

«Детектив — это расследование тайны, замешанной на преступлении, насилии, крови. Как насчет аналитического детектива, завязанного на самых интригующих тайнах нашей души?» — профессор К. Анохин

# Свобода ВОЛИ

*Иллюзия или возможность*



Дмитрий Волков



Центр исследования  
сознания

ПРИ ФИЛОСОФСКОМ ФАКУЛЬТЕТЕ МГУ

0+

**Дмитрий Волков**  
**Свобода воли. Иллюзия**  
**ИЛИ ВОЗМОЖНОСТЬ**

*[http://www.litres.ru/pages/biblio\\_book/?art=56036454](http://www.litres.ru/pages/biblio_book/?art=56036454)  
SelfPub; 2020*

**Аннотация**

Многие открытия в нейронауке и социальной психологии ставят под сомнение наличие свободы воли. Они показывают, что поступки людей определяются социальным контекстом и бессознательными состояниями мозга. Но не следует ли из этого, что человек не может принимать решения сам? А если так, если свобода воли – иллюзия, возможно ли считать других людей ответственными за поведение? В книге автор защищает современную компатибилистскую теорию – позицию, согласно которой свобода и ответственность совместимы с современными научными представлениями об устройстве мира. Кроме того, в ней предложены решения и других философских головоломок, связанных со свободой: проблемы тождества личности и ментальной каузальности. Эта книга – результат многолетних исследований автора. Дмитрий Борисович Волков – доктор философских наук, содиректор Московского центра исследования сознания при философском факультете МГУ,

автор книги «Бостонский зомби: Д. Деннет и его теория сознания» (2012).

# Содержание

Предисловие	8
Введение	13
Глава 1. Проблема ментальной каузальности и возможность действий	42
I. Аргументы в пользу ментальной каузальности	42
1. Связь проблем ментальной каузальности и свободы воли	42
2. Корреляция ментальных состояний и действий	44
3. Два типа возможных связей между ментальными состояниями и действиями	49
4. Доводы против каузальной связи действий и ментальных состояний	52
А. Трудности с доказательством психофизических законов	52
В. Аргумент аномальности ментального	55
С. Аргумент логической связи	56
5. Доводы Д. Дэвидсона в пользу каузальной связи	58
II. Проблема ментальной каузальности в современной аналитической философии	67
1. Аргумент исключения ДЖ. Кима против	67

ментальной каузальности	
2. Аргумент супервентности против каузального влияния одних ментальных состояний на другие	73
III. Решение проблемы ментальной каузальности в локальном интеракционизме	78
1. Ключевые положения локального интеракционизма	78
2. Аргумент каузальных траекторий в пользу ментальной каузальности и против локальной супервентности	80
А. Локальная или глобальная супервентность?	86
В. Разъяснение двусмысленности в аргументе	91
3. Возражения на аргумент каузальных траекторий	94
А. Аргумент-двойник	94
В. Память не является качественным состоянием	100
С. Решение парадокса сближающихся в ментальных состояниях двойников	103
IV. Решение проблемы ментальной каузальности в биологическом натурализме Дж. Сёрла	108
1. Ключевые положения биологического	108

натурализма Дж. Сёрла	
2. Иллюстрации к каузальным связям в биологическом натурализме	114
3. Внутренние противоречия теории Дж. Сёрла	120
4. Недостатки проведенных аналогий и иллюстраций	122
V. Решение проблемы ментальной каузальности в телеофункционализме Д. Деннета	125
1. Ключевые положения телеофункционализма Деннета	125
2. «История двух черных ящиков» как иллюстрация каузальной эффективности высокоуровневых свойств	134
3. Интерпретация мысленного эксперимента	139
4. Критика телеофункционализма и возможные ответы	144
А. Невыполнимость условий мысленного эксперимента	144
Конец ознакомительного фрагмента.	147

*Моей дочке Анастасии Волковой*

# Предисловие

«Папа, у меня появилась идея. Я хочу сделать в школе доклад по твоей философии. Ты можешь мне рассказать про сознание? Что это такое и почему тебя это интересует?» Эта неожиданная просьба застает меня врасплох. Ей всего десять лет, и мне трудно объяснить ей суть философских проблем, которыми я занимаюсь. Мне потребовалось десять лет только для того, чтобы познакомиться с темой, увидеть и сформулировать для себя основные вопросы, разобраться в имеющихся стратегиях ответов на них, ключевых проблемах и аргументах, а теперь мне нужно буквально в нескольких предложениях, да еще и без использования специфической терминологии, самыми простыми словами, изложить все содержание моей работы. И тем не менее я пытаюсь. Я пытаюсь рассказать ей о том, чем я занимаюсь, почему это важно для меня и почему это важно вообще. И эта книга, хотя она, конечно, совсем не похожа на школьный доклад, – это благодарность за ее живой интерес. Я начал писать эту книгу очень давно и для себя, а заканчиваю теперь – для Насти, надеясь, что когда-нибудь она ее прочитает.

Кроме дочки в моей жизни есть и другие важные люди, которые воодушевляли меня во время создания этой книги, без которых она никогда бы не появилась. Это – прежде всего мой наставник, друг и главный собеседник в философии,

Вадим Васильев. Именно он привил мне вкус к философским аргументам, мудро направлял мои исследования, всегда очень деликатно корректировал мои недочеты и на собственном примере, в своих статьях и книгах демонстрировал образец честной и глубокой философской работы. Вместе с Васильевым мы почти десять лет назад создали Московский центр исследования сознания (МЦИС), который сегодня объединил вокруг себя многих специалистов по аналитической философии сознания. Начинался же он с того, что к нам присоединились Роберт Хауэлл, Антон Кузнецов и Артем Беседин.

С их появлением в Центре мы стали настоящим философским сообществом – пусть небольшим, но очень активным. Работать стало намного интересней – появилась возможность дискутировать и получать быстрый живой, эмоциональный отклик. Коллеги – Васильев, Кузнецов, Беседин – слушали мои выступления, читали черновики работ и критиковали предлагаемые мной решения. Благодаря им я уточнил определения, избавился от многих затруднений, лучше определился с собственной позицией и отточил аргументы. Этому способствовали также диалоги с нашими коллегами из Санкт-Петербурга: Данилом Разеевым, Сергеем Левиным, Марией Секацкой, Дарьей Чирвой. Мы познакомились на одной из совместных конференций в Москве, а потом они пригласили нас к себе в Петербург. С этого момента завязалась дружба, и наше профессиональное сообщество расши-

рило свои границы. Впрочем, в связи с работой Центра у меня появилось много собеседников и за рубежом.

Особо продуктивными были международные Летние школы по проблемам свободы воли и моральной ответственности, организованные Центром. Во время этих двухнедельных выездных семинаров мы собирали группы молодых ученых из разных стран, специализирующихся на этой проблематике, и под руководством именитых профессоров и авторитетных специалистов обсуждали основные проблемы и аргументы, полемизировали и делились идеями. На этих Летних школах и различных конференциях я познакомился со многими западными аналитическими философами. Это прежде всего Дерк Перебум, Джон Фишер, Ричард Суинбёрн, Пол и Патрисия Чёрчленд, Джесси Принц и Дэвид Чалмерс. Одно дело, когда читаешь философские работы автора, о котором знаешь только из энциклопедических статей. Совсем другое дело – когда эти авторы становятся твоими собеседниками. Общение с этими философами значительно повлияло на стиль моего мышления. Особую роль в этом сыграл Дэниел Деннет. С ним меня связывает долгая дружба. Для меня Деннет стал образцом современного философа: ходячая энциклопедия (причем по многим отраслям знаний от музыки до биологии), необыкновенно острый математический ум и прекрасный оратор и публицист. Влияние его концепций хорошо прослеживается в тексте всей книги.

В ходе работы над текстом мне удалось защитить доктор-

скую диссертацию на философском факультете МГУ по проблеме свободы воли. На защите возникла острая дискуссия. Не всем членам совета понравился аналитический подход к проблеме свободы воли. Впрочем, как ни странно, во многом успеху способствовали выступления оппонентов Данила Разеева, Светланы Нагумановой и Дмитрия Иванова. Несмотря на то что они не были согласны со всеми выводами работы, они доказали, что аналитическая философия в целом и ее подход к современным метафизическим проблемам заслуживает самого большого внимания. Именно современная аналитическая философия дает лучший образец философского критического мышления и продолжает традиции классического философствования. Я очень благодарен им за оппонирование и критику. Некоторые их замечания я учел при работе над этой монографией.

Кроме того, я благодарен философскому факультету МГУ. Здесь я получал образование, здесь я впервые познакомился с современными философскими вопросами, здесь работают мои учителя и коллеги Дмитрий Бугай, Егор Фалев, Галина Стрельцова, Владислав Коровин и Владимир Миронов. И здесь я воплотил мечту своей юности – стал профессиональным философом. Спасибо за помощь, критику и замечательные дискуссии. И наконец, огромное спасибо моему редактору Андрею Мерцалову, который работал над этим текстом с огромной кропотливостью и усердием иногда в выходные дни и даже по ночам. Благодаря ему мы испра-

вили множество недочетов, и текст читать стало легче. Он был моим верным соратником в этой трудной, некогда казавшейся бесконечной работе. Теперь она закончена.

# Введение

*Если вол забодает мужчину или женщину до смерти, то вол побить камнями и мяса его не есть; а хозяин вола не виноват...*

*Библия. Исход. 21: 28*

Бартоломе Шасене, французский юрист, практиковавший в XVI в., сделал себе имя в судах, выступая представителем довольно необычных клиентов. Он защищал интересы животных. Наиболее известна его защита крыс, обвинявшихся церковным судом в уничтожении урожая ячменя в епархии Отена. На основании поступившей жалобы было открыто дело, обвиняемые вызваны в суд к определенной дате, а Шасене, талантливый и знающий юрист, назначен их защитником. Причастность крыс была очевидна ввиду множества свидетельских показаний, да и репутация грызунов не была безупречной. Поэтому Шасене пришлось прибегнуть к разным юридическим уловкам, чтобы отстаивать права своих клиентов.

В первую очередь он обратил внимание судьи на то, что подзащитные расселены на значительной территории страны и не были в равной мере проинформированы о вызове в суд. Эту техническую тонкость приняли во внимание, и были сделаны повторные вызовы крыс в суд, на этот раз с кафедр всех церковных приходов епархии. По истечении отпу-

ценного времени Шасене, однако, обратился с новой жалобой. Он заметил, что явка в суд для крыс была сопряжена с риском для жизни, так как на каждом шагу их подстерегали злейшие враги – кошки. Адвокат утверждал, что любой имеет право воздержаться от появления в суде, если это угрожает его безопасности. Это обстоятельство также приняли во внимание, благодаря чему адвокату удалось избежать дефолта из-за неявки клиентов. Впрочем, окончательный вердикт этого трибунала остается неизвестным – он не отражен в доступных источниках. Известно, однако, что именно этот процесс принес Шасене славу и способствовал его карьере, что в итоге позволило ему занять пост первого председателя высшего суда Прованса. Известно и то, что проблема защиты животных в судах интересовала Шасене и дальше. Он даже написал на эту тему объемную работу. Основным ее выводом служило то, что животные могут представлять перед судом, однако в процессе этих судов должны выполняться все юридические нормы и правила. До тех пор пока эти процедуры не соблюдены, к обвиняемым не могут применяться никакие наказания. Мнение Шасене о наличии вины и подсудности животных разделяли и некоторые другие средневековые юристы.

