

Свобода воли, физикализм и моральная ответственность

Секацкая Мария Александровна

Стенограмма доклада:

Видеозапись доклада можно прослушать:

<http://tube.sfu-kras.ru/video/2077?playlist=2074>

Уважаемые коллеги, добрый день! Во-первых, спасибо большое организаторам конференции, за то, что вы нас пригласил, это для нас большая честь! Спасибо тем, кто пришел! Надеюсь, это окажется крайне интересным и важным мероприятием. Надеюсь внести в него свою лепту. Итак, мой доклад называется «Свобода воли, физикализм и моральная ответственность» и в этом докладе я бы хотела проанализировать два вопроса, а именно: инкомпатибилистскую позицию в современных аналитических дебатах о свободе воли и релевантность данных современной нейронауки для истинности или ложности инкомпатибилистской позиции.

Соответственно содержание доклада можно видеть на экране в презентации (введение: постановка проблемы; два определения свободы воли; нейронаука и свобода воли; интерпретация экспериментальных данных).

Вначале хочу определить используемые в моем докладе термины, поскольку это термины нагруженные большим количеством значений, итак: физикализм – философский тезис об онтологии, подтверждает, что в мире существуют только физические объекты и явления, подчиняющиеся законам физики, в соответствие с таким образом понимаемым тезисом физикализма все ментальные психологические объекты и явления являются физическими объектами и явлениями и подчиняются физическим законам. Инкомпатибилизмом относительно физикализма и моральной ответственности буду называть философский тезис, который состоит в том, что моральная ответственность человека за свои действия несовместима с истинностью физикализма. Противоречия между истинностью тезиса физикализма и возможностью существования свободы воли возникает в связи с тем, что в аналитической философской традиции есть две доминирующие концепции свободы воли, но каждая из них, как кажется инкомпатибилистам, она не совместима с тезисом физикализма. Противоречие между истинностью тезиса физикализма и возможностью существования моральной ответственности вызвана тем, что свобода воли представляется необходимой для моральной ответственности, таким образом, если свобода воли не совместима с физикализмом, но необходима для моральной ответственности это значит, что моральная

ответственность не совместима с физикализмом. И это тот важнейший аргумент, в связи с которым инкомпатибилисты настаивают на том, что мы должны отказаться от физикализма относительно вопроса о человеке, неважно насколько все остальные причины принять физикализм были бы убедительными. То есть для нас очень важна моральная ответственность, а что если физикализм действительно с ней не совместим? Получается, что дело плохо... Я хочу подставить под вопросом это рассуждение. Я собираюсь последовательно рассмотреть и подставить под вопрос эти противоречия, а затем привести несколько соображений о том, что именно данные нейронаук говорят о проблеме свободы воли. Итак, свобода воли в этих дебатах традиционно понимается как: первое – как способность выбора между альтернативными вариантами действия, либо как второе – присущая агенту способность быть автором собственных действий. Инкомпатибилисты утверждают, что ни то ни другое определение свободы не совместимо с истинностью физикализма. Противоречие между свободой воли в первом смысле и физикализмом можно выразить следующим образом: в соответствии с тезисом физикализма, каждое действие человека является действием определенной физической системы. И это действие следует из предыдущего состояния этой системы. То есть, состояния человеческого тела, мозга и всех релевантных факторов среды, действующих в момент осуществления действия и каузально влияющих на эту систему, определяется из естественнонаучных законов, по которым функционирует эта система. Формальное изложение несовместимости первого определения и тезиса физикализма дано в работе Ван Инваген (1975). Если законы физики имеют детерминистский характер, то любое действие человека с необходимостью должно следовать из состояния тела и мозга, факторов среды и естественнонаучных законов. Но, если действие следует с необходимостью, то человек не может выбрать поступить ли ему так или иначе, он может поступить одним единственным способом, следовательно, свобода воли в первом смысле не совместима с тезисом физикализма, если законы физики имеют детерминистский характер. С другой стороны, если законы физики и имеют индетерминистский характер и этот индетерминизм распространяется на макро объекты, например, на работу нейронов в головном мозге, ситуация может показаться столь же плачевной. В самом деле, индетерминизм означает случайность. Если действие человека является таким состоянием физической системы, которая лишь вероятно наследуется из предшествующего состояния этой физической системы, то человек точно так же не способен выбрать, какое действие ему совершить, потому что у него отсутствует контроль над случайными событиями. Таким образом, инкомпатибилисты говорят, что даже если физика имеет индетерминистский характер, то свобода воли в первом смысле, все равно не совместима

