тождестве личности во времени и заменить его вопросом о принадлежности. Она позволяет снять требование нумерического тождества во времени и заменить его требованием связанности, что успешно обеспечивается текстом. Структура текста и его смысловая направленность позволяют интерпретировать события, описываемые в тексте, как результат деятельности какого-то определенного агента. Единый агент — это допущение, результат интерпретации, виртуальная, теоретическая сущность.

Этот агент является абстрактной, а не конкретной сущностью. Поэтому найти его ни интроспективно, ни объективными методами не удастся. Личность как фиктивная сущность наделяется фантазийными характеристиками, такими как полная рациональность, целостность, неделимость, уникальность. Но эти характеристики в абсолютном смысле недостижимы. И личность всегда обладает ими только «отчасти».

В концепции Деннета, личность — это фантазия, порожденная структурой текста. Но это полезная, удобная фантазия, так как она позволяет объяснить и предсказать поведение биологического организма и наладить сотрудничество между людьми. Ее удобством обуславливается повсеместное принятие этой теоретической конструкции. А повсеместное принятие является основанием для возможности ответных реакций и моральной ответственности.

## ЗАКЛЮЧЕНИЕ

Результатом данной работы является представление и оценка теорий моральной ответственности и свободы воли в аналитической философии конца XX — начала XXI вв. В ходе исследования выявлены ключевые фигуры дискуссий, проведена систематизация и сравнение позиций, определены наиболее важные работы и представленные в них аргументы, осуществлен анализ этих аргументов и мысленных экспериментов, оценена их достоверность, а также представлены собственные варианты усовершенствования наиболее перспективных позиций. В работе также проверена гипотеза о том, что препятствиями для моральной ответственности и необходимой для неё свободы воли является не только детерминизм, но и каузальная замкнутость физического мира, а также проблема асинхронного тождества личности. В результате исследования сделано заключение о том, что современным аналитическим философам в целом удастся преодолеть эти препятствия, т.е. представить рациональные основания моральной ответственности и показать возможность свободы воли. Этот вывод получен в несколько этапов.

Первым этапом работы был анализ современных теорий ментальной каузальности. Было показано, что каузальная эффективность ментального является определяющим элементом действия, поэтому ментальная каузальность служит одним из метафизических условий свободы воли и моральной ответственности. К этому выводу я пришел, рассмотрев сначала аргументы Д. Дэвидсона в пользу «тезиса о ментальной каузальности», а затем Аргумент исключения, направленный против возможности ментальной каузальности, сформулированный Дж. Кимом. В поисках ответа на этот аргумент я обратился к наиболее репрезентативным концепциям ментальной каузальности. Мной были проанализированы решения этого противоречия в биологическом натурализме, телеофункционализме и локальном

интеракционизме, а также критика позиции Кима со стороны Н. Блока с помощью Аргумента обобщения.

Вторым этапом работы было исследование возможности соблюдения двух других принятых в аналитической философии условий свободы воли и моральной ответственности — принципа автономии и принципа альтернативных возможностей. Традиционно считается, что совместимость этих принципов с детерминизмом и даже с индетерминизмом является сложной философской проблемой. Во второй главе я рассмотрел альтернативные позиции современных философов по этому вопросу: либертарианство Р. Кейна и твердый индетерминизм Перебума. Были исследованы основные аргументы и проведена их систематизация. В частности, были проанализированы Аргумент последствий и Аргумент манипуляций. В конце главы были представлены основные доводы в пользу компатибилистской позиции, подобной позициям Д. Деннета, Г. Франкфурта и Дж. Фишера.

Третий этап исследования был посвящен современным теориям тождества и существования личности, а также проблеме атрибуции действий агенту. Атрибуция действий также является необходимым условием свободы воли и моральной ответственности, так как без связи действия с личностью невозможно приписывать ответственность. Первая часть третьей главы была сфокусирована на проблеме идентификации личности, или определения условий для включения существа в категорию личностей. В этой части проанализирована концепция личности у Д. Деннета и Г. Франкфурта. Вторая часть главы посвящена проблеме реидентификации личности. В качестве предлагаемых решений реконструированы субстанциональная теория Р. Суинбёрна, психологическая теория С. Шумейкера и Дж. Перри, биологический подход Э. Олсона, нарративная теория Д. Деннета. В ходе работы были проанализированы основные критические аргументы: Аргумент Редупликации Б. Вильямса и мысленный эксперимент «Телетранспортер» Д.

Парфита. Анализ аргументов показал, что наиболее перспективным является нарративный подход. Доводы сторонников нарративного подхода были соотнесены с эмпирическими исследованиями Д. Вегнера и М. Газзаниги, а также дополнены собственными соображениями. В итоге работы были обоснованы следующие положения.

Наиболее убедительной теорией действия является стандартный каузализм, предложенный Д. Дэвидсоном в 60-х гг. XX в. Согласно стандартному каузализму, действие определяется по каузальным основаниям. На стандартном каузализме основывается «тезис о ментальной каузальности» (или ментальный каузализм) — утверждение о том, что ментальная каузальность необходима для свободного и ответственного действия. Наиболее существенным препятствием на пути к возможности ментальной каузальности является «тезис о физикализме». Это утверждение о том, что все события, в том числе действия, являются физическими событиями и в качестве причины имеют только другие физические события. «Тезис о физикализме» лежит в основе Аргумента исключения, используемого Дж. Кимом для опровержения ментальной каузальности. Таким образом, «тезис о ментальной каузальности» и «тезис о физикализме» составляют противоречие. Многие современные аналитические философы критикуют Аргумент исключения и пытаются снять выявленное Кимом противоречие.

Сторонники локального интеракционизма и биологического натурализма представляют альтернативные решения проблемы ментальной каузальности. Сравнительный анализ этих подходов показывает, что их общность — в утверждении нередуцируемости ментальных свойств. Оба этих подхода ставят задачей вписать ментальную каузальность в физическую картину мира. Но как локальный интеракционизм, так и биологический натурализм обладают существенными недостатками. Локальный интеракционизм во многом зависит от успешности Аргумента каузальных траекторий. Однако этот аргумент уязвим для критики с помощью



предложенного мной Аргумента-двойника. Биологический натурализм содержит внутренние противоречия и слишком широко определяет понятие причины. Как было показано, каузальное отношение в интерпретации Дж. Сёрла включает в себя отношения, которые в аналитической традиции принято называть «реализацией» и «супервентностью». Иллюзия решения проблемы ментальной каузальности в биологическом натурализме существует, таким образом, только благодаря путанице в определениях.