В исторической хронике, созданной швейцарским летописцем Шиллингом, сохранился отчет о суде, состоявшемся в 1479 г. под председательством епископа Лозанны. Дело было открыто в связи с жалобой местного приходского свя-

щенника на неких насекомых, называемых *inger* (по-видимому, зерновки). Согласно свидетельству священника, насекомые уничтожили значительную часть урожая. Епископ не остался равнодушным к описанию преступлений, совершенных вредителями, и издал указ, в котором наказывал им в шестидневный срок покинуть поля, сады и деревья. В случае неповиновения насекомые должны были предстать перед судом города Вифлисабурга (совр. Аванш).

Из хроники следует, что насекомые не испугались указа. На судебных заседаниях интересы насекомых представлял адвокат Жан Перродэ из Фрибурга. Он не был таким же изобретательным, как Шасене, поэтому суд завершился достаточно быстро и не в пользу обвиняемых. Судья счел преступления доказанными, а вину очевидной. В качестве наказания насекомые были преданы анафеме, и им было запрещено впредь появляться на полях. Впрочем, анафема не была самым жестоким наказанием, применявшимся к животным. По итогам множества средневековых судебных разбирательств и судов Нового времени животных нередко пытали, отправляли в тюрьмы, вешали, душили, сжигали и закапывали заживо.

Так, за убийство ребенка в 1456 г. в городке Оппенхайм и в 1557 г. в коммуне Сен-Кантен на смертную казнь были осуждены свиньи – их похоронили заживо. В 1314 г. бык, принадлежавший фермеру в селении Муази, на дороге напал на мужчину. Мужчина погиб от ранений, а животное было

поймано и посажено в тюрьму, после чего состоялся суд, на котором быка приговорили к повешению. В 1474 г. в Базеле на смерть за необычное преступление был осужден петух – он якобы снес яйцо. Это преступление было приравнено к колдовству, и птицу сожгли заживо. Американский историк Э. Эванс приводит больше сотни таких случаев за XVI–XVII вв. Основанием для таких действий, вероятно, было убеждение в универсальности закона – так называемого *lex talionis*, закона равного возмездия. Удивительным образом он распространялся даже на неодушевленные предметы.

В Древней Греции, например, существовала специальная категория разбирательств – суды над вещами. Так, согласно Драконовым законам в Древней Греции оружие и любые предметы, задействованные в убийстве, должны были быть публично осуждены и выброшены за пределы Афин. Из Античности известен случай осуждения статуи атлета Никона с Тасоса. Подвергшись нападению завистников спортсмена, статуя была свергнута с пьедестала, но, падая, обрушилась на одного из нападавших. В результате статуя была представлена перед трибуналом и осуждена на утопление в море.

Любопытна в этом отношении и знаменитая расправа над угличским колоколом. С помощью этого колокола в 1591 г. жители города были оповещены о гибели царевича Дмитрия. Угличане расправились с предполагаемыми убийцами наследника престола, а царь Борис Годунов решил наказать их за самосуд. Наказали не только людей, но и колокол: его

сбросили со Спасской колокольни, вырвали ему язык, отрубили ухо, принародно на площади он был избит плетьюми, а после – отправлен в сибирскую ссылку. В Тобольске воевода, князь Лобанов-Ростовский, велел запереть колокол в приказной избе и сделать на нем надпись «первосыльный неодушевленный с Углича». Об этом свидетельствуют «Сибирские летописи» и «Статейный список сибирских воевод».

Однако наиболее жутким примером работы системы правосудия был «трупный синод», состоявшийся в январе 897 г. в Латеране. Этот трибунал был проведен над эксгумированным трупом папы Формоза, который умер в апреле 896 г. В Латеранской базилике Св. Иоанна новый папа Стефан велел посадить на трон полуразложившийся труп своего предшественника и в присутствии высшего духовенства подверг его посмертному допросу. Формозу вменялось вероломство, переход с одной епископской кафедры на другую вопреки установленному Никейским собором запрету, а также совершение им, мирянином, религиозных таинств. Итогом разбирательства стало признание указов бывшего папы недействительными, и... наказание тела. У трупа были отрублены три пальца, которыми Формоз совершал крестное знамение. После чего его проволокли по улицам Рима и с прикрепленным грузом сбросили в Тибр.

В некотором смысле подобные суды носили символический характер. Их целью была в том числе демонстрация

универсальности закона для всех и всего, и в этом смысле их можно обосновать. Но, судя по всему, одной из мотиваций таких судов было возмездие. И в этом смысле подобные процессы абсурдны. Возмездие, как и вознаграждение, можно применять только к объектам, которые могут нести ответственность. Только такие объекты способны совершать преступления и быть виновными, поступать хорошо и быть достойными благодарности. Этим требованиям иногда соответствуют люди. Но животные и неодушевленные предметы, скорее всего, для этого не подходят. Объяснить причину такого несоответствия можно, выявив условия ответственности.

Понятие ответственности родственно понятию причины. В связи с этим иногда говорят, что «лаваина несет ответственность за гибель альпинистов», или что «ураган ответственен за многочисленные разрушения». Но это – переносное значение слова. Ответственность в прямом смысле, с точки зрения современной юридической практики и большинства современных этических теорий, – это нечто большее. Она предполагает выполнение особых условий для агента и поступка. Этими условиями являются зависимость действий от намерений, наличие свободы воли и какой-то особой связи, сохраняющейся во времени между агентом и совершенным действием.

Во-первых, ответственность предполагает, что действие в качестве своего источника имеет намерение. События, не

связанные с намерениями агентов, как и случайные события, не влекут за собой обвинения или благодарность. Конечно, существуют, например, преступления, совершенные неумышленно. Однако в таком случае преступный поступок является следствием другого, намеренного действия, которое привело к преступлению. Так, если пьяный водитель совершает наезд на прохожего, мы считаем водителя виновным, даже если он совершил наезд непреднамеренно. Здесь вина проистекает из того, что водитель намеренно употреблял алкоголь и после этого сел за руль. Таким образом, намеренное действие все равно является необходимым в цепочке ответственного действия.

Вторым условием ответственности является свобода воли, понимаемая в первую очередь как возможность поступать иначе. Если возможности поступать иначе у агента нет, его действие считается вынужденным. Но вынужденные действия обычно не относятся к категории тех, за которые агент может нести ответственность. Так, за действия, совершенные под дулом пистолета или под угрозой расправы над личностью или его родственниками, человек обычно не несет ответственность. В подобных случаях ответственность возлагается на тех, кто вынудил совершить действие, и снимается с выполнившего чужое распоряжение агента. Только за действия, продиктованные свободным волеизъявлением, человека можно считать ответственным.

Наконец, необходимо, чтобы агент, совершивший дей-

ствие, был особым образом связан с действием, за которое ему полагается воздаяние. В качестве этой связи подходит отношение принадлежности: действие должно было принадлежать тому же агенту, кому полагается воздаяние. Связь с другими агентами не подходит для сохранения ответственности. Очевидно, например, что если преступление совершил один близнец, то его брат не несет ответственность. Не может нести ответственность сын за поступки своего отца. Конечно, сын может испытывать угрызения совести или осуждать поступки отца, но это не то же самое, что ответственность. Оснований наказывать или поощрять сына за поступки отца нет.

Таковы три важнейших условия ответственности. Возможно, существуют и другие условия, но я сфокусирую внимание именно на этих трех. Эти три условия не выполняются в случаях, о которых шла речь выше, в ситуациях с неодушевленными предметами и нерациональными существами. Однако все эти условия должны быть выполнимы в принципе и фактически выполняться хотя бы иногда в жизни нормальных личностей. В противном случае мы бы лишились оснований приписывания моральной ответственности, а вместе с ней – и того фундамента, на котором строится большая часть межличностных отношений.

События у людей вызывают различные реакции, положительные или отрицательные. Люди радуются или огорчаются тому, что пошел дождь. Они сожалеют, что дерево упа-

ло на машину, или довольны, что в этом году поспело много яблок. Но есть специфические реакции, которые применяются только к определенному роду событий – ответственным человеческим действиям. Этими специфическими реакциями является негодование, осуждение, благодарность и т. п. На дождь или дерево нерационально злиться и негодовать. Неуместно благодарить или восторгаться поведением яблонь и светофоров. Негодовать, злиться, испытывать благодарность можно только в отношении ответственных действий агентов.

Представьте, что эти реакции исчезли бы совсем: каким был бы мир без них? Отношения между людьми стали бы полностью «объективными». К примеру, с человеком, совершившим дурной поступок, поступили бы так, словно он – упавшее на машину дерево или плохо работающий пылесос. Его бы починили, исправили, привели в пригодное состояние. Но по отношению к нему не было бы гнева и осуждения. Да и сам человек не мог бы испытывать раскаяния. Был бы такой мир лучшим? Мнения об упразднении отрицательных ответных реакций могут расходиться. Кто-то может считать, что, возможно, было бы даже очень неплохо, если бы в отношениях между людьми исчезли гнев, жажда мести и даже чувство вины. Но упразднение положительных ответных реакций – думаю, с этим согласятся все – сделало бы отношения бессодержательными, утилитарными. Можно ли помыслить отношения без чувств благодарности, симпатии, люб-

ви, уважения? Нельзя! Эти чувства принципиально важны и наполняют жизнь смыслом. Но все эти реакции продиктованы представлением об ответственности. И можно попытаться показать, что условия ответственности невыполнимы в принципе.

Существует точка зрения, что признание нормальных здравомыслящих личностей ответственными так же необоснованно, как и приписывание вины животным и неодушевленным предметам. То есть предполагается, что действия и поступки людей принципиально не отличаются от любых других событий в природе. Это мнение поддерживается некоторыми современными научными и философскими теориями. Может показаться, что среди научных теорий опасность для ответственности должны представлять социально-психологические и нейрофизиологические теории. Так, против ответственности могут быть использованы теории, в которых доказывается определяющая роль ситуации или авторитета в поведении людей. Доводом может служить, в частности, знаменитый Стэнфордский тюремный эксперимент, проведенный в 1971 г. психологом Филипом Зимбардо<sup>1</sup>.