с физикализмом. Их противники по дебатам о свободе воли, позиция которых носит название компатибилизм, то есть тезис о совместимости свободы воли и физикализма, они традиционно критиковали это рассуждение, исходя из лингвистического анализа смысла слов «может поступить иначе». Классическая компатибилистская стратегия такого анализа состоит в том, чтобы понимать фразу «некий человек мог поступить иначе», если бы у него были другие мотивы, желания и принципы, и т. д.». Если анализировать это высказывание, в таком образе, оно уже не противоречит детерминизму, хотя наличие свободы при индетерминизме остается проблематичным. Другая компатибилистская стратегия связана с именами Питера Стросона, Франкфурта, и состоит в том, что признать физикализм не совместимым со свободой в первом смысле, но при этом он не противоречит существованию моральной ответственности потому, что для моральной ответственности не требуется способность в каждой конкретной ситуации поступить так или иначе, что тогда требуется, вероятно, понимание свободы во втором смысле. То есть, как присущую агенту способность быть причиной собственных действий. Здесь инкомпатибилисты полагают, что свобода во втором смысле тоже несовместима с физикализмом, потому что быть причиной собственных действий с их точки зрения означает быть изначальной причиной, то есть такой причиной, которая сама не является следствием никакой другой причины. Является *causa sui*, как говорил Спиноза. Это еще напоминает нам представление о перводвигателе у Аристотеля, который движет все сам, будучи недвижимым. Есть такие философы, которые рассматривают свободу воли во втором смысле в качестве перводвигателя. Из тех, кто перечислен в презентации это, например, Т. О'Коннор. Однако большинству философов такая онтология представляется слишком экстравагантной, в частности, не понятно, как именно такая свобода воли появилась в истории развития нашего вида, и не понятно как она взаимодействует с физическими процессами в мозге и других частях нашего тела. Большинство сторонников существования свободы во втором смысле сходятся на том, что достаточно, если человек будет ближайшей причиной своих действий. Человек может быть ближайшей причиной своих действий, даже если физикализм истинен. В том случае, если причиной человеческих поступков является мотивы, цели и ценности, присущие человеку, мы можем сказать, что сознательная интенция, которая формируется у человека на основании этих мотивов, целей, ценностей - это и есть причина действий человека, лежащая в нем самом. Такого взгляда придерживается целый ряд философов, которые в принципе являются инкомпатибилистами относительно первого тезиса. Например, Дж. Сёрль, А. Миле и Роберт Кени. Однако, все это хорошо, но может показаться, что некоторые данные современной нейронауки опровергают свободу во втором смысле. Большую известность

получил эксперимент Бенджамина Либета (1983). Этот эксперимент выглядел следующим образом. Испытуемых просили следить за перемещением красной точки на миллисекундном циферблате, их задача состояла в том, чтобы в какой-то момент, по произвольно возникшему желанию, сжать руку в кулак, и, что особенно важно – запомнить положение красной точки на циферблате в тот момент, когда у них возникло желание сжать кулак. Этот момент Либет именуется буквой W. Эксперимент построен на том, что как еще важно, в 65-х годах еще показали Кон Хуберт и Дэке, что за 800 миллисекунд до действия в мозге возникает, так называемый, потенциал готовности, то есть повышается электрическая активность моторных областей мозга, что можно измерить при помощи электроэнцефалограммы. Эксперимент Либета показал, что потенциал готовности предшествует действию, принимаемому как спонтанное в среднем на 55 мс. Но, самое интересное не в этом, а в том, что на основе этих данных Либет приходит к выводу, что испытуемые осознают желание сжать руку в кулак приблизительно за 200 миллисекунд до того, как они ее действительно сжимают. То есть, W наступает в среднем через 350 миллисекунд после того, как приборы фиксируют возникновение потенциала готовности, но за 200 миллисекунд до самого действия. То есть, значит, что нейроученые могут предсказать действие человека на 350 миллисекунд раньше, чем сам этот человек мог бы предсказать, что он сейчас почувствует желание осуществить это действие, которое ему кажется спонтанным, возникающим свободно. Последующие эксперименты других нейрофизиологов, например, эксперименты Сум показали, что временной интервал предсказания может быть расширен, на основании компьютерного анализа полученного при помощи ФМРТ данных об активности соответствующих зон коры головного мозга, движение испытуемого может быть предсказано за 10 секунд. Правда, всего с 60% точностью. В целом следует отметить, что на данный момент проведено большое количество эмпирических исследований возможности предсказания действий на основании фиксации предшествующей мозговой активности. В целом, можно заключить, что исследование Либета и других нейроученых показывают, как вместо предсказаний действий на языке мотивов, предсказывать действие на языке физических причин. Более того, именно физические причины кажутся истинными причинами, потому что они предшествуют мотивам по времени. Испытуемые осознают желание поднятия руки уже после того, как начался нейрофизиологический процесс, ведущий к поднятию руки. То есть, осознание мотивов исходит уже после того, как процесс движения запущен, отсюда некоторыми учеными и философами делается вывод, что мотив не играет каузально значимой роли, или, используя более сильные выражения, наше сознание не является причиной наших действий. Существует большое