Тем не менее, ментальная каузальность может быть спасена. По моему мнению, решение должно состоять из трех шагов. Первым шагом следует считать критику Аргумента исключения с помощью Аргумента обобщения, предложенного Н. Блоком. Этот аргумент удачен и показывает, что Аргумент исключения приводит к абсурдным следствиям. Следующий шаг — выявление ошибки в рассуждении. По моему мнению, эта ошибка заключается в формулировке принципа каузальной замкнутости физического мира. Последний этап — это позитивное решение, демонстрация возможности ментальной каузальности. Такая демонстрация предложена в телеофункционализме Д. Деннета. Это решение наиболее перспективно и может быть усовершенствовано. Мысленный эксперимент Библиотека первых изданий, предложенный мной, является полезным дополнением к телеофункционалистским иллюстрациям. Таким образом, было показано, что современная аналитическая философия обладает достаточными аргументами в пользу возможности совмещения убеждения о ментальной каузальности и истинности физикализма. Тем не менее, для возможности существования свободы воли и моральной ответственности есть и другие препятствия. Традиционно таким препятствием считается детерминизм.

Современная аналитическая философия выступает преемницей классической философии в том смысле, что ее представители продолжают опираться на два принципа свободы воли и ответственности: принцип альтернативных возможностей и принцип автономии. То есть современные

философы считают, что для наличия свободы воли необходимы возможность поступать иначе и автономия принятия решений. Однако в связи с работами Г. Франкфурта, в частности его статьей «Альтернативные возможности и моральная ответственность», первый принцип утратил решающее значение. Для современных инкомпатибилистов именно принцип автономии личности (принцип быть ответственным в конечном счете) является более фундаментальным. Такое убеждение разделяют как сторонник либертарианства Р. Кейн, так и сторонник твердого инкомпатибилизма Д. Перебум. Сравнительный историко-философский анализ показывает, что основным источником их разногласий, однако, служит вопрос совместимости ответственности и свободы воли с индетерминизмом. Перебум и Кейн считают свободу и ответственность несовместимыми с детерминизмом, но Перебум также считает, что они несовместимы и с индетерминизмом. Ключевым рассуждением здесь является Аргумент исчезающего агента, предложенный Д. Перебумом. Этот аргумент следует отнести к традиции Аргументов случайности, восходящих еще к Д. Юму, однако в формулировке Д. Перебума он содержит ряд существенных усовершенствований.

Впрочем, оба варианта инкомпатибилизма, либертарианство и твердый инкомпатибилизм, являются не до конца убедительными. В ходе исследования показано, что современные компатибилисты могут парировать основные доводы противников: Аргумент манипуляций в формулировке Д. Перебума и Аргумент последствий в формулировке П. ван Инвагена. Аргумент манипуляций активно критикуется Д. Нелкин и С. Вульф. Но с моей точки зрения, наиболее перспективной является предложенная мной критика с помощью дополнительных сценариев манипуляций. Эти дополнительные сценарии показывают, что в отдельных случаях манипуляций у агента возникает моральная ответственность. Аргумент последствий также может быть успешно атакован. В пользу этого вывода

могут быть использованы рассуждения Г. Франкфурта и Д. Деннета. Они показывают, что ошибка аргумента, в частности, в том, что он слишком узко интерпретирует понятие «иметь возможность поступить иначе». Исходя из этих рассуждений, можно сделать вывод о совместимости свободы воли и моральной ответственности с детерминизмом. Однако важно прояснить также их отношение с индетерминизмом.

Ключевым основанием считать индетерминизм и свободу несовместимыми является аргумент Исчезающего агента в формулировке Д. Перебума. Этот аргумент, по мнению его автора, демонстрирует недостаточность контроля в случае индетерминированных действий. Однако многие аналитические философы сомневаются в справедливости такого вывода. В работе проанализирована критика аргумента Исчезающего агента А. Миле и К. Франклином и выявлены ее сильные и слабые стороны. Этот анализ позволил показать, что в сравнении с их возражениями наиболее перспективным направлением критики служит предложенный авторский Аргумент четырехмерного агента. Этот аргумент основан на идеях четырехмерного существования личности, предложенных Д. Льюисом, а также вбирает в себя интуиции, защищаемые Миле и Франклином. Результатом этого и предыдущих рассуждений служит обоснование суперкомпатибилизма — наиболее распространенной среди современных компатибилистов позиции. Суперкомпатибилизм — это позиция, согласно которой не только детерминизм, но и индетерминизм совместим с моральной ответственностью и свободой воли. Правда, необходимо преодолеть еще одно философское препятствие — проблему тождества личности во времени и атрибуции действия такой личности.

В результате исследования было показано, что основным препятствием для существования тождественной во времени личности является событийный каузализм. Событийный каузализм — это теория действия, согласно которой причиной действий может быть только событие. Этот вид

каузализма совместим как с позицией Д. Дэвидсона, так и с позициями его оппонентов. Противниками событийного каузализма можно считать только некоторых либертарианцев, таких как К. Лерер, Т. О'Коннор и Р. Кларк. В диссертации показано, что теории этих философов неправдоподобны. Однако из тезиса о «событийном каузализме», похоже, следует исчезновение делящегося во времени и ответственного агента. В поисках способов преодоления этой проблемы я обратился к современным теориям личности.

Исследование и систематизация теорий личности показали превалирование психологического подхода. Он разделяется такими философами, как П. Грайс, Дж. Перри, С. Шумейкер. Современный психологический подход является развитием классической теории Дж. Локка, однако решает множество сопутствующих ей трудностей, обнаруженных еще Дж. Батлером и Т. Ридом. Тем не менее, психологические теории остаются уязвимыми для критики. Брешь в этом подходе обнаруживается с помощью Аргумента редупликации Б. Вильямса и мысленного эксперимента «Телетранспортер», предложенного и подробно проанализированного Д. Парфитом в фундаментальной работе «Причины и личности». В поисках удовлетворительного решения проблемы тождества личности я проанализировал альтернативные теории личности: субстанциональный подход Р. Суинбёрна и биологический подход Э. Олсона. Анализ этих теорий, однако, показал, что и в них не обнаруживается приемлемого решения проблемы тождества личности.

Наиболее удачным выходом из этой ситуации я считаю нарративный подход. Его представителями являются Д. Деннет и М. Шехтман. Нарративный подход является неклассическим в том смысле, что он заменяет вопрос об отождествлении личностей в разное время вопросом атрибуции и характеристики личности. Он позволяет снять требование нумерического тождества во времени и заменить его требованием связанности, что успешно обеспечивается текстом. Это дает возможность обосновать существование

ответственности личности за действия в прошлом. Однако возникает коллизия между нарративным подходом и обыденными интуициями в отношении природы личности. Эта коллизия может быть снята посредством развития интуиции с помощью метафор и аналогий, которые использует Деннет. Уподобление личности центру нарративной гравитации может поддерживаться также эмпирическими открытиями, которые приведены в научных работах Д. Вегнера и М. Газзаниги.