---

<sup>1</sup> Эксперимент является одним из классических в современной социальной психологии. Однако он часто критикуется как в этическом плане, так и в плане соблюдения научной методологии. Некоторые критики считают, что Зимбардо фактически доказал не то, что поведение личности определяется ситуацией, а то, что конформность с авторитетом определяют поведение. Однако для целей нашего исследования это не имеет существенного значения. Эксперимент все

Эксперимент начался с объявления в газете, в котором молодым мужчинам предлагалось за небольшую плату две недели прожить в тюремных условиях в роли «охранника» или «заключенного». «Тюрьма» была создана на территории Стэнфордского университета, в подвале здания психологического факультета. Помещение состояло из коридора и нескольких небольших кабинетов. Часть из них была превращена в камеры. Из комнат вынесли мебель, поставили простые деревянные кровати, двери заменили решетками. Чулан в коридоре стал одиночным изолятором. В конце коридора несколько дополнительных комнат отвели для охранников и наблюдателей. Во всех помещениях было смонтировано аудио- и видеонаблюдение, так чтобы все события фиксировались.

Роли «заключенных» и «охранников» были распределены случайным образом. Те, кому выпала роль «охранников», должны были работать в три смены по 8 часов. После работы они возвращались домой. «Охранникам» объяснили, что целью эксперимента является симуляция тюрьмы в рамках практических и этических ограничений. Главной их задачей было поддержание порядка и дисциплины. Для этого, в частности, они должны были регулярно пересчитывать «заключенных». Никто не должен был пропасть. «Заключенным» специальных инструкций не давали. Предполагалось, что все время они проведут в камерах. Может быть, именно из-за от-

сутствия обязанностей большее число участников изначально изъявляли желание быть именно «заключенными». Впрочем, энтузиазм тех, кому выпала эта роль, скоро убавился. Организаторы постарались, чтобы условия эксперимента не сильно отличались от реальности.

Утром первого дня экипаж полицейских произвел задержание «подозреваемых». Это оказалось довольно неожиданным. Будущим «заключенным» были предъявлены обвинения, на них надели наручники и с завязанными глазами на полицейской машине привезли в помещение «тюрьмы». Там «охранники» по протоколу провели прием поступивших: их обыскали, раздели догола, обрызгали из баллончика с дезодорантом, якобы для дезинфекции, одели в новую тюремную форму. Этой формой были робы с номерами, под которыми запрещалось что-либо носить. На щиколотку «заключенным» надели цепочку с замком, а волосы закрыли нейлоновой шапочкой. Выполнение процедур предназначалось для того, чтобы придать еще большего реализма тюремной обстановке, а форма и присвоение номеров способствовали «деперсонализации» «заключенных».

«Деперсонализации» были подвергнуты и «охранники». Им были выданы одинаковые рубашки цвета хаки, деревянные полицейские палки, свистки и черные очки. Эти атрибуты носили символический характер. Палки применять запрещалось, однако все это должно было способствовать вживанию волонтеров в свои роли, а также формированию но-

вой идентичности. Как показал эксперимент, это произошло даже быстрее и радикальней, чем могли предположить организаторы. Пожалуй, наиболее явно это стало проявляться во время «пересчетов».

Эта рутинная процедура формально предназначалась для проверки наличия всех «заключенных» на месте. Первые пересчеты также предназначались для ознакомления заключенных со своими номерами. Но «охранники», ощутив свою власть над «заключенными», превратили процедуру в психологическое истязание. С шутками и издевками они заставляли «заключенных» десятки раз повторять номера, в разных порядках, с разными интонациями, пропевая их, делая различные жесты. «Заключенные» недоумевали. Однако любые попытки сопротивляться этим издевательствам стали вызывать агрессию. «Охранники», освоившись со своей ролью, осознавшие собственную безнаказанность, заставляли десятками раз повторять одно и то же. Тех, кто не выполнял команды или делал это неохотно, раздевали догола, заставляли приседать или отжиматься. Через некоторое время и это стало казаться недостаточным: «охранники» стали ставить на отжимающихся ноги или заставляли других «заключенных» садиться сверху. Пересчеты, которые должны были длиться десять минут, стали растягиваться на часы.

Издевательское поведение «охранников» вызвало протест среди «заключенных». Им было трудно смириться с подобным произволом, ведь в их представлении «охранники» не

отличались от «заключенных»: и те и другие были просто участниками психологического эксперимента, и роли были определены случайным образом. На второй день мятежники забаррикадировали собственные камеры и отказались выходить. Но сопротивление вызвало еще большую агрессию со стороны «охранников». Мятеж был подавлен с крайней жестокостью. «Охранники» вызвали дополнительную смену в подмогу и стали распылять огнетушители через решетки в камеры. Задыхающиеся «заключенные» вынуждены были сдаться, но им не удалось избежать наказания. Наиболее непокорных заключили в изолятор, у остальных забрали кровати и постельное белье. Но даже на бетонном полу им не удавалось выспаться.

«Охранники» стали проводить пересчеты и ночью с новыми и новыми издевательствами. Нервные срывы и беспомощность «заключенных», казалось, вызывали еще больше энтузиазма. «Охранники» стали раньше приходить на работу и придумывали все больше и больше бесполезных и абсурдных действий, которые они вынуждали совершать «заключенных» (например, заправлять и разглаживать постель, перетаскивать с места на место пустые коробки и т. п.). Более того, они запретили бунтарям пользование туалетом – это стало привилегией наиболее послушных «заключенных». Остальным в камеры поставили ведра, так что вскоре запах и гигиеническое состояние «тюрьмы» стало совершенно безобразным. Когда издевательства стали носить порно-

графический характер, руководитель эксперимента вынужден был вмешаться. За пять дней молодые, психически здоровые ребята из благополучных семей, никогда не проявлявшие склонность к насилию, превратились в садистов. Как это могло случиться? Откуда возникла такая изощренность и изобретательность в способах унижения?

По мнению Ф. Зимбардо, основной причиной этого была Ситуация. Сама сущность «тюремного порядка» и распределение ролей оказались достаточным основанием для возникновения аберрантного, антисоциального поведения «охранников». Введенные символические элементы и процедуры, из которых состоял «порядок», спровоцировали агрессию. В этих условиях индивидуальный моральный характер оказался подавленным. В очень короткий срок внешние обстоятельства трансформировали личности даже тех, кто не имел никаких первоначальных склонностей к садизму. Таким образом, эксперимент показал, что тюремный порядок ведет к моральной деградации стражей. Однако, как заметили наблюдатели, моральная деградация распространяется и на заключенных. С течением времени даже они начинают копировать поведение стражников – отношение заключенных друг к другу также зависит от общей динамики. То есть все участники, независимо от их роли, подпадают под контроль ситуации и это ведет к моральной деградации участников. Из эксперимента следует, вероятно, что тюремный порядок не соответствует поставленным обществом целям.

Впрочем, из Стэнфордского тюремного эксперимента, как я уже говорил, можно сделать и более общий вывод. Возможно, из него следует, что внешние обстоятельства определяют действия индивидов и что разница между преступниками и жертвами – только в ситуации, в которой они оказываются. Одним выпадает роль преступников, другим – стражей, и в этом кроется объяснение различий. Подобный вывод можно попробовать обосновывать и результатами другого знаменитого эксперимента. Его провел один из самых влиятельных психологов XX в., Стэнли Милгрэм из Йельского университета. Милгрэм показал, что под воздействием авторитета люди способны совершать аморальные поступки, которые они никогда не совершили бы самостоятельно.

В Эксперименте о подчинении 1963 г. участвовали сорок мужчин в возрасте от двадцати до пятидесяти лет из разных социальных групп. Среди них были инженеры, школьные учителя, продавцы, рабочие и даже ученые. Участникам объяснили, что цель эксперимента в том, чтобы исследовать влияние боли и наказаний на память. Эксперимент начинался с распределения ролей: участники тянули жребий и, как они думали, могли получить роль либо учителя, либо ученика. На самом деле все участники получали роль учителя. Роль ученика всегда выпадала приглашенному актеру.

По формальным условиям эксперимента «ученик» должен был запоминать пары слов из длинного списка. После чего его на глазах «учителя» привязывали к креслу с элект-

тродами «для того, чтобы ограничить нежелательные движения». «Учитель» переходил в другое помещение и садился за стол с контрольной панелью, которая должна была управлять электрическим генератором. Панель состояла из 30 переключателей. На каждом переключателе были подписаны значения вольт от 15 до 450. Переключатели были объединены в несколько групп: «легкий шок», «умеренный шок», «сильный шок», «очень сильный шок», «чрезвычайно сильный шок», «опасность: максимальный шок». Согласно условиям эксперимента, «учителя» должны были проверять знание списка у «ученика» и в случае неправильного ответа наказывать «ученика» электрическим разрядом. За каждый последующий неправильный ответ уровень шока должен был повышаться.

Перед испытанием эксперты сделали прогноз, что не более 3 % участников будут «наказывать» «учеников» электрическим шоком с высокими значениями. Их ожидало ужасающее открытие: почти все участники в конечном итоге использовали шок в 450 вольт. То есть почти все здоровые здравомыслящие люди готовы были подвергнуть «ученика» сильнейшим страданиям в условиях лабораторного эксперимента. Может быть, они не понимали, что должно было происходить, или не верили в реалистичность эксперимента? Может быть, они подверглись принуждению со стороны организаторов?

«Ученик» в действительности не испытывал никакого шо-

ка – оборудование лишь симулировало электрические разряды. Но эксперимент был поставлен так, что участники никаким образом не могли догадаться об этом. Для убедительности «учителям» вначале самим предлагалось протестировать аппарат, и каждый из них получал довольно сильный разряд в 45 вольт. Они также слышали реакцию «ученика»: после шока в 300 вольт «жертва» начинала стучать ногами в стену, требуя прекратить эксперимент. То же повторялось и после разряда в 315 вольт. Перед каждым использованием переключателя участник должен был вслух объявлять значение электрического разряда. То есть участники были уверены, что причиняют боль. Это подтверждалось и их реакцией. Большая часть испытуемых были крайне эмоционально напряжены. Они краснели, кусали губы, тяжело дышали, дрожали, царапали себя ногтями и тем не менее продолжали эксперимент. При этом никакого физического принуждения они не испытывали. У них вообще не было обязательств продолжать «пытку».

В комнате с испытуемым находился ведущий эксперимента. Он объяснял цели исследования и его правила. Когда испытуемый «учитель» останавливался (например, после стука в стену и протестов «ученика»), ведущий говорил, чтобы он продолжал. Когда от «ученика» вообще переставали поступать сигналы, ведущий указывал на то, что отсутствие ответа должно приравниваться к неправильному ответу и что за это должно следовать «наказание». Среди стандартных заго-

товленных фраз ведущего на возражения со стороны «учителей» были: «пожалуйста, продолжайте», «эксперимент требует, чтобы вы продолжили», «абсолютно важно, чтобы вы продолжали», «у вас нет иного выбора, кроме как продолжать». Но никаких других способов воздействия на испытуемого у ведущего не было. Участник в любой момент мог просто прекратить пытку. Это теоретически мог сделать каждый испытуемый, при этом большинство из них продолжали до самого конца, вопреки убеждению, что причинение боли другому является злом. Почему?