количество литературы, критикующей экспериментальную парадигму Либета, как в связи с тем, что момент осознания решения очень сложно зафиксировать, так и в связи с тем, что условия этого эксперимента искусственны и отличаются от обычного процесса принятия решения. Однако, вне зависимости от того, верна ли эта критика, я бы хотела подчеркнуть следующее, такие эксперименты докажут, что наши мотивы не имеют каузальной силы только в том случае, если мы будем считать, что мотивы существуют только в тот момент, когда мы их осознаем. Если мотив существует только в момент переживаемого настоящего, тогда действительно, весь простор для принятия решения сужается до крошечного окна возможностей, о котором говорит Либет, что у нас якобы есть якобы 50 миллисекунд на то, чтобы сознательно наложить вето на этот процесс. Новейшие исследования нейрофизиологов, например, Лиза Филевич, опубликовавшая в 2013 г. статью, где она говорит, что если это единственная возможность, где мы можем наложить вето на наши действия, то ее тоже нет, поскольку из предыдущих наблюдений уже видно собираемся мы это сделать или нет. Но соответственно моя критика основывается не на конкретных соображениях, а на соображениях более общего порядка. А именно, на том, что мы по-другому должны интерпретировать мотивы, и в другой интерпретации мотивов, я предлагаю, ориентироваться на понятия, разработанные Гилбертом Райлом и Уилфредом Селларсом. Уилфред Селларс, разрабатывая идею диспозициональных свойств, выделил два типа ментальных состояний: диспозициональные и проявляющиеся. Диспозициональные свойства - это свойства, которые проявляются в подходящий момент. Так, например, диспозициональное свойство стекла - хрупкость. Проявляющиеся свойства - это свойства, которые проявляются сейчас. Хрупкость стекла станет проявляющимся свойством, если стеклянный стакан разобьется, упав на пол. Не всякое имеющееся у нас ментальное свойство должно быть проявляющимся. Так, например, я диспозиционально знаю русский язык, даже когда сплю без сновидений, и я диспозиционально знаю английский язык, даже когда говорю по-русски. В целом, проявляющиеся свойства моего сознания сейчас, это всё то, что я субъективно ощущаю, находясь в так называемом феноменальном поле моего сознания. Наши мотивы, как мне кажется, нередко являются диспозициональными свойствами. Они присущи нам, как черты характера, привычки и пристрастия, но не проявляются, пока не настанет подходящая ситуация. Поэтому, если мы будем считать мотивы диспозициональными свойствами, имеющими определенную физическую реализацию в мозге, выраженную, например, в специфической архитектуре нейронных сетей, сформировавшихся в результате опыта, накопленного конкретным человеком, то предсказания поведения человека на основе анализа активности его нейронных сетей не

будет опровержением того, что этот человек поступает в соответствии со своими мотивами. Конкретно по поводу экспериментов Либета можно предположить, что в этом случае так называемое желание поднять руку, зафиксированное в момент W , было просто проявлением диспозиционального мотива поднять руку в произвольный момент, сформированного на основании инструкции экспериментатора и вполне осознаваемого испытуемым. Вывод несколько футуристического характера, что когда-нибудь возможно нейроучёные смогут выявлять, чьи-то диспозициональные мотивы и предсказывать следующие за ними действия, на основании анализа активности нейронов. Подобные технологии, если они будут возможны, имеют огромный потенциал по изменению нашей жизни и в них будут большие этические последствия, именно поэтому нам важно понимать проблему. Проблема состоит не в том, что нейронаука докажет, что у нас нет свободы воли, и поэтому мы не действуем на основании мотивов и не несём ни за что моральной ответственности. Она состоит в том, что одни люди получают непосредственный доступ к диспозициональным мотивам других людей, и чтобы понять, что это значит, нужно тщательно продумать последствия. Спасибо за внимание!