Резюмировать исследование можно выводом о том, что современные аналитические философы значительно продвинулись в решении классических философских вопросов и в текущий момент значительно ближе подошли к формированию теории свободы и ответственности, а также решению большинства связанных с этим философских загадок, чем их предшественники. Свобода воли, скорее всего, реализуема в мире вне зависимости от того, является наш мир детерминистическим или индетерминистическим, а применение моральной ответственности к личностям оправдано.

## БИБЛИОГРАФИЯ

### Тексты на русском языке

1. Брентано Ф. Психологии с эмпирической точки зрения. Т.2, раздел 7. Представления и суждения как два различных класса психических феноменов // Вопросы философии. 1995. № 2. С. 148-171.
2. Беседин А. может ли детерминизм нами манипулировать // Логос. 2016. №5. С. 201- 212.
3. Беседин А. Натуралистический аргумент П. Стросона в пользу моральной ответственности // Вестник Московского Университета. 2015. №6. С. 8-16.
4. Васильев. В.В. Трудная проблема сознания. М.: Прогресс-Традиция, 2009.
5. Васильев В.В. Сознание и вещи: Очерк феноменалистической онтологии. М.: Книжный дом Либроком, 2014.
6. Васильев В.В. В защиту классического компатибилизма // Вопросы философии. 2016. № 2. С. 64–76.
7. Васильев В.В. В защиту классического компатибилизма: Эссе о свободе воли. М.: Ленанд, 2017.
8. Волков Д.Б. Бостонский зомби: Д. Деннет и его теория сознания. М.: Либроком, 2012.
9. Волков Д.Б. Психологический подход как решение проблемы тождества личности // Философская мысль. № 2. 2017. С.13-25
10. Волков Д.Б. Решение проблемы ментальной каузальности в биологическом натурализме Дж. Сёрла // Философская мысль. № 2. 2017. С.1-12.
11. Волков Д.Б. Современный субстанциональный подход к проблеме тождества личности // Философия и культура. № 1. 2017. С.77-85.

12. Волков Д.Б. Рецензия на книгу Д. Перебума "Свобода воли, агент и смысл жизни" // Вестник томского университета. № 4 (36) 2016. С. 404-409.
13. Волков Д.Б. Аргумент исключения Дж. Кима и проблема ментальной каузальности. // Вестник МГУ. № 6. 2016. С. 15-33.
14. Волков Д.Б. Нарративный подход как решение вопроса о тождестве личности // Вестник СПбГУ. № 4. 2016. С. 21-32.
15. Волков Д.Б. "Тезис о сознании" и моральная ответственность в исследованиях Нила Леви // Логос. № 5. 2016. - С. 213-226.
16. Волков Д.Б. Являются ли психопаты морально ответственными личностями? // Вестник Томского государственного университета. 2016. № 2 (34). С. 212-222.
17. Волков Д.Б., Деннет Д. Об организации конференции в Гренландии «Проблемы сознания и свободы воли в аналитической философии» // Вестник Московского Государственного университета. 2015. № 6. С. 3-7.
18. Волков Д.Б. Опровергает ли Аргумент каузальных траекторий локальную супервентность ментального над физическим // Эпистемология и философия науки. 2015. Т. 2 (44). С. 166-182.
19. Волков Д.Б. Проблема ментальной каузальности: обзор новейших исследований // Философия и культура. 2016. № 6. С. 673-682.
20. Волков Д.Б. Проблема свободы воли. Обзор ключевых исследований конца XX - начала XXI вв. в аналитической философии // Философский журнал. 2016. № 3. С. 175-189.
21. Волков Д.Б. Свобода воли в либертарианстве Кейна // Вестник Московского Государственного университета. 2015. № 6. С. 32-51.
22. Волков Д.Б. Споры о сознании и аналитическая философия // Эпистемология и философия науки. 2007. № 3. С. 176-181.

23. Волков Д.Б. Что доказывают манипуляции с «Аргументом манипуляции» Д. Перебума? // *Философия и культура*. 2015. № 6. С. 933-942.
24. Дубровский Д. И. Проблема идеального. Субъективная реальность. М.: Мысль, 1983 (М: Канон+, 2002).
25. Дубровский Д.И. Сознание, мозг, искусственный интеллект. М.: ИД Стратегия-Центр, 2007.
26. Иванов Д. Природа феноменального сознания. М.: Либроком, 2013.
27. Кузнецов А. Локальный интеракционизм. Новое решение проблемы ментальной каузальности // *Вопросы философии*. 2016. №1. С. 148-161.
28. Кузнецов А.В. Конститутивный панпсихизм и проблема ментальной каузальности // *Вестник Московского Университета. Серия 7: Философия*. 2015. №6. С. 106-117.
29. Кузнецов А.В. Проблема ментальной каузальности в аналитической философии сознания. Диссертация на соискание степени кандидата философских наук: 09.00.03 – История философии. М., 2016.
30. Левин С.М. Физикализм, субъект и метафизическая свобода воли // *Вестник ленинградского государственного университета имени А.С. Пушкина*. 2013. № 2 (2). С. 67-77.
31. Левин С.М. Коллективное сознание и моральная ответственность // *Вестник Томского государственного университета*. 2016. №3. С. 46-53.
32. Левин С.М. Моральная ответственность и личность корпораций: нарративный подход // *Мысль: Журнал Петербургского философского общества*. 2016. № 21. С. 63-76.
33. Локк Дж. Сочинения в 3-х т. Т. 1. Опыт о человеческом разумении / Под ред. И.С. Нарского. М.: Мысль. 1985.
34. Мареева Е.В. Дж. Сёрл: старое и новое в понятии сознания // *Философия сознания: история и современность*. М., 2003. С. 216-231.



35. Нагуманова С.Ф. Материализм и сознание: анализ дискуссии о природе сознания в современной аналитической философии. Казань: Казан ун-т, 2011.
36. Проблема сознания в философии и науке. Под ред. Дубровского Д. И. М.: Канон+, 2009.
37. Разеев Д. Н. Ложные воспоминания и тождество личности // Вопросы философии. 2015. №1. С. 45-50.
38. Разин А.В. Свобода и необходимость // Вестник Российского университета дружбы народов. Серия: Философия. 2015. №2. С. 9-15.
39. Секацкая М.А. Свобода воли и предсказуемость. Философский анализ современных исследований в нейронауке. // Вопросы философии. 2016. №3. С. 163–169.
40. Секацкая М. А. Интенциональность, свобода воли и моральная ответственность коллективов // Журнал Петербургского философского общества. 2016. № 21. С. 77-86.
41. Секацкая М. А. Пересадка мозга и тождество личности: альтернативная интерпретация одного мысленного эксперимента // Эпистемология и философия науки. 2014. Т. 42. №4. С. 67-76.
42. Секацкая М.А. Физический детерминизм и свобода воли: картезианское решение // Философские науки. 2010. №10. С. 65-80.
43. Сёрл Дж. Открывая сознание заново. М.: Идея-Пресс, 2002.
44. Сёрль Дж. Рациональность в действии. М.: Прогресс-Традиция, 2004.
45. Скворцов А.А. Этика : учебник для бакалавров. Под ред. Гусейнова А.А. М.: Издательство Юрайт, 2016.
46. Чирва Д.В. «Правильная» постановка вопроса о свободе: деятельностный индетерминизм // Этносоциум и межнациональная культура. 2015. №10 (88). С. 87-93.
47. Юлина Н.С. Философский натурализм: о книге Дэниела Деннета «Свобода эволюционирует». М.: Канон+, 2007.