По мнению Милгрэма, в данном случае определяющими факторами были авторитет и склонность подчиняться. Участники эксперимента понимали, что он проводится таким авторитетным учреждением, как Йельский университет, они предполагали, что его сотрудники компетентны и ответственны за условия, и поэтому, вопреки внутренним моральным соображениям, следовали всем указаниям ведущего. Удивительным было лишь то, что этот авторитет продолжал играть решающую роль даже тогда, когда действия начинали носить абсолютно аморальный характер. Они нарушали прививаемую всем с детства максиму о непричинении боли и страдания другому. И при этом участники действовали в отсутствие внешнего принуждения. Несмотря на то что неподчинение ведущему эксперименту не грозило никакими последствиями испытуемым, они почти все избрали путь конформизма. Они выбрали подчинение даже несмотря на

то, что испытывали сильнейшие угрызения совести и эмоциональный стресс. Таким образом, описанный выше эксперимент демонстрирует преобладание силы конформизма и привычки к подчинению над моральными ценностями личности.

В этом отношении выводы Милгрэма созвучны результатам, полученным Зимбардо: внешние обстоятельства играют большую роль, чем мы ожидаем. Они во многом определяют поведение личностей. Таким образом, описанные выше эксперименты могут заставлять усомниться в выполнимости условий ответственности. Впрочем, наука выявляет и другие, *внутренние* обстоятельства, которые играют важнейшую роль в определении поведения – это неосознаваемые процессы, протекающие в мозге. Нейрофизиология раскрывает механизмы, определяющие поведение человека. Одной из лучших иллюстраций в этой области является эксперимент Бенджамина Либега 1983 г.

Ученый предложил испытуемым двигать пальцем и/или запястьем правой руки тогда, когда им этого хотелось. Движение не должно было зависеть от каких-либо внешних факторов, но совершаться в произвольный момент. Участников предупредили, что они не должны планировать движение заранее. При этом нужно было точно фиксировать момент принятия решения о движении руки. Для этого использовались специальные точные часы: световая точка на циферблате осциллоскопа пробегала полный круг каждые примерно

2,5 секунды, поэтому с помощью положения точки на циферблате можно было определить время с точностью до доли секунды. С помощью этих часов испытуемые фиксировали момент возникновения желания пошевелить рукой. Но параллельно с этим Либет записывал еще один параметр – потенциал готовности.

К голове над двигательной областью коры головного мозга испытуемым прикрепляли электроды. С помощью этих электродов безопасным для участников способом ученые снимали показания об изменении электрической активности в мозге и возникновении потенциала готовности, который всегда предшествует двигательной активности, сокращению мышц. Время возникновения потенциала готовности также записывалось для каждого «произвольного» действия, а потом сравнивалось с временем фактического действия и моментом возникновения желания, фиксируемым самим участником эксперимента. Как соотносились эти показания по времени?

Логичным было бы предположить, что первым событием должно было бы быть возникновение желания, фиксируемое самим испытуемым с помощью часов, ведь по условиям эксперимента решение о движении руки должно было быть произвольным и осознанным. Следом стоило бы ожидать возникновения потенциала готовности, так как он необходим для инициации действия. И наконец, само движение должно было замыкать каузальную цепочку. К удивлению уче-

ных, последовательность оказалась другой. У испытуемых первым возникал потенциал готовности, и только затем следовала фиксация желания. То есть к тому моменту, как у испытуемого возникало желание совершить движение, его мозг уже принял это решение. Это открытие может быть еще одним доводом в пользу утверждения об отсутствии у агента свободы, а значит, и ответственности. Из некоторых интерпретаций этого эксперимента следует, что свобода воли – это не более чем удобная иллюзия.

Из описанных выше экспериментов может следовать, что поведение личности определяется факторами, лежащими вне ее контроля: социальными силами и нейрофизиологическими процессами. Социальная динамика направляет личность, и поведение оказывается обусловлено не контролируемым агентом факторами. Сознание лишь регистрирует решение, устанавливаемое неосознаваемыми процессами в мозге, а агент испытывает иллюзию выбора. Но если это так, то три описанных выше условия ответственности не выполняются даже для здоровых нормальных личностей.

Эксперименты Зимбардо, Милгрэма и Либега, как кажется, демонстрируют, во-первых, что намерения агентов не являются источником их действий – этим источником служат нейрофизиологические бессознательные процессы или социальный контекст. Во-вторых, они, вероятно, показывают, что социальный контекст определяет поведение личности таким образом, что у нее стираются индивидуальные особен-

ности, то есть у личности не остается альтернатив. И в-третьих, они, возможно, свидетельствуют, что подлинным субъектом действия является мозг человека, его неосознаваемая активность, а не сознательная личность, выступающая объектом воздаяния и наказания. Таким образом, ни одно из указанных условий не выполняется, а если это так, то никто и никогда не должен нести ответственность, и наказание и благодарность в отношении людей столь же абсурдны, как наказание и благодарность в отношении животных или даже неодушевленных объектов. Следует ли принять такой вывод на основании эмпирических данных? Приводит ли наука к такому заключению?

Голландский философ и натуралист Б. Спиноза считал, что люди верят в свободу воли потому, что осознают собственные желания, но не осознают реальные причины этих желаний. По этой логике, сейчас, после прояснения некоторых причин поведения, у свободы воли и связанной с ней ответственности не остается оснований. Такую позицию разделяют некоторые выдающиеся современные западные ученые, среди которых Д. Вегнер, С. Харрис, М. Газзанига, Б. Либет, Ф. Крик и др. Есть и российские ученые, занимающие подобную позицию, например В. В. Никулин и В. А. Ключарев. Я, однако, не согласен с такого рода воззрениями. Я считаю, что любые эмпирические открытия сами по себе не могут доказать отсутствие или наличие ответственности и свободы. Открытия новых, ранее неизвестных и неосознавае-

мых причин действий не представляют действительной угрозы для этих важных убеждений. Эти причины могут представлять препятствие для ответственности только в составе философских теорий. Почему научные теории сами по себе не показывают, что свободы не существует? Почему ответственность и необходимая для нее свобода воли – это философские проблемы?

Предположим, эксперименты показывают, что источником действий личностей действительно являются неосознаваемые нейрофизиологические процессы. Означает ли это, что в таком случае намерения не могут быть источником действий? Ответ на этот вопрос лежит явно за пределами нейрофизиологии, ведь для этого нужно соотнести понятия «намерение» и «неосознаваемый нейрофизиологический процесс». «Намерение» входит в язык бытовой психологии и не имеет строгого определения. Чтобы произвести соотношение с терминами нейрофизиологии, нужно сначала прояснить его с помощью концептуального анализа, осуществить перевод, расшифровку. Это – задача для современных философов, занимающихся проблемой сознания. Выходит, чтобы выяснить, возможна ли свобода, нужно решить вопрос о соотношении ментального и физического – одну из ключевых философских загадок, разрешить многовековой философский спор. Но даже после этого решения философская работа не заканчивается – нужно разработать теорию причинности.

Предположим, намерение и его нейронный коррелят различаются между собой. Это по-прежнему не гарантирует, что оба явления одновременно не могут быть источниками действия. Источник действия – это причина действия, но именно философия изучает свойства причин. Может ли у явлений быть одновременно несколько причин, могут ли эти причины систематически коррелировать – все это вопросы теории причинности. Они должны быть определенным образом решены, чтобы понять, действительно ли неосознаваемые нейронные процессы являются препятствием для свободы воли. Впрочем, и само представление о свободе требует философского исследования. Что под ним следует подразумевать?

Согласно распространенному мнению, свобода предполагает наличие альтернативных возможностей. Это правдоподобное утверждение, но что именно здесь предполагается? Мы не можем непосредственно наблюдать альтернативные возможности, живя в мире актуальном. Тогда как определить их наличие или отсутствие? Здесь опять нам приходится прибегать к философии. Эмпирическая наука изучает природу преимущественно в том виде, в котором она существует, а альтернативные возможности – это гипотетические ситуации, «что, если бы...» ситуации. Мир гипотетических возможностей – это мир философа. Существует множество философских теорий, в которых разъясняется, что значит быть возможным, существовать как реальная альтерна-

тива. Философы различают логические и естественные возможности. Некоторые утверждают, что альтернативных возможностей вообще не существует, а некоторые – что они так же актуальны, как и фактический мир. Это значит, что, даже зная, что все поведение человека диктуется ситуацией, нельзя утверждать отсутствие у него альтернатив *tout court*.

И наконец, третий вопрос – об особой связи действующего агента и агента, несущего ответственность, – вопрос тоже философский. Для ответа на него нужно в первую очередь прояснить, какого рода связь необходима, а потом определить, может ли она быть реализована. Некоторые считают, что действующий и несущий ответственность агент должны быть тождественны. Отношение тождества в различные моменты времени является философским понятием, так как относится к наиболее общему типу отношений. И именно это отношение необходимо прояснить, прежде чем делать выводы о возможности агента быть ответственным. Более того, для отождествления действующего и несущего ответственность агента необходимо понять, что вообще следует считать агентом, какого рода эта сущность и из чего она может состоять. Таким образом, даже имея эмпирические доказательства, что мозг или какая-то его часть является действующим началом, а ответственным – целостная сознательная личность, без философского анализа нельзя утверждать, что действующее начало и личность не совпадают. Все эти соображения приводят нас к выводу о необходимости уча-

ствия философии в решении вопроса об ответственности и свободе. К сожалению, порой философским анализом пренебрегают. И даже выдающиеся ученые, совершившие значимые открытия в науке, не освоив философской методологии, делают ошибки, свойственные неофитам. Методология, которую я считаю наиболее эффективной, – это методология аналитической философии.

Истоки аналитической философии восходят к диалектике Платона в древности и размышлениям Декарта и Гоббса в Новое время. Наиболее влиятельными фигурами, сформировавшими аналитическую философию XX в., считаются Б. Рассел и Л. Витгенштейн. Оба этих философа стремились добиться максимальной ясности и математической точности в своих размышлениях. Эти принципы нашли отражение в ключевом методе аналитической философии – концептуальном анализе. Это исследование сложных понятий с помощью разложения их на более простые. Философский концептуальный анализ близок научному анализу. Так, развитие физики выражается, в частности, в возможности декомпозиции атома на протоны, нейтроны и электроны, развитие химии – в разложении сложных химических соединений на простые элементы. Другим методом аналитиков является эксперимент. Как и ученые, современные аналитические философы ставят эксперименты, но эти эксперименты – не физические, а мысленные. Мысленные эксперименты позволяют выявить базовые интуиции и показать, как они работа-

ют в той или иной гипотетической ситуации.

Еще одной важной чертой аналитической философии является ее связь с эмпирическими данными. Современная аналитическая философия не существует в изоляции, а является фундаментом для междисциплинарных обобщений. Это взгляд на научные открытия с высоты птичьего полета. Необходимость такой перспективы я показал выше. В этой книге я пытаюсь применить комплекс методов аналитической философии к проблемам ответственности и свободы воли.