Вопросы из зала:

(Слушатель) - У меня есть комментарий и возражение. Первый комментарий, мне кажется, что вы очень хорошо показали, почему философы не должны, так сказать плестись за нейроучёными, почему мы не должны смотреть им в рот, что там психологи, психиатры и нейроучёные открывают. Примеры с экспериментом Либета про качели, то признаются, то опровергаются и это очень хороший пример. Сначала после представления экспериментов была такая очень восторженная реакция на эти результаты, потом начали опровергать и он сам сменил свою позицию, потом, как вы отметили, и наложение вето можно предсказать. И я хочу указать на статью 2015 года в материалах национальной академии наук США, в этой статье показывается, что предыдущие исследования неверны. Это статья прошлого года, которая доказывает, что у нас есть возможность накладывать вето. То есть мы как философы, как должны реагировать на эти разногласия?

И собственно возражение, мне кажется, чтобы ваша аргументация по поводу понятия физикализма работала, необходим ещё один дополнительный ход. Физикализм утверждает, что существует только физические объекты, значит, агенты не существуют, а этого физикализм не утверждает.

(Мария) – Нет, значит, агенты являются физическими объектами. Во-первых, спасибо за ваш комментарий по статье, а во вторых, я полностью согласна с уточняющим вопросом, если нет противоречия. Физикалисты, кроме некоторых случаев, принимают компатибилизм, считая, что там свобода воли совместима с каузальной замкнутостью

физического мира, но их смущает аргумент от нейронаук, то есть их устраивает философская аргументация о том, почему это совместимо, потому что им кажется, что, если причины этого действия могут быть предсказаны до того, как сам человек осознаёт это действие, то причины не могут быть мотивами, потому что мотивы по существу являются тем, что я осознаю. И я как раз хочу показать, что это не является проблемой для компатибилизма.

(Слушатель) – Можем ли мы сказать, что в данном случае проблема свободы воли является проблемой инерции автомобиля, то есть у него есть водитель, он может что-то делать, но его нельзя развернуть моментально, то же самое как отношение неопределённостей, мы можем предсказать отношение частицы в каком-то диапазоне поневоле. То есть это значит, если мы поставим цель предсказать поведение человека с абсолютной точностью на год вперед, то это будет абсурд.

(Мария) – Предсказать поведение человека на год вперед, это значит предсказать поведение всей той части вселенной, в которой он будет находиться, потому что на наше поведение влияет всё. Никто не говорит о предсказании на год вперед, говорится о предсказании, допустим за 10 секунд.

(Слушатель) – Я думаю можно и больше предсказать, я предскажу, что жулик украдёт и через год и через два и через двадцать лет. Также человек высокой моральной ответственности, он защитит другого. Я очень благодарен вам за доклад, и за то, что вы включили понятие этики, а вопрос мой в следующем. Вы как то упустили момент о теории Узнадзе, психологических установках и предпосылках. Моральная ответственность у разных людей разная в одних и тех же условиях. Когда она возникает и почему она возникает, за что ответственен человек? Если я не управляю объектом и не могу его реализовать, то я не могу за него отвечать. Если я не влияю на него, то, как быть?

(Мария) – Есть конкретные теории моральной ответственности, которые говорят, что это такое, за какие поступки человек несёт ответственность, а за какие нет, это в целом предмет этики. А люди, которые пишут про проблемы свободы воли, говорят, что у них есть следующий формальный критерий, что как бы мы не трактовали моральную ответственность, она уже требует свободы, но если она уже требует свободы, и свобода не совместима с физикализмом, то получается, что физикализм не совместим с моральной ответственностью. Это и есть аргумент инкомпатибилистов, которые говорят, что мы тогда должны выбрать одно из двух, либо физикализм, либо моральную ответственность. Я попыталась высказать пару аргументов, которые говорят, что мы не сразу должны выбрать одно из двух.