48. Юм Д. Исследование о человеческом познании. // Д. Юм. Собрание сочинений в 4-х т. Т. 4. М.: Канон+РООИ Реабилитация, 2009.

49. Юм Д. Трактат о человеческой природе // Сочинения в 2-х тт. М.: Мысль, 1996.

#### Тексты на английском языке

1. Anscombe G. Intention. New York: Ithaca, Cornell University Press, 1963.
2. Anscombe G. The Collected Philosophical Papers of G.E.M. Anscombe // Metaphysics and the Philosophy of Mind. Vol. 2. Oxford: Basil Blackwell Publisher, 1981.
3. Antony L. I'm a Mother, I Worry // Philosophical Issues. 1995. No. 6. P. 6-160.
4. Antony L. The Causal Relevance of the Mental: More on the Mattering of Minds // Mind & Language. 1991. No. 6. P. 295–327.
5. Antony L. and Levine J. Reduction with Autonomy // Philosophical Perspectives. 1997. No. 11. P. 83–105.
6. Antony L. Anomalous Monism and the Problem of Explanatory Force // The Philosophical Review. 1989. Vol. 48. No.2, April. P. 153-187.
7. Atkins K. Narrative Identity and Moral Identity. A Practical Perspective. Oxford: Routledge, 2010.
8. Austin J. Ifs and Cans // Proceedings of the British Academy. 1956. Vol. 42. P. 109–132.
9. Beckermann A., Flohr H., Kim J. Downward Causation in Emergentism and Non-Reductive Physicalism // Emergence or Reduction? Essays on the Prospects of Nonreductive Physicalism. Berlin: Walter de Gruyter, 1992.
10. Block N. Anti-Reductionism Slaps Back // Philosophical Perspectives. 1997. No. 11. P. 107-132.
11. Block N. Can the Mind Change the World? // Boolos G. (ed.). Meaning and Method: Essays in Honor of Hilary Putnam. Cambridge: Cambridge University Press, 1990. P. 137-170.

12. Block N. Do Causal Powers Drain Away? // *Philosophy and Phenomenological Research*. 2003. No. 67. P. 133-150.
13. Bourget D., Chalmers D. What Do Philosophers Believe? // *Philosophical Studies*. 2014. No. 170. P. 465 – 500.
14. Chalmers D. Facing up the Problem of Consciousness // *Journal of Consciousness Studies*. 1995. No. 2 (3). P. 200-219.
15. Chalmers D. *The Conscious Mind. In Search of a Fundamental Theory*. N.Y.: Oxford University Press, 1997 (1996).
16. Chisholm R. *Human Freedom and The Self // The Lindley Lecture*, University of Kansas, April 23, 1964.
17. Christman J. Narrative Unity as a Condition of Personhood // *Metaphilosophy*. 2004. 35, No. 5. P. 695-713.
18. Colyvan, M. Can the Eleatic Principle Be Justified? // *Canadian Journal of Philosophy*. 1998. Vol. 28, No. 3. P. 313–336.
19. *Companion to Analytic Philosophy*. Ed. Martinich A., Sosa D. Malden: Blackwell Publisher, 2001.
20. Cornell D., Warren, J., Hawk, G. Psychopathy in Instrumental and Reactive Violent Offenders // *Journal of Consulting and Clinical Psychology*. 1996. No. 64. P. 783 -790.
21. Crane T. Causation and Determinable Properties: On the Efficacy of Colour, Shape, and Size // Hohwy J. and Kallestrup J. (eds.). *Being Reduced: New Essays on Reduction, Explanation, and Causation*. Oxford, 2008. P. 176-195.
22. Crane T. Mental Causation and Mental Reality // *Proceedings of the Aristotelian Society*. 1992. Vol. 92. P. 185–202.
23. Crane T. Why Indeed? Papineau on Supervenience // *Analysis*. 1991. Vol. 51. No. 1. P. 32-37.
24. Crane T. and Mellor D. There is No Question of Physicalism // *Mind*. 1990. No. 99. P. 185–206.

25. Crane T. *Elements of Mind: An Introduction to the Philosophy of Mind*. Oxford: Oxford University Press, 2001.
26. Davidson D. *Thinking Causes* // Heil J. and Mele A. (eds). *Mental Causation*. Oxford: Clarendon Press. 1993. P. 3–17.
27. Davidson D. *Actions, Reasons, and Causes* // *Journal of Philosophy*. 1963. No. 60. P. 685–700.
28. Davidson D. *Essays on Actions and Events*. Oxford: Clarendon Press, 1980.
29. Davidson D. *Mental Events* // *The Essential Davidson*. Oxford, New York: Oxford University Press, 2006. P. 105-121.
30. Dennett D. *From Bacteria to Back and Back. The Evolution of Minds*. N.Y.: Norton and Company, 2017.
31. Dennett D. *Brainchildren: Essays on Designing Minds*. Cambridge, Massachusetts: The MIT Press, 1998.
32. Dennett D. *Brainstorms: Philosophical Essays on Mind and Psychology*. Cambridge, Massachusetts: The MIT Press, 1981.
33. Dennett D. *Conditions of Personhood* // Rorty A (ed.). *The Identities of Persons*. Berkley: University of California Press, 1976.
34. Dennett D. *Consciousness Explained*. Boston, Massachusetts: Little Brown & Co, 1991.
35. Dennett D. *Darwin's Dangerous Ideas. Evolution and the Meaning of Life*. New York: Touchstone, 1995.
36. Dennett D. *Elbow Room: The Varieties of Free Will Worth Wanting*. Cambridge, Massachusetts: MIT Press, 2015.
37. Dennett D. *Freedom Evolves*. N.Y.: Penguin Books, 2003.
38. Dennett D. *Intuition Pumps and Other Tools for Thinking*. N.Y.: W.W. Norton & Company, 2013.
39. Dennett D. *Mechanism and Responsibility* // Honderich T. (ed.). *Essays on Freedom of Action*. London: Routledge & Kegan Paul, 1973. P. 159-184.
40. Dennett D., Taylor C. *Who's Still Afraid of Determinism? Rethinking Causes and Possibilities* // Kane R. (ed.). *The Oxford Handbook of Free*