Цель размышлений, изложенных в этой книге, – решение загадок о свободе воли и ответственности. В этом тексте я попытаюсь проанализировать наиболее значимые философские позиции и ответить на вопрос, как возможны свобода и ответственность. Выше было показано, что одни только эмпирические открытия, научные эксперименты, раскрывающие причины поведения человека, не способны убедить в отсутствии этой свободы и связанной с ней ответственности. Но в совокупности с некоторыми философскими теориями подобные научные открытия могут представлять подлинную опасность для убеждений в наличии у нас ответственности и свободы воли. Такими теориями являются физикализм, детерминизм и событийный каузализм. Каждой из этих теорий и диалогу с ее сторонниками в современной философии посвящена отдельная глава книги. В каждой из глав представлены альтернативные позиции по отдельной философ-

ской проблеме и намечено ее решение. В первой главе обосновывается возможность сочетания физикализма и ментальной каузальности, во второй – детерминизма и возможности альтернативных действий, в третьей – событийного каузализма и возможности приписывания действий личности. Итогом книги является доказательство возможности свободы и моральной ответственности, то есть вывод о том, что события нашей жизни – не просто ирония судьбы.

Неискушенный читатель может встретить в книге много незнакомых терминов. Но пусть это не вызывает удивления. Философия, как и любая другая наука, имеет свой словарь. Он необходим для точности выражения мысли. Но этот лексикон достаточно легко освоить. Сложнее с понятиями, встречающимися как в философии, так и в обыденном языке. У философов эти понятия часто имеют специальное значение (например, понятия «сознание», «причина», «агент»). В книге я не всегда разъяснял их, так как писал этот текст в основном для профессиональных философов, работающих в аналитической традиции. Надеюсь, впрочем, что большая часть терминологии будет понятна из контекста и чтение доставит вам удовольствие, даже если это – первая философская книга на тему свободы воли, которую вам доводилось держать в руках.

# **Глава 1. Проблема ментальной каузальности и возможность действий**

## **I. Аргументы в пользу ментальной каузальности**

### **1. Связь проблем ментальной каузальности и свободы воли**

Ответственность предполагает существование свободы воли и возможность действий. Действия не существовали бы без ментальной каузальности – способности ментальных состояний, убеждений, желаний, эмоций быть причинами поведения. Ведь действия отличаются от обычных событий в силу этих особых причин. Но могут ли ментальные состояния быть причинами поведения? Обосновано ли различие между действиями и обычными событиями? Обладаем ли мы свободой и есть ли основание у ответственности за поступки? Как ни странно, это не пустые риторические вопросы. В начале этой главы я расскажу о главном основа-

нии для сомнений в ментальной каузальности – физикализме. Физикализм – это убеждение в том, что все события в качестве своих причин имеют только другие физические события. Это фундаментальный принцип современной науки. При этом физикализм создает серьезную угрозу для возможности существования ментальной каузальности. Кажется, утверждение о том, что только физические события могут быть причинами любых других событий, ставит под сомнение возможность ментальной каузальности. Эта интуиция выражается в одном из важнейших аргументов современной аналитической философии – Аргументе исключения.

Из Аргумента исключения следует невозможность ментальной каузальности, отрицание свободного и ответственного действия. Такое следствие противоречит интуициям и показывает, что наше отношение к поступкам других людей иррационально. Неужели люди в принципе не могут поступать свободно? Неужели нет оснований для наказаний и похвалы? Надеюсь, что это не так. В поисках решения проблемы ментальной каузальности, разрешения противоречия между нашими интуициями и фундаментальным научным принципом, я обращусь к наиболее значимым, по моему мнению, теориям современных аналитических философов: локальному интеракционизму, биологическому натурализму и телеофункционализму. После исследования этих подходов я представлю собственное дополнение к одному из

них – телеофункционалистскому решению проблемы ментальной каузальности. Итогом главы будет телеофункционалистское обоснование возможности ментальной каузальности в условиях физикализма. Таким образом, будет преодолено первое препятствие на пути к возможности свободы и ответственности. Другие я рассмотрю во второй и третьей главах. А теперь сделаем шаг назад и попробуем убедиться в том, насколько действительно ментальная каузальность необходима, ведь само по себе это не очевидно.

## **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? Американский философ Д. Дэвидсон задается подобными вопросами в эссе «Агентность»: «В каких событиях в жизни личности проявляется агентность? Что является ее действиями и поступками в отличие от того, что просто случается в ее истории? Какая отличительная черта характеризует ее действия?» [Davidson 1980, 43]. Сегодня я проснулся рано утром, разбуженный звуком скрипки, на которой кто-то

занимался, умылся, побрился, 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, мотивов, стремлений и целей агента-человека» [MacIntyre 2007, 209]. 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 придется. Скажем, рефлекторные движения рук или ног явно нельзя отнести к действиям. Вздрагивания, кашель и храп не являются действиями, как и пробуждение или спотыкание [Davidson 1980, 44]. А полет радиоуправляемого самолета, скорее всего, – это действие, хотя сам пилот остается 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 прочих процессов также отличаются возможностью их 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. 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>2</sup>.*

В формулировке Кима неизбежной оказалась ссылка на оптимальность. Без нее 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 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 чаш-

---

<sup>2</sup> В таком виде этот принцип приводит Дж. Ким в книге «Философия сознания» [Kim 2006, 126].

ку чая, идете 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 выпить чаю будет 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 дополни-

тельных 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 остается 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 вскипятил воду? Потому что хотел выпить чаю. Почему он совершил убийство? Потому что он надеялся попасть в историю. Но тот факт, что объяснение действия предполагает указание на эти состояния, не дает 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) кау-

зальную связь, то есть 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 свершения 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, на-

пример логическая [Ginet 2002]. До публикации Дэвидсоном статьи «Действия, мотивы и причины» (1963) доминирующей позицией среди аналитических философов был антикаузализм. После публикации Дэвидсона ситуация радикально поменялась. О результате изменения философ, профессор Университета Флориды

А. Мили 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) характерными объяснениями действий являются каузальные объяснения» [Mele 2009, 554]. 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 1963 г. Любопытно, что Дэвидсон косвенным образом внес вклад в развитие и этой позиции.

## 4. Доводы против каузальной связи действий и ментальных состояний

### А. Трудности с доказательством психофизических законов

Традиционной моделью каузальности является номологическая модель<sup>3</sup>. Ее принимают многие современные философы. Модель предполагает, что каждый случай причинной цепи представляет собой частный случай действия общих законов. То есть 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 определенной температуры, к примеру, является причиной ее кипения, так как существует общая закономерность парообразования в жидкости при 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 действия-

---

<sup>3</sup> К примеру, ее принимали К. Гемпель и Д. Дэвидсон. «Согласно как Гемпелю, так и Дэвидсону, каузальность – это вопрос номологической достаточности:  $C$  является причиной  $E$ , если и только если существует закон, который соединяет события, подобные  $C$ , с событиями, подобными  $E$ » [Psillos 2009, 149].

ми, 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 невозможно: не получается 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 замрет 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, внутреннего самонаблюдения? 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, то есть 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 выпить горячий 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» [Davidson 2006b].

## **В. Аргумент аномальности ментального**

Отличительной чертой ментальных содержаний является рациональность и согласованность. 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 Херст» мы не будем интерпретировать как свидетельство наличия противоречащих убеждений, но как выражение того, что некоторые работы Херста ценителю современного искусства нравятся, а некоторые – нет. Итак, требования логической непротиворечивости и связности являются обязательными для ментальных содержаний. Но подобных ограничений не существует для физических событий. Одни физические события нельзя лишь с помощью одной только логики вычислить из других. Физические законы случайны, а не необходимы. Из этих различий, по мнению Дэвидсона, 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  $Y$ , то есть 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 концептуальное отношение. Для того чтобы опровергнуть каузализм, необходимо показать, что именно концептуальная 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 за-

ключается суть Аргумента логической связи (АЛС). Считается, что «это самый значимый аргумент против каузализма» [Stoecker 2010, 603]. Его отстаивали Г. Райл и Э. Энском<sup>4</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 другой 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 передаете солонку без желания это делать (ведь никто вас не заставляет передавать соль, и это действие – не

---

<sup>4</sup> Подробно Аргумент разбирается в статье Дж. Оттена «Возвращение к жизни Аргумента логической связи» [Otten 1977].

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 желание действия логически связаны. И все же это далеко не окончательный вывод. Ключевым оппонентом этого аргумента является Д. Дэвидсон.

## **5. Доводы Д. Дэвидсона в пользу каузальной связи**

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 мотива-

ми» (*primary reasons*). Философ утверждает: «1. Для того чтобы понимать, как какой-либо мотив рационализует 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» [Davidson 1980, 4]. 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 выпить чаю 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 убеждение просто подразумевается. Знание основного мотива рационализует действие, так как оно по-

казывает его телеологическую направленность. «В свете основного мотива действие предстает соответствующим определенным качествам агента (долгосрочным или краткосрочным, характерным для него или нет), и агент представляется в качестве Рационального Животного» [Davidson 1980, 8]. Но на каком основании автор считает, что основные мотивы являются причинами действий?

Ссылка на основной мотив является способом объяснения действия. Это объяснение рационализует 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» [Davidson 1980, 9].

Для примера возьмем случай, когда человек А случайно получает перевод средств, предназначавшийся человеку Б. А сообщает банковским служащим о том, что произошла ошибка [Lepore, Ludwig 2006, 4]. Основным мотивом является желание вернуть деньги тому, кому они предназна-

чались. Однако может быть такая ситуация, что существует еще один 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 мотивов может быть причиной. Именно этот мотив будет основанием 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 (АЛС). И во-вторых, необходимо продемонстрировать, как соотносится убеждение о ментальной каузальности 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 влечет 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 определенным 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, оправданно утверждать их логическую независимость и допустимость использования первых для объяснения вторых [Davidson 2006a, 25].

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 то же самое, что объединять событие и его причину, и объяснение с помощью такого описания не исключает каузальное объяснение» [Davidson 1980, 14]. 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 кау-

зальная 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, – классификацию можно провести и в нейрофизиологических, химических или физических терминах» [Davidson 1980, 17].

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 построены на других концепциях. Однако это общее описание может быть переформулировано с большей точностью и с помощью других концепций, и тогда оно будет частным

случаем общего закона. Таким образом, Дэвидсон снимает противоречие психофизической аномальности представлениям номологической модели и обосновывает ментальную каузальность. Впрочем, есть и другие аргументы в ее пользу.

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 их за действия. 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, что практика применения моральных категорий «очевидным образом предполагает, что наши ментальные состояния имеют каузальный эффект в физическом мире» [Kim 2008, 9]. 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: жела-

ниями и убеждениями? И как агент может нести ответственность за действия, если он является пассивным наблюдателем по отношению к ним?

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 действия от всех прочих событий. 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>5</sup>. Поэтому вопрос о ментальной каузальности представляется в большей степени не эмпирическим вопросом, существует она или нет, а философским: как возможно ее существование.

Относительно важности ментальной каузальности категорически высказался другой современный аналитический философ, Дж. Фодор: «Я не полностью убежден, насколько важно, что ментальное – это физическое; и еще меньше в том, что важно, чтобы мы могли это доказать. Однако если неверно то, что мое желание каузально отвечает за

---

<sup>5</sup> Принцип сформулирован по статье [Colyvan 1998].

мое действие, щекотка за чесотку, мое убеждение за мои выражения... если все это неверно, тогда все, в чем я убежден, – ложь, и это конец света» [Цит. по: Kim 2006, 181]. 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 [Беседин, Васильев, Волков, Кузнецов 2017]. С этим согласен и я. Однако это 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 [Taylor 1964] и Н. Малкольмом [Malcolm 1968]. В конце 80-х он был актуализирован американским метафизиком Дж. Кимом, а затем подробно разработан им же в статье «“Нисходящая каузальность” в эмерджентизме и нередуктивном физикализме» в 1992 г. Как пишет А. Кузнецов в своей диссертации, «Ким – философ, вокруг которого строятся дискуссии о проблеме ментальной каузальности» [Кузнецов 2016b, 52]. В наиболее поздней монографии «Физикализм или что-то около того» Ким представляет обновленную версию Аргумента, вписанную в более широкую теорию отношения ментального и физического [Kim 2008, 13].

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:

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

Согласно этому принципу утверждается, что достаточная причина у события должна быть одна, кроме специальных случаев сверхдетерминации. Под случаями сверхдетерминации 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 сверхдетерминированное событие, так как 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, взлет 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 этот 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 привел пример, когда она имеет место. Но подчеркну, что сверхдетерминация – это особенный случай. Систематическая, регулярная сверхдетерминация представляется невероятной. Одно дело – допускать единичный случай наступления смерти в результате расстрела стрелковым подразделением, а другое – 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 при-

чиной 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  $M$ . То есть  $P \rightarrow P^*$ . Если  $M \neq P$ , то у  $P^*$  две причины: ментальная и физическая. То есть  $P$  – 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*, то есть не имеют каузальной силы. События развивались бы точно так же, даже 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 те-

ла, это лишает агента возможности существования...» [Kim 2006, 217].

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 разных случаях и у разных индивидов разными физическими состояниями. К примеру, боль как ментальное свойство может быть реализована 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, то есть 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). 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 эмерджент-

тисты. 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*, то есть зависимо и определено базовыми физическими свойствами. ТС имеет множество формулировок, к некоторым из которых мы обратимся ниже, однако для целей своего аргумента Ким представляет его в самом общем виде:

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

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 должно иметь некое физическое событие  $P^*$ , ведь ментальное не существует автономно. Теперь применим ПКИ, точнее вариант ПКИ. В исходном виде этот принцип не совсем подходит, ведь он относится к случаям каузальной связи. Но его можно обобщить до любых генерирующих (определяющих) связей, в том числе и отношения супервентности. Тогда он будет звучать так:

(ПКСИ) *Если событие  $e$  определено событием  $c$  – каузально или отношением супервентности, –*

*тогда событие  $e$  не может быть определено другим, отличающимся от  $c$  событием, кроме случаев сверхдетерминации.*

Применив этот принцип, получим: если событие  $M^*$  имеет достаточное основание в событии  $c$ , это и есть его единственное основание (кроме случаев сверхдетерминации). Но которое из двух перечисленных оснований,  $M$  или  $P^*$ , 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» [Kim 2008, 36]. Тогда остается 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 вы-

несен отрицательный вердикт в связи с Аргументом исключения. Круг замкнулся, и теперь мы в полной мере наблюдаем трудность философской проблемы ментальной каузальности.

Похоже, что ментальные состояния не могут каузально 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 проблема сознание-тело 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 среду благодаря сознанию и ментальности. Аргументы, представленные выше, уже показали, что физикализму не удастся сохраниться без изменений...» [Kim 2008, 31].

За последние несколько десятилетий вопросу о соотношении ментального и физического (проблема «сознания») и проблеме ментальной каузальности были посвящены мно-

гие сотни работ. Значительная их часть написана философами. Но растет также количество исследований, предпринятых учеными-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 все в окружающем нас мире в конечном итоге состоит из физических объектов и зависит от физических свойств. С появлением новых книг растет количество аргументов, и авторы с критикой набрасываются на тексты своих коллег, так что у любопытствующей публики создается 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>6</sup>, опубликованным в последнее время, можно проследить оригинальную философскую позицию, которую защищает российский философ В. В. Васильев. Автор называет ее «*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 ре-

---

<sup>6</sup> См. работы Васильева В. В. «Трудная проблема сознания» [2009], «Сознание и вещи: Очерк феноменалистической онтологии» [2014], «В защиту классического компатибилизма: Эссе о свободе воли» [2017], а также [Васильев 2016] и др.

ально 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 2004 г. при размышлении над книгой «Сознающий ум» Д. Чалмерса [Васильев 2009, 216]. На этом аргументе основывается важная часть теории локального ин-

теракционизма и решение проблемы ментальной каузальности.

## **2. Аргумент каузальных траекторий в пользу ментальной каузальности и против локальной супервентности**

Аргумент сформулирован в виде трех посылок и вывода. Первая посылка – это утверждение о том, что 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 память отражает события, происходившие с субъектом в прошлом, то есть отражает каузальную историю субъекта. Конечно, память может представлять каузальную историю ошибочным образом, но это скорее исключение, чем норма [Васильев 2009, 217]. 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, 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 (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...» [Васильев 2009, 217–218].

Итак, из аргумента следует, что, находясь в том же самом физическом состоянии, агент мог бы вести себя различным образом. И эти различия обуславливаются только его качественными ментальными содержаниями, различными убеждениями и желаниями. И никакой анализ наличного физиче-

ского состояния не способен дать точный прогноз действий субъекта. Ментальные состояния действительны, 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 ввел 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 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» [Davidson 2006b, 111]. 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 разделяет 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 супервернтно на физическом? Тем важнее оказывается аргумент, который подвергает этот тезис сомнению, отрицает супервернтность. Впрочем, здесь необходимо сделать некоторые уточнения.

Возможны как минимум три формулировки Тезиса о супервернтности. (1) Ментальное супервернтно 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 супервернтно на физическом в том смысле, что объекты, совершенно одинаковые в своих физических свойствах, не могут отличаться относительно ментальных свойств. (3) Ментальное супервернтно на физическом в том смысле, что миры, одинаковые во всех физических отношениях, одинаковы и во всех ментальных отношениях; то есть миры, одинаковые физически, совершенно одинаковы в целом<sup>7</sup>. Сила этих утверждений различна. Наиболее сильным является утверждение (1), так как оно фиксирует супервернтность даже одного изолированного свойства.

---

<sup>7</sup> Варианты Тезиса о супервернтности приведены по книге [Kim 2006, 9–10].

А 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 возьмем 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 о супервентности с выводом Аргумента каузальных траекторий?

Очевидно, реальная возможность существования двух идентичных людей с разными ментальными состояниями, допускаемая в качестве результата этого аргумента, противоречит варианту (1) ТС. Но сам Васильев сообщает, что его выводы не противоречат «глобальной супервентности»: «последняя формулировка выражает так называемую “глобаль-

ную супервентность”, и мы увидим, что предложенные мною аргументы не отрицают ее. Вторая формулировка тоже не опровергается ими, если она допускает глобальную интерпретацию. Ошибочно лишь первое понимание» [Васильев 2009, 243]. 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 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 сохранен, 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, но также и 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, вернемся к аргументу еще 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 лю-

дей. 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?

Васильев 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. То есть 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 Васильева, эти положения не могут быть оспорены, так как отражают устройство наших познавательных способностей. Одним из базовых положений Васильев считает убеждение «в соответствии прошлого и будущего» [Васильев 2009, 206]. На нем 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» [Васильев 2009, 208], Принцип каузальной замкнутости физического. По мнению автора, аксиома соответствия прошлого и будущего обуславливает также веру в возможность существования только одной причины, только одного События А, предшествующего Событию В. Мне кажется, что следование здесь далеко не очевидно. Но в данном случае я не берусь оценивать умозаключения Васильева, так как 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 Васильева, «Сознание и вещи» [Васильев 2014].

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 положение 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 antecedentные события. Отскок мяча на двести метров мог иметь гораздо меньше причин, чем просто от-

скок 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>8</sup> может иметь только одну причину? Возьмем сингулярное событие и создадим для него исчерпывающее конкретное описание. Это описание будет бесконечно длинным, пусть так. Но будет ли у этого конкретного события возможность в качестве антецедента 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 ан-

---

<sup>8</sup> Васильев использует термин примерно в том же смысле, «что Витгенштейн понимал под миром» [Васильев 2009, 241].

тецедентов 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, то есть все три варианта формулировки ТС.

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  $C$  в результате различных событий ( $A$  или  $B$  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  $C$ , но моя память при этом могла бы отражать одно из различных предшествующих событий ( $A$  или  $B$  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 послылки становятся верными. Мозг может прийти к некоторому состоянию  $C$  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  $C$  в результате разных событий, реально возможно, чтобы память отражала произошедшие с субъектом события. Но в этих двух послылках – можно возразить – речь идет 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. Значит, связь между послылками 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 пришел к состоянию  $C$  в результате различных нейронных событий ( $A$ ,  $B$  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  $C$ , но моя память отражала различные биографические события ( $A$ ,  $B$  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 конем, но нельзя начать игру сразу с двух ходов, конем 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 пришел 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 казуальную 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 пришел к нему, он, будучи чисто меха-

нической конструкцией, должен в этом состоянии одинаково реагировать на свое окружение. Итак, для поведения нашего робота-зомби значима не его индивидуальная история, а исключительно его наличное состояние» [Васильев 2009, 221]. Наличное состояние робота может отражать прошлые события. Но, по мнению российского философа, механическая память робота не чувствительна к его фактической индивидуальной истории. На чем основана такая уверенность?

Возьмем убеждение, что я родился в 1976 г., и запись компьютера о том, что первый файл на его диск записан в 1976 г. В чем принципиальная разница? Оба «воспоминания» содержат хронологическую метку, оба могут 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 жесткий диск. Повреждения или значимые изменения в физическом носителе в обоих случаях приводят к изменениям или исчезновениям воспоминаний.