- Will: Second Edition (2 ed.). Oxford: Oxford University Press, 2011. P. 221-242.
41. Dennett D. Real Patterns // *The Journal of Philosophy*. 1991. Vol. 88, No. 1. P. 27-51.
42. Dennett, D. The Self as a Center of Narrative Gravity // Kessel F., Cole P. and Johnson D. (eds.) *Self and Consciousness: Multiple Perspectives*. Hillsdale, NJ: Erlbaum, 1992. P. 103-115.
43. Dretske F. Mental Events as Structuring Causes of Behavior // Heil J. and Mele A. (eds.). *Mental Causation*. Oxford: Clarendon Press. 1993. P. 121-136.
44. Dretske F. Reasons and Causes // *Philosophical Perspectives*. 1989. No. 3. P. 1-15.
45. Dretske F. *Explaining Behavior: Reasons in a World of Causes*. Cambridge, MA: MIT Press, 1991.
46. Duff A. Psychopathy and Moral Understanding // *American Philosophical Quarterly*. 1977. Vol. 14, No. 3. P. 189-200.
47. Ekstrom L. Free Will is Not a Mystery // Kane R. (ed.). *The Oxford Handbook of Free Will: Second Edition (2 ed.)*. Oxford: Oxford University Press, 2011. P. 366-380.
48. Ekstrom L. Volition and the Will // T. O'Connor and C. Sandis (eds.) *A Companion to the Philosophy of Action*. New Jersey: Wiley Blackwell. 2010. P. 99-107.
49. Faigenbaum G. *Conversation with John Searle*. Mexico: LibrosEnRed, 2003.
50. Fischer J. *Deep Control: Essays on Free Will and Value*. Oxford: Oxford University Press, 2012.
51. Fischer J. Frankfurt-Type Examples and Semi-Compatibilism // Kane R. (ed.), *The Oxford Handbook of Free Will: Second Edition (2 ed.)*. Oxford: Oxford University Press, 2011. P. 243-265.
52. Fischer J. *My Way: Essays on Moral Responsibility*. Oxford: Oxford University Press, 2006.

53. Fischer J., Kane R., Pereboom D., Vargas M. *Four Views on Free Will*. M. Malden: Blackwell, 2007.
54. Fischer J., Ravizza M. *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press, 1998.
55. Fischer J. *Recent Work on Moral Responsibility // Ethics*. 1999. No. 110 (10). P. 93–139.
56. Fischer J. and Ravizza M. *Responsibility and Control: A Theory of Moral Responsibility*. New York: Cambridge University Press, 1998.
57. Fischer J. and Tognazzini N. *The Physiognomy of Responsibility // Philosophy and Phenomenological Research*, 2011. Vol. 82. No. 2. P. 381–417.
58. Fischer J. *Moral Responsibility*. Ithaca: Cornell University Press, 1986.
59. Fischer J. Ravizza M. *Perspectives on Moral Responsibility*. Ithaca: Cornell University Press, 1993.
60. Fischer J. *The Frankfurt Cases: The Moral of the Stories // Philosophical Review*. 2010. Vol. 119. No. 3. P. 315–336.
61. Fischer J. *The Metaphysics of Free Will: An Essay on Control*. Cambridge, MA: Wiley-Blackwell, 1994.
62. Fodor J. *Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology // Behavioral and Brain Sciences*. 1980. Vol. 3, No. 1. P. 63–73.
63. Fodor J. *Special Sciences: Still Autonomous After All These Years // Philosophical Perspectives*, 1997. No. 11. P. 149-163.
64. Fodor J. *The Mind–Body Problem // Scientific American*. 1981. No. 244. P. 114-125.
65. Fodor J. *Special Sciences (or: The Disunity of Science as a Working Hypothesis) // Synthese*. 1974. Vol. 28. No. 2. P. 97-115.
66. Foot P. *Natural Goodness*. Oxford: Clarendon Press, 2001.
67. Frankfurt H. *Alternate Possibilities and Moral Responsibility // The Journal of Philosophy*. 1969. Vol, 66. No. 23. P. 829–839.

68. Frankfurt H. *The Importance of What We Care About: Philosophical Essays*. Cambridge: Cambridge University Press, 1988.
69. Franklin C. *Event-Causal Libertarianism, Functional Reduction and the Disappearing Agent Argument // Philosophical Studies*. 2014, Vol. 170. Issue 3. P. 413–432.
70. Ginet C. *On Action*. Cambridge: Cambridge University Press, 1990.
71. Goodman M. *A Sufficient Condition for Personhood // The Personalist Forum*. 1992. Vol. 8 (Supplement). No. 1. P. 75-81.
72. Haji I. *Control Conundrums: Modest Libertarianism, Responsibility and Explanation // Pacific Philosophical Quarterly*. 2001. Vol. 82. No. 2. P. 178-200.
73. Hardcastle V. G. *On the Matter of Minds and Mental Causation // Philosophy and Phenomenological Research*. 1998. No. 58. P. 1–25.
74. Hare R. *Inaugural Address: Supervenience // Proceedings of the Aristotelian Society Supplementary*. 1984. Vol. 58. P. 1-16.
75. Hare R. *Without Conscience: The Disturbing World of the Psychopaths Among Us*. N.Y.: Guilford Press, 1999.
76. Hitchcock C. *Three Concepts of Causation // Philosophy Compass*. 2007. No. 2/3. P. 508-516.
77. Hume D. *An Enquiry Concerning Human Understanding*. Oxford: Clarendon Press, 2000.
78. Jackson F. *Essentialism, Mental Properties and Causation // Proceedings of the Aristotelian Society*. 1995. Vol 69. P. 253-268.
79. Jackson F. *Mental Causation // Mind*. 1996. No.105. P. 377–413.
80. Jackson F. and Pettit P. *Functionalism and Broad Content // Mind*. 1988. No. 97. P. 381–400.
81. Jackson F. and Pettit P. *Program Explanation: A General Perspective // Analysis*. 1990. No. 50. P. 107-117.
82. James R., Blair R. *Neurobiological Basis of Psychopathy // The British Journal of Psychiatry*. 2003. No. 182. P. 5-7.