Может, различие в том, что механизм компьютерной памяти относительно хорошо изучен, а механизм работы человеческой памяти все еще отчасти остается загадочным для ученых? Но отсутствие понимания всех механизмов работы человеческой памяти не является доводом в пользу нали-

чия у нее нефизических свойств и отсутствия локальной супервентности. Или различие 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 еще одно, возможно, ключевое отличие механической памяти от па-

мента человека. Видимо, именно оно должно делать АКТ 1 правдоподобным, а его механическую вариацию, АКТ 2, – нет. В представлении автора память человека – это приватные качественные состояния. И поэтому «индивидуальное прошлое не растворяется в нейтральных нейронных кодах, а продолжает существовать на уровне ментальных образов» [Васильев 2009, 221]. Это утверждение нужно использовать как еще 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 Васильев говорит о том, что человеческая память состоит из приватных качественных 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 скошенной травы, слегка греет солнце, чувствую удовольствие с оттенком сожаления. Воспоминания как бы оживляют пережитые ощущения, мои личные, те, что никто другой, кроме ме-

ня, 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, то есть помнит тысячи слов, выражений и грамматических 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 Бенджамина 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 мое поведение. Но чтобы иметь кажимость влияния, они не должны быть как-то качественно окрашены.

Теперь попробуем расставить все 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 действенности «кавали-

тативно окрашенных» и «квалитативно пустых» воспомина- ний? Или углубиться в интроспективные исследования, что- бы все-таки найти «вкус» всех без исключения записей па- мяти? Мне кажется, однако, что обе альтернативы беспер- спективны. И вернее считать, что память не имеет «квали- тативных качеств». Тем более что такую точку зрения могут занимать не только элиминативисты (противники существо- вания квалиа), но даже такие квалиафилы, как Д. Чалмерс.

## **В. Память не является квалитативным состоянием**

В своей основной монографии, «Сознающий ум» (1996), Д. Чалмерс, один из самых известных современных аналити- ческих философов, утверждает, что ментальное исчерпыва- ется психологическим и феноменальным (квалитативным). Некоторые ментальные свойства сводятся только к психоло- гическим, некоторые – только к феноменальным, а некото- рые имеют двойственный характер. Боль, к примеру, име- ет как психологические, так и феноменальные аспекты. Пси- хологической составляющей является состояние, связанное обычно с повреждением организма и ведущее к особому по- ведению, реакции уклонения. Как мы видим, это описание функциональное – оно фиксирует входящие параметры и со- ответствующее этим параметрам поведение-реакцию. Фено- менальным аспектом боли является ее мучительное пережи-

вание. Концептуально это не связанные 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 ее 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...» (курсив мой. – Д. В.) [Chalmers 1997, 24]. И причина в том, что процесс запоминания и хранения в памяти можно легко представить неосознаваемым, лишенным феноменальных свойств. Почему тогда автор Аргумента каузальных траекторий считает, что память состоит из приватных качественных состояний?

Вероятно, Васильев, когда пишет о памяти, имеет в виду не процесс запоминания и хранения информации о прошлом, а воспроизведение этой информации в текущем опыте субъекта. Только в текущем опыте часть воспоминаний действительно представляется феноменально окрашенной. Но тогда точнее будет сказать, что не память обладает качественностью, а текущее состояние, объектом которого являются события из прошлого. Содержание памяти может быть воспроизведено в сознании с феноменальными свойствами, но не имеет их в процессе хранения. Качественные аспек-

ты 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 не осознаваемых 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 улетучиваться 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 фено-

менальных 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 «придется забыть очень многое или даже почти все, если двойников очень много, а ведь это тоже реально возможно» [Васильев 2009, 219]. Чем больше двойников с альтернативными траекториями, тем больше воспоминаний должны быть стерты. 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» [Васильев 2009, 219]. 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 большими подробностями. Итак, есть два двойника, два почти неотличимых по физическому строению человека, живущие в разных мирах. Пусть один из них, Дэн, родился в Бостоне, а другой, Фил, в Бейруте. И у них разные занятия: Дэн бизнесмен, а Фил – музыкант. Один проводит все время в офисе, другой – играет джаз на фортепиано в маленьком ресторанчике. В какой-то момент параллельных, но не пересекающихся жизнью 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 провел 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, ведет 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

большого 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 нечетных чисел, останется бесконечное множество четных. 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 приобрел 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, зная 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. И что Тезис о супервентности не ведет к противоречивым последствиям. Это означает, что локальный интеракционизм, скорее всего, не дает бесспорного решения проблемы ментальной каузальности. Но тогда разве не возвращаемся мы к исходной дилемме? После критики локального интеракционизма разве не вынуждены мы отказываться от ментальной каузальности вообще? Или придется 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>9</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>10</sup>. И также утверждает каузальную эффективность ментальных состояний. Но проблему ментальной каузальности американский философ

---

<sup>9</sup> См. [Сёрл 2002; Сёрль 2004; Searle 1997, 1998, 1999, 2004, 2006, 2008 и др.].

<sup>10</sup> В одной из своих монографий Сёрл пишет: «Я настаиваю на том... что мы можем принять очевидные факты физики... не отрицая в то же самое время очевидные факты нашего внутреннего опыта – например, что все мы обладаем сознанием и что наши сознательные состояния имеют достаточно специфические *нередуцируемые* феноменологические свойства» (курсив – авторский) [Searle 1994, 28].

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.

*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) качественный характер, (в) интенциональность и (г) единство опыта [Searle 1994]. Под субъективностью ментального имеется в виду именно онтологическая, не эпистемическая субъективность. Сёрл пишет: «ментальные состояния обладают нередуцируемой субъективной онтологией» [Searle 1994, 19].

Для Сёрла ментальные феномены существуют только как часть внутреннего опыта человека (или животного).

Когда Сёрл говорит о качественном 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 свежеспеченного 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, запахов и проч. Сёрл подчеркивает здесь холистическое свойство сознательного опыта<sup>11</sup>.

Итак, каузальная эффективность ментального тоже утверждается обоими философами. Только российский философ считает это утверждение выводом из истинных посылок, а Сёрл – очевидностью. По мнению Сёрла, ментальная каузальность – это эмпирический факт, с которого и нужно начинать. «Я думаю, что мы ощущаем необходимую [каузаль-

---

<sup>11</sup> Резюмирующее описание теории Сёрла можно найти в его статье «Биологический натурализм» [Searle 2007].

ную] связь почти всю нашу сознательную жизнь...» [Searle 2004, 203]. 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 еще, 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 мистическим или необъяснимым феноменом. Она 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 ее 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. Такой вывод следует из открытий Д. Юма. Фактом является не наличие причинно-следственной связи, в частности, между желанием и действием, а регулярное следование одного за другим во времени. 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 днем 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 разберемся в пунктах, в которых сосредоточены основные различия. Сёрл утверждает, что ментальные состояния полностью обусловлены процессами в мозге. То есть 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. «Обобщая... мы могли бы сказать, что два феномена могут быть как связаны каузальной связью, так и реализованы один другим, если они находятся на разных уровнях описания» [Searle 1999, 266]. 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 подчинительного союза «потому» и каузальной. Кажется, Сёрл демонстрирует, что союз «потому» может указывать сразу на оба отношения: и концептуальное, и каузальное. Примеры самого Сёрла лучше проясняют 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>12</sup>. Если отбросить связь с агентом (пока оставим ментальную каузальность в стороне и рассмотрим только очевидно физические процессы), кипение можно объяснять так: вода закипает потому, что разогревается зажженной конфоркой до соответствующей температуры. То есть (1) разогрев воды на огне приводит к переходу ее 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 кинети-

---

<sup>12</sup> Этот пример Сёрл рассматривает в книге «Открывая сознание заново» [Searle 1994, 87].

ческая энергия, передающаяся в результате окисления углеводородов (горения топлива) молекулам  $H_2O$ , заставляет их двигаться так быстро, что между ними разрываются (полярные и водородные) связи. То есть (2) повышение кинетической энергии до определенного уровня приводит к разрыву связей между молекулами воды. Таким образом, у нас два совершенно законных каузальных объяснения перехода воды 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 antecedentом и consequentом. Эти описания отражают последовательность событий во времени, поэтому вслед за Сёрлом назовем их описаниями «слева направо» [Searle 1994, 87]. 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. То есть 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 кинетической энергии молекул воды до определенного уровня является причиной кипения, (4) разогрев воды до определенного 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 [Searle 1994, 87]. Эти способы возникают в силу того, что одно и то же явление можно описывать на разных уровнях. И, по всей видимости, конкуренция за каузальную эффективность между этими описаниями снимается через идентификацию явлений нижнего и верхнего уровней.

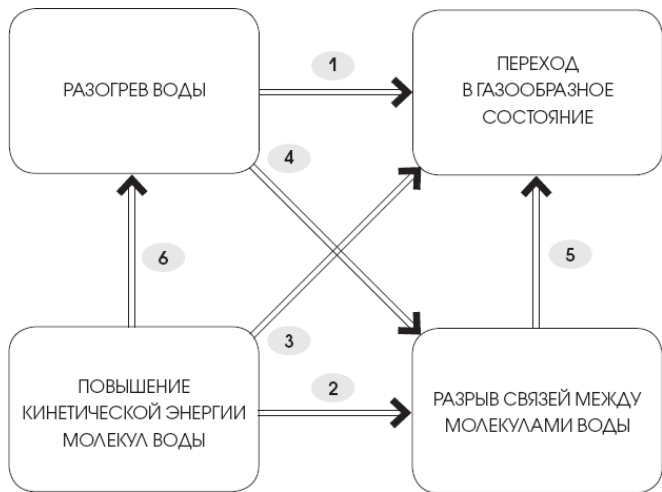
Но с позиции Сёрла в описанном явлении существует еще один тип каузальных связей – «снизу вверх». Этот тип связи наиболее экстравагантен. Судя по его рассуждениям<sup>13</sup>, Сёрл утверждает наличие каузальности между одномоментными, синхронными событиями на разных уровнях. То есть повышение кинетической 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

---

<sup>13</sup> Эти рассуждения Сёрла о ментальной каузальности приведены в Эпilogue к его 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 1999, 262–272].

причины кипения воды: снизу вверх, микро-макро. Конкуренция в этом случае, по его мнению, также не является проблемой. Ведь речь идет об одном и том же явлении.

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 ее и предлагает.



## *Рис. 1.* Схема каузальных связей процесса кипения воды по Сёрлу

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, так же 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 события 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.

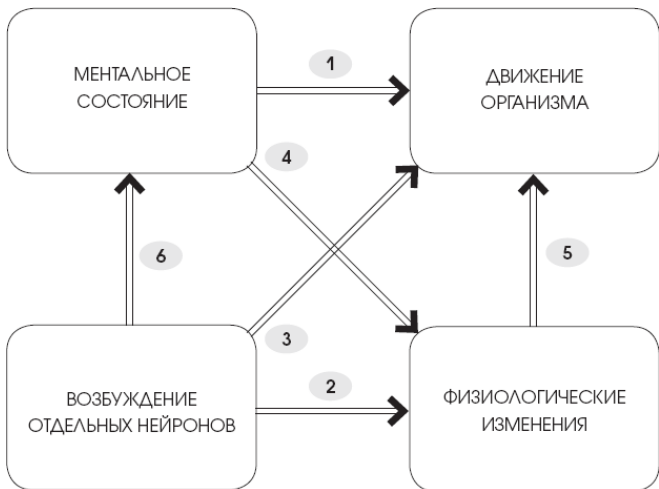


Рис. 2. Схема каузальных связей движения организма по Сёрлу<sup>14</sup>

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?