83. Kane R. *Free Will: New Directions for an Ancient Problem* // Kane R. (ed.). *Free Will*. Malden, MA: Blackwell 2002. P. 222-248.
84. Kane R. *Responsibility, Luck and Chance: Reflections on Free Will and Indeterminism* // Russell P. and Deery O. *The Philosophy of Free Will: Essential Readings From the Contemporary Debates*. Oxford: Oxford University Press, 2013. P. 193-214.
85. Kane R. *Rethinking Free Will: New Perspectives on an Ancient Problem* // Kane R. (ed.). *The Oxford Handbook of Free Will: Second Edition* (2 ed.). Oxford: Oxford University Press, 2011. P. 381-404.
86. Kane R. *The Oxford Handbook of Free Will*. Oxford University Press, 2002.
87. Kane R. *The Significance of Free Will*. N.Y.: Oxford University Press, 1996.
88. Kim J. *Mechanism, Purpose, and Explanatory Exclusion* // *Philosophical Perspectives*. 1989. No. 3. P. 77–108.
89. Kim J. *Can Supervenience and ‘Non-Strict Laws’ Save Anomalous Monism?* // Heil J. and Mele A. (eds.). *Mental Causation*. Oxford: Clarendon Press, 1993. P. 19–26.
90. Kim J. *Causation, Nomic Subsumption, and the Concept of Event* // *Journal of Philosophy*. 1973. No. 70. P. 217-236.
91. Kim J. *Epiphenomenal and Supervenient Causation* // *Midwest Studies in Philosophy*. 1984. Vol. 9. No. 1. P. 257-270.
92. Kim J. *Essays in the Metaphysics of Mind*. Oxford: Oxford University Press, 2010.
93. Kim J. *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. Cambridge MA: Bradford Book, 2000.
94. Kim J. *Multiple Realization and the Metaphysics of Reduction* // *Philosophy and Phenomenological Research*. 1992. Vol. 52. No. 1. P. 1-26.
95. Kim J. *Philosophy of Mind*. Boulder: Westview press, 2010.
96. Kim J. *Physicalism, or Something Near Enough*. Princeton: Princeton University Press. 2007.



97. Kim J. Psychophysical Supervenience // *Philosophical Studies*. 1982. No. 41. P. 51-70.
98. Kim J. The Non-Reductivist's Troubles with Mental Causation // Heil J. and Mele A. (eds.). *Mental Causation*. Oxford: Clarendon Press, 1993. P. 189–210.
99. Kim J. *Supervenience and Mind: Selected Philosophical Essays*. Cambridge: Cambridge University Press, 1993.
100. Kind A. *Persons and Personal Identity*. Malden MA: Polity Press. 2015.
101. King P. Why Isn't the Mind-Body Problem Medieval? // Lagerlund H. (ed.). *Forming the Mind: Essays on the Internal Senses and the Mind-Body Problem from Avicenna to the Medical Enlightenment*. Dordrecht: Springer, 2007. P. 187–205.
102. LePore E. and Loewer B. Mind Matters // *Journal of Philosophy*. 1987. No. 84. P. 630-642.
103. LePore E. and Loewer B. More on Making Mind Matter // *Philosophical Topics*. 1989. No. 18. P. 175-191.
104. Levine J. *Purple Haze: The Puzzle of Consciousness*. New York: Oxford University Press, 2001.
105. Levy N. The Good, the Bad, and the Blameworthy // *Journal of Ethics and Social Philosophy*. 2005. Vol. 1, No. 2. P. 2–16.
106. Levy N. The Responsibility of the Psychopath Revisited // *Philosophy, Psychiatry, and Psychology*. 2007. Vol. 14. P. 129-138.
107. Lewis D. Causation as Influence // *The Journal of Philosophy*. 2000. Vol. 97, No. 4. P. 182-197.
108. Lewis D. Extrinsic Properties // *Philosophical Studies*. 1983. Vol. 44, No. 2. P. 197–200.
109. Libet B. Consciousness, Free Action and the Brain // *Journal of Consciousness Studies*. 2001. No. 8. P. 59–65.

110. Libet B. Unconscious Cerebral Initiative and the Role of Conscious Will in Voluntary Action // Behavioral and Brain Sciences. 1985. No. 8. P. 529-566.
111. Lowe E. Causal Closure Principles and Emergentism // Philosophy. 2000. No. 75. P. 571-585.
112. Lowe E. Physical Causal Closure and the Invisibility of Mental Causation // S. Walter and H. Heckmann, eds., Physicalism and Mental Causation. Exeter, UK: Imprint Academic, 2003. P. 137-154.
113. Lowe E. The Problem of Psychophysical Causation // Australasian Journal of Philosophy. 1992. No. 70. P. 263-276.
114. Lowe E. Subjects of Experience. Cambridge: Cambridge University Press, 1996.
115. Macintyre A. After Virtue: A Study in Moral Theory. Notre Dame, Indiana: University of Notre Dame Press, 2007.
116. Malcolm N. The Conceivability of Mechanism // Philosophical Review. 1968. No. 77. P. 45-72.
117. McKenna M. The Limits of Evil and the Role of Moral Address: A Defense of Strawsonian Compatibilism // Journal of Ethics. 1998. No. 2. P. 123–142.
118. McKenna M. A Hard-Line Reply to Pereboom’s Four-Case Manipulation Argument // Philosophy and Phenomenological Research. 2008. Vol.77. No. 1. P. 142-159.
119. McKenna M. and Russell P. Free Will and Reactive Attitudes: Perspectives on P.F. Strawson's «Freedom and Resentment». Burlington, VT: Ashgate Publishing, 2008.
120. McKenna M. Conversation and Responsibility. N.Y.: Oxford University Press, 2012.
121. McKenna M., Pereboom D. Free will: A Contemporary Introduction. N.Y.: Routledge Taylor and Francis Group, 2016.

122. Mele A. A Critique of Pereboom's 'Four-Case Argument' for Incompatibilism // *Analysis*. 2005. Vol. 65. No. 285. P. 75-80.
123. Mele A. *Dialog on Free Will and Science*. Oxford University Press, 2013.
124. Mele A. On Pereboom's Disappearing Agent Argument // *Criminal Law and philosophy*. 2015.P. 1-14.
125. Mele A. Causation, Action, and Free Will // H. Beebe, C. Hitchcock, and P. Menzies (eds.). *The Oxford Handbook of Causation*. Oxford, New York: Oxford University Press, 2009. P. 554-574.
126. Mele A. *Free Will and Luck*. New York: Oxford University Press, 2008.
127. Murphy J. Moral Death: A Kantian Essay on Psychopathy // *Ethics*. 1972. Vol. 82. No. 4. P. 284-298.
128. Nelkin D. Do We Have a Coherent Set of Intuitions about Moral Responsibility? // *Midwest Studies in Philosophy*. 2007. No. 31. P. 243–259.
129. Nelkin D. *Making Sense of Freedom and Responsibility*. New York: Oxford University Press, 2011.
130. Nozick R. *Philosophical Explanations*. Cambridge, MA: Harvard University Press, 1983.
131. Olson E. *Human Animal: Personal Identity Without Psychology*. Oxford: Oxford University Press, 1999.
132. Olson E. Personal Identity // *The Stanford Encyclopedia of Philosophy* (Spring 2016 Edition) / Edward N. Zalta (ed.). 2016. URL: <http://plato.stanford.edu/archives/spr2016/entries/identity-personal/> (09.07.2015).
133. *Oxford Handbook on Free Will*. Ed. Kane R. Oxford University Press, 2005.
134. *Oxford Handbook of the Self*. Ed. Gallagher S. Oxford: Oxford University Press, 2011.