<sup>14</sup> Подобный рисунок Сёрл приводит в том же эпилоге [Searle 1999, 270].

### 3. Внутренние противоречия теории Дж. Сёрла

Само описание причинных связей в примере Сёрла вызывает у меня сомнение, в особенности 5-е 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 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 редуکتивно-супервентное отношение явлений разного уровня или логическая связь, а не каузальная. То есть газообразное состояние воды редуцируется к свободному и быстрому движению ее молекул и является супервентным 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>15</sup>. Причинная связь асимметрична, причины опре-

<sup>15</sup> К подобным выводам приходит в своей диссертации и А. В. Кузнецов. Он

деляют 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, оправданно свойства целого выводить из свойств его частей. Физика считается самой фундаментальной наукой именно потому, что изучает 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 каузальной зависимости, мне кажется, предполагает асимметричность, а Сёрл ее нарушает. По-

---

пишет: «Понятие эмерджентной причинности [Сёрла] 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 2016b, 109].

этому 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, качественность, 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 поменим кратность увеличения, то увидим несколько молекул,

их скопление и общее движение. Если еще уменьшить кратность, то, наконец, станет видно движущуюся каплю воды, то есть 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, то есть существует как приватные субъективные состояния. Но это принципиально отличает сознание от физического. И следовательно, аналогия в теории Сёрла не проходит. Из все-

го этого можно сделать вывод о том, что решение проблемы ментальной каузальности в биологическом натурализме не оптимально. 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 и известных современных аналитических философов. Он – не только автор множества монографий и сотен статей, но и сам герой множества книг. 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>16</sup>. Разработанный Деннетом подход можно назвать телеофункционалистским (ТФ). «Я, конечно же, – своего рода “телеофункционалист”, возможно, даже первый телеофункционалист» [Dennett 1991b, 460], – пишет он в монографии «Объясненное сознание». Теория сознания Деннета оснащена множеством эмпирических положений, но в ней

---

<sup>16</sup> Этот призыв Деннет сделал в статье «От кабинетной философии – в поле» [Dennett 1998, 289].

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 [Dennett 1991b, 281]. 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 высокоуровневые – в низкоуровневых. Деннет также, судя по всему, разделяет Принцип каузальной замкнутости физического мира, но допускает наличие высо-

коуровневых 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 входящих 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, определенные 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>17</sup> ментальных состояний. То есть, по мнению

---

<sup>17</sup> Это понятие ввел в философию сознания Х. Патнэм. В серии статей, направ-

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

Позицию Деннета можно для контраста сравнить с другим видом функционализма – с функционализмом машины Тьюринга (ФМТ)<sup>18</sup>. Эта теория несколько более ранняя и, возможно, менее совершенная. Для иллюстрации общности и 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 одинаковым для обоих вариантов функционализма. Он служит идентификатором *монизма*. «Каждое ментальное состояние – это какое-то физическое состояние», – считает большинство

---

ленных против теории тождества, он показал, что ментальные состояния могут быть реализованы различными физическими состояниями, поэтому ментальное и физическое не могут быть в строгом смысле тождественны друг другу [Putnam 1967].

<sup>18</sup> В «Объясненном сознании» Д. Деннет пишет: «Тогда являюсь ли я функционалистом? И да, и нет. Я не функционалист машины Тьюринга» [Dennett 1991b, 460].

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 скажет, что общность между двумя такими системами в том, что с учетом этого убеждения поведение обеих систем (в той мере, в которой оно определяется этим убеждением) можно будет надежно предсказать.

Таким образом, (1) телеофункционалист не требует ни общности в физической реализации, ни общности «программы» (мыслящие существа могут в силу образования, воспитания и наследственных данных значительно различаться «алгоритмами»). Могут быть запущены разные презентации на разных 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 телеофунк-

ционалистской. 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, ее 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 – произвольной функцией. К тому же устройство термостата

не позволяет ему самому определять ответные «действия». Термостат можно использовать как для регулирования потока хладагента, поступающего в радиатор, так и для запуска ядерной ракеты. Биологические сенсорные системы 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 надежные 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 существованием 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» [Dennett 1981, 9].

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 XIX в. Интенциональность (от латинского *intendo* – указывать на что-то, целить во что-то), эта отличительная черта психического, как было подмечено австрийским философом Ф. Brentano, указывает на способ-

ность 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»: «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 – любит, в ненависти – ненавидит, в желании – желает и т. д.» (курсив мой. – Д. В.) [Brentano 1924, 124–125; цит. по: Молчанов 2007, 412–413; см. также: Brentano 1996, 33]. Деннет сохраняет эту 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 реше-

ние проблемы ментальной каузальности. Однако в данном случае решение Деннета не представляет собой строгий аргумент и уж тем более развернутую теорию. Когда философа спрашивают о ментальной каузальности, он всегда приводит свой мысленный эксперимент, который называется «История двух черных ящиков» [Dennett 1995, 412]. Деннет считает, что описание этой гипотетической 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 черных ящика. На первом расположени 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 чем состояло отличие, ученые подключили к кабелю «прослушивающее» устройство. И вскоре обнаружили, что, когда нажималась любая из кнопок, по кабелю передавался поток импульсов с интервалами, как бы нулей и единиц, всего 10 000 бит. 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 бит 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 памя-

ти компьютера никогда не оставалось прежним. Наблюдая за работой компьютера, ученые решили, что любая последовательность из 10 000 бит 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, то есть 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 еще 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 фи-

зическими законами. 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 зеленой. 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 нем 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 псевдопроизвольности всегда, когда нажималась кнопка  $\alpha$ , загоралась красная лампочка, а когда  $\beta$  – 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 еще более загадочным для ученых.

Но однажды в лабораторию зашли два программиста, Эл и

Бо, которые создали этот аппарат, и все 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 истинные утверждения, а при нажатии  $\beta$  – ложные. Вторая машина сопоставляла утверждение на входе с собственными «убеждениями» и, если получала истинное утверждение, подтверждала его красным сигналом, а если ложное – откликалась зеленым. Когда данные на входе невозможно было обработать, то есть 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 распространенной реакцией. Так раскрылось, что удивительное свойство красных после-

довательностей было свойством «быть истинным утверждением», а зеленых – «быть ложным утверждением». Неожиданно, поняв устройство аппарата, ученые сами смогли создавать сколько угодно зеленых и красных последовательностей. Они даже открыли последовательности, которые сначала были красными, а потом становились зелеными, как, например, сообщение: «это сообщение передавалось менее 100 раз». Ведь на 101-й раз это сообщение должно было становиться зеленым.

### **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, была опубликова-

на в двух книгах Деннета: «Опасной идее Дарвина» (1990) и монографии «Помпы для интуиции» (2013)<sup>19</sup>. И в обоих случаях автор представлял лишь очень краткое разъяснение этого мысленного эксперимента. Поэтому я приведу собственную интерпретацию. С моей точки зрения, мысленный эксперимент Деннета иллюстрирует каузальную эффективность высокоуровневых свойств. Только этим можно объяснить функционирование этой сложной системы.

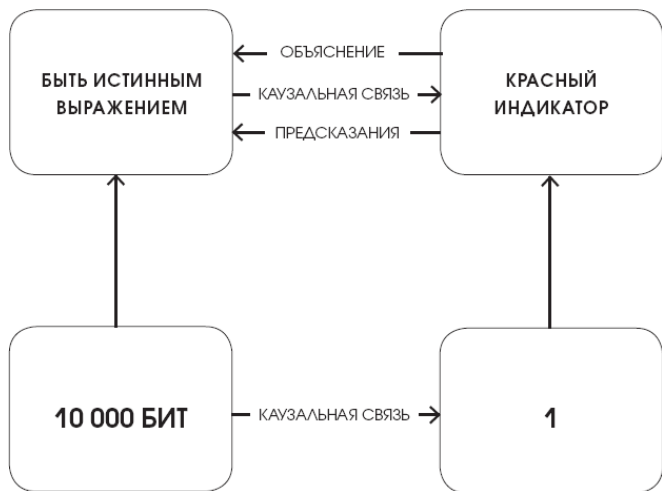


Рис. 3. Два черных ящика Д. Деннета

Описанная Деннетом система состоит из двух уровней. На

<sup>19</sup> Также эта история была опубликована в журнале *Activitas Nervosa Superior* [Dennett 2010].

базовом уровне ее функционирование определяется последовательностями из символов: нулей и единиц. А на высоком уровне последовательностями из утверждений: утверждения могут быть ложными или истинными. Поскольку каждый раз из ящика *A* в ящик *B* передается 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, а не ложные 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. Таким образом, результатом мысленного эксперимента можно считать 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, то есть последовательности битов, не содержат какого-то общего паттерна и потому не могут обосновывать каузальную закономерность. Синтаксис и семантика одновременно присутствуют в системе, но только семантические свойства имеют каузальную силу. Это не уникальный случай такого сочетания свойств. 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. В предшествующем событии реализованы разные свойства, но только некоторые из них каузально эффективны. Вымышленная ситуация Деннета построена так, чтобы продемонстрировать, как семантическое содержание может быть каузально релевантным свойством, как форма или вес, а не как цвет или материал по отношению к следу на песке. Но, вероятно, каузальную действенность семантических свойств в механической системе можно было продемонстрировать как-то проще?

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 та-

кого «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» [Dennett 2013, 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 – «зеленой». Мне кажется, установка действительно соответствует минимальным требованиям и не чересчур усложнена. Но достигает ли Деннет с ее помощью поставленных целей? Удастся 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 опро-

вергнута, во-вторых, что-то должно было измениться с момента создания базы и, в-третьих, то, что было истинным для машины *A*, могло не быть истинным для машины *B*. К примеру, высказывание «Изотоп антигелия-4 был зарегистрирован на ионном коллайдере RHIC» могло быть ложным до 2011 г., истинным после 2011 г. 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 *B* 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 истинным» или «быть ложным выражением» в качестве таких свойств уже не работают. Так как в случае с предварительно согласованной базой высказываний программисты могли бы допустить совместные ошибки. Также в ходе исследования ящиков ученые могли случайно набрать последовательность символов, которая была бы истинной, 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  $A$

# Конец ознакомительного фрагмента.

Текст предоставлен ООО «ЛитРес».

Прочитайте эту книгу целиком, [купив полную легальную версию](#) на ЛитРес.

Безопасно оплатить книгу можно банковской картой Visa, MasterCard, Maestro, со счета мобильного телефона, с платежного терминала, в салоне МТС или Связной, через PayPal, WebMoney, Яндекс.Деньги, QIWI Кошелек, бонусными картами или другим удобным Вам способом.