135. Parfit D. *Reasons and Persons*. Oxford: Oxford University Press. 1984.
136. Pereboom D. *A Defense of Free Will Skepticism: Replies to Commentaries by Victor Tadros, Saul Smilansky, Michael McKenna, and Alfred L. Mele on Free Will, Agency, and Meaning in Life* // *Forthcoming in Criminal Law and Philosophy*. 2017. P. 1-20.
137. Pereboom D. *A Hard-Line Reply to the Multiple-Case Manipulation Argument* // *Philosophy and Phenomenological Research*. 2008. Vol. 77. No. 1. P. 160-170.
138. Pereboom D. *Free Will* // Crisp R. (ed.). *The Oxford Handbook of the History of Ethics*. Oxford: Oxford University Press, 2013. P. 606-637.
139. Pereboom D. *Free Will, Agency, and Meaning in Life*. N.Y.: Oxford University Press, 2016.
140. Pereboom D. *Free Will*. Indianapolis: Hackett Publishing Co, 2009.
141. Pereboom, D. *Living Without Free Will*. Cambridge: Cambridge University Press, 2001.
142. Pereboom, D. *Living Without Free Will: The Case for Hard Incompatibilism* // Kane R. (ed.). *The Oxford Handbook of Free Will*. NY: Oxford University Press, 2002. P. 477-488.
143. Pereboom, D. *Optimistic Skepticism about Free Will*. // Russell P. and Deery O. (eds.). *The philosophy of Free Will*. NY: Oxford Univeristy Press, 2013 P. 421-449.
144. Pereboom D. *Reasons Responsiveness, Alternative Possibilities, and Manipulation Arguments Against Compatibilism; Reflections on John Martin Fischer's My Way* // *Philosophical Books*. 2006. Vol. 47. No. 3. P. 198-212.
145. Pereboom D. *Responses to John Martin Fischer and Dana Nelkin* // *Science, Religion, and Culture*. Vol. 1. No. 3. 2014. P. 218-225.
146. Perry J. *A Dialogue on Personal Identity and Immortality*. Indianapolis: Hackett Publishing, 1978.

147. Perry J. *Identity, Personal Identity, and the Self*. Indianapolis, Cambridge: Hackett Publishing Company, 2002.
148. Pinker S. *The Blank Slate: The Modern Denial of Human Nature*. London: Penguin Books, 2002.
149. Psillos S. Regularity Theories // Beebe H., Hitchcock C., and Menzies P. (eds.) *The Oxford Handbook of Causation*. Oxford, New York: Oxford University Press, 2009. P. 131-157.
150. Reid M. Narrative and Fission: A Review of Marya Schechtman's *The Constitution of Selves* // *Philosophical Psychology*. 1997. Vol. 10. No. 2. P. 211-219.
151. Roberts J. Aristotle on Responsibility for Action and Character // *Ancient Philosophy*. 1989. Vol. 9. No. 1. P. 23–36.
152. Ross D., Spurrett D. What to Say to a Skeptical Metaphysician: A Defense Manual for Cognitive and Behavioral Scientists // *Behavioral and Brain Sciences*. 2004. Vol. 27. No. 5. 2004. P. 603-627.
153. Ross D. Rainforest Realism: A Dennettian Theory of Existence // *Dennett's Philosophy: A Comprehensive Assessment*. Cambridge: MIT Press, 2000. P. 147-166.
154. Rust J. John Searle. London, N.Y.: Continuum, Bloomsberry Academic, 2009.
155. Sapolsky R. The Frontal Cortex and the Criminal Justice System // *Philosophical Transactions of the Royal Society of London B. Biological Science*. 2004. No. 359. P. 1787-1796.
156. Schechtman M. Personality and Persistence: The Many Faces of Personal Survival // *American Philosophical Quarterly*. 2004. Vol. 41, No.2. P. 87-105.
157. Schechtman M. *Staying Alive: Personal Identity, Practical Concerns, and the Unity of a Life*. Oxford: Oxford University Press, 2014.
158. Schechtman M. *The Constitution of Selves*. Ithaca: Cornell University Press, 2007.

159. Schechtman M. *The Narrative Self* // S. Gallagher (ed.) *The Oxford Handbook of the Self*. Oxford: Oxford University Press, 2011. P. 394-416.
160. Searle J. *Freedom and Neurobiology: Reflections on Free Will, Language, and Political Power*. N.Y.: Columbia University Press, 2006.
161. Searle J. *Mind, Language and Society: Philosophy in the Real World*. N.Y.: Basic Books, 1998.
162. Searle J. *Mind: A Brief Introduction*. Oxford: Oxford University Press, 2004.
163. Searle J. *Philosophy in a New Century: Selected Essays*. Cambridge: Cambridge University Press, 2008.
164. Searle J. *Rationality in Action*. Cambridge: Bradford Book, 2003.
165. Searle J. *The Rediscovery of the Mind*. Massachusetts: Bradford Book, 1992.
166. Searle J. *Biological Naturalism* // Velmans M. and Schneider S. (eds.). *The Blackwell Companion to Consciousness*. Malden, MA: Blackwell. 2007. P. 325-334.
167. Searle J. *Intentionality: An Essay in the Philosophy of Mind*. Cambridge: Cambridge University Press, 1983.
168. Searle J. *The Mystery of Consciousness*. N.Y.: New York Review of Books, 1997. Shoemaker D. *Attributability, Answerability, and Accountability: Toward a Wider Theory of Moral Responsibility* // *Ethics*. 2011. Vol. 121. No. 3. P. 602-632.
169. Shoemaker S. *Causal and Metaphysical Necessity* // *Pacific Philosophical Quarterly*. 1998. Vol. 79. No. 1. P. 59–77.
170. Shoemaker S. *Causality and Properties* // P. van Inwagen (ed.), *Time and Cause: Essays Presented to Richard Taylor*, Dordrecht: D. Reidel Publishing, 1980. P. 109–135.
171. Shoemaker S. *Identity, Cause, and Mind, Expanded Edition*. Oxford: Clarendon Press, 2003.

172. Shoemaker D. Moral Address, Moral Responsibility, and the Boundaries of the Moral Community // *Ethics*. 2007. Vol. 118. No. 1. P. 70–108.
173. Shoemaker D. *Oxford Studies in Agency and Responsibility*. Volume 1. New York: Oxford University Press, 2013.
174. Shoemaker D. and Tognazzini N. *Oxford Studies in Agency and Responsibility*. Volume 2. New York: Oxford University Press, 2014.
175. Shoemaker S. Realization and Mental Causation // Gillett C. and Loewer B. (eds.), *Physicalism and its Discontents*. Cambridge/New York: Cambridge University Press. 2001. P. 74–98.
176. Shoemaker D. Responsibility and Disability // *Metaphilosophy*. 2009. Vol. 40. No. 3-4. P. 438-461.
177. Shoemaker S., Swinburne R. *Personal Identity*. Oxford: Wiley-Blackwell, 1991.
178. Smith A. Attributability, Answerability, and Accountability: in Defense of a Unified Account // *Ethics*. 2012. Vol. 122. No. 3. P. 575-589.
179. Sosa E. Davidson's Thinking Causes // Heil J. and Mele A. (eds.). *Mental causation*. Oxford: Clarendon Press, 1993. P. 41–50.
180. Sosa E. Mind-Body Interaction and Supervenient Causation // *Midwest Studies in Philosophy*. 1984. No. 9. P. 271-281.
181. Stoecker R. Davidson // *A Companion to the Philosophy of Action Theory*. O'Connor T. and Sandis C. (eds.). Oxford: Wiley-Blackwell, 2010. P. 598-605
182. Strawson G. The Impossibility of Moral Responsibility // *Philosophical Studies*. 1994. Vol. 75. No. 1-2. P. 5–24.
183. Strawson G. *Freedom and Belief*. New York: Oxford University Press, 2010.
184. Strawson G. (ed.) *The Self?* Malden: Blackwell publishing, 2005.
185. Strawson P. F. Freedom and Resentment // *Proceedings of the British Academy*. 1962. No. 48. P. 1-25.

186. Strawson P.F. *Freedom and Resentment and Other Essays*. London: Routledge, 2008.
187. Strawson P.F. Reply to Ayer and Bennett // Zak van Straaten (ed.). *Philosophical Subjects: Essays Presented to P.F. Strawson*. Oxford: Clarendon Press, 1980. P. 260-266.
188. Swinburne R. *Mind, Brain and Free Will*. Oxford: Oxford University Press, 2013.
189. Talbert M. Blame and Responsiveness to Moral Reasons: Are Psychopaths Blameworthy? // *Pacific Philosophical Quarterly*. 2008. No. 89. P. 516-535.
190. Talbert M. Accountability, Aliens, and Psychopaths: A Reply to Shoemaker // *Ethics*. 2012. Vol. 122. No. 3. P. 562-574.
191. Talbert M. *Moral Responsibility: An Introduction (Key Concepts in Philosophy)*. Cambridge: Polity, 2016.
192. Talbert M. The Significance of Psychopathic Wrongdoing // Schramme T. (ed.). *Being Amoral: Psychopathy and Moral Incapacity*. Cambridge: MIT Press, 2014. P. 275-300.
193. Talbert M. Unwitting Wrongdoers and the Role of Moral Disagreement in Blame // Shoemaker D. (ed.). *Oxford Studies in Agency and Responsibility. Volume 1*. Oxford: Oxford University Press, 2013. P. 225-245.
194. Taylor C. *Explanation of Behaviour*. N.Y.: The Humanities Press, 1964.
195. *The Oxford Handbook of Causation*. Beebee H., Hitchcock C., and Menzies P. (eds.). Oxford, N.Y.: Oxford University Press, 2009.
196. Timpe K. *Free Will: Sourcehood and Its Alternatives*. London: Bloomsbury Academic, 2008.
197. Van Gulick R. Who's in Charge Here? And Who's Doing All the Work? // Heil J. and Mele A.(eds.). *Mental Causation*. Oxford: Clarendon Press, 1993. P. 233-256.



198. Van Inwagen P. *An Essay on Free Will*. Oxford: Clarendon Press, 1983 (1986).
199. Van Inwagen P. A Modal Argument for Incompatibilism // Russell P. and Deery (eds.). *The Philosophy of Free Will: Essential Readings from the Contemporary Debates*. Oxford: Oxford University Press, 2013. P. 117-125.
200. Van Inwagen P. Free Will Remains a Mystery // *Philosophical Perspectives*. 2000. No. 14. P. 1-19.
201. Vasilyev V. «The Hard Problem of Consciousness» and Two Arguments for Interactionism // *Faith and Philosophy*. 2009. Vol. 26. No. 5. P. 514-526.
202. Wallace R. Dispassionate Opprobrium: On Blame and the Reactive Sentiments // Wallace R.J., Kumar R., and Freeman S. (eds.). *Reasons and Recognition: Essays on the Philosophy of T.M. Scanlon*. N.Y.: Oxford University Press, 2011. P. 348-372.
203. Wallace R. *Responsibility and the Moral Sentiments*. Cambridge, MA: Harvard University Press, 1998.
204. Watson G. Responsibility and the Limits of Evil: Variations on a Strawsonian Theme // Watson G. *Agency and Answerability: Selected Essays*. Oxford: Oxford University Press. 2004. P. 219-259.
205. Watson G. Two Faces of Responsibility // *Philosophical Topics*. 1996. Vol. 24. No. 2. P. 227–248.
206. Watson G. Free Action and Free Will // *Mind*. 1987. Vol. 96, No. 382. P. 145-172.
207. Watson G. The Trouble with Psychopaths // Wallace R.J., Kumar R., and Freeman S. (eds.). *Reasons and Recognition: Essays on the Philosophy of T.M. Scanlon*. N.Y.: Oxford University Press, 2011. P. 307-331.
208. Wegner D. *Précis of The Illusion of Conscious Will* // *Behavioral and Brain Sciences*. 2004. No. 27. P. 649-692.
209. Wegner D. *The Illusion of Conscious Will*. Cambridge, MA: MIT Press, 2003.

210. Williams B. *Shame and Necessity*. Los Angeles: University of California Press, 2008.
211. Wolf S. The Importance of Free Will // *Mind*. 1981. No. 90. P. 366-378.
212. Wolf S. *Freedom within Reason*. N.Y.: Oxford University Press, 1993.
213. Yablo S. Mental Causation // *The Philosophical Review*. 1992. Vol. 101. No. 2. P. 245-280.
214. Zimmerman M. Responsibility, Reaction, and Value // *Journal of Ethics*. 2010. Vol. 14. No. 2. P. 103–115.
215. Zimmerman M. *An Essay on Moral Responsibility*. Totowa, NJ: Rowman and Littlefield, 1988.