Человеческое сознание и его носители: можем ли мы превратить себя в роботов?*

© 2021 г. В.В. Васильев

МГУ им. М.В. Ломоносова,
Москва, 119991, Ломоносовский пр-т, д. 27, к. 4.

E-mail: vadim.v.vasilyev@gmail.com

Поступила 15.02.2021

В статье рассматривается вопрос о носителях человеческого сознания и личности. Автор показывает, что представление о тождественном Я конструируется рефлексией над воспоминаниями, что истинность этого представления гарантируется непрерывным потоком перцепций, удерживаемых в воспоминаниях, и что поток перцепций предполагает наличие нормально функционирующего мозга как его носителя. Поэтому тождество личности оказывается зависимым от тождества мозга во времени. Попытка копирования структур сознания и личности на другие носители может, таким образом, приводить лишь к созданию двойников изначальной личности, но не к продолжению ее существования на другом носителе. Более перспективным путем трансформации биологических носителей человеческой личности автор считает постепенную замену их компонентов на искусственные аналоги. Чтобы оценить возможные последствия такой замены, автор анализирует аргументы Дж. Серла и Д. Чалмерса, призванные показать, соответственно, исчезновение сознания и личности при подобной замене и, наоборот, сохранение их в неизменном виде. Демонстрируется неубедительность аргументов Серла и зависимость аргументов Чалмерса от предпосылки, экспликация и обоснование которой возможны лишь в контексте спорных теорий тождества ментального и физического, эпифеноменализма или панпсихизма. Делается вывод о непредсказуемости изменений сознания при постепенной замене биологического субстрата.

Ключевые слова: сознание, личность, тождество личности, носители сознания, Д. Чалмерс, Дж. Серл, Р. Суинберн.

DOI: 10.21146/0042-8744-2021-9-105-117

Цитирование: *Васильев В.В.* Человеческое сознание и его носители: можем ли мы превратить себя в роботов? // Вопросы философии. 2021. № 9. С. 105–117.

* Исследование проведено при финансовой поддержке гранта Министерства науки и высшего образования РФ (проект «Новейшие тенденции развития наук о человеке и обществе в контексте процесса цифровизации и новых социальных проблем и угроз: междисциплинарный подход», соглашение № 075-15-2020-798).

Human Mind and Its Carriers: Is It Possible to Transform Ourselves into Robots?*

© 2021 Vadim V. Vasilyev

Lomonosov Moscow State University,
24/7, Lomonosovsky av., Moscow, 119991, Russian Federation.

E-mail: vadim.v.vasilyev@gmail.com

Received 15.02.2021

In this paper I discuss some aspects of the problem of carriers of human mind and person. The main emphasis is placed on the origin of our idea of the identical self in the stream of perceptions, the need for a physical carrier of our self and person, and on possibility of replacing the biological carriers of self and person with artificial analogues. I argue that the idea of identical self is constructed by reflection on memories, that its truth is guaranteed by continuous stream of perceptions kept in memories, and that the stream of perceptions presupposes the presence of a normally functioning brain, which can be considered as a carrier of our mind and person. Therefore, personal identity turns out to be dependent on the identity of the brain in time. An attempt to copy the structures of mind and person onto other possible carriers can thus only lead to creation of duplicates of the original person, but not to the continuation of its existence on another carrier. I argue that the gradual replacement of their components with artificial analogues is a more promising way of transforming the biological carriers of human person. To access the possible consequences of such a replacement I analyze arguments of John Searle and David Chalmers, designed to show, respectively, the disappearance of consciousness and person with such a replacement and, on the contrary, their preservation in a previous state. I explain why Searle's arguments are unconvincing, and demonstrate that Chalmers' arguments are based on a hidden premise, the confirmation of which is possible in the context of dubious theories of mind-body identity, epiphenomenalism or panpsychism only. I conclude that in the current situation it is impossible to predict which consequences for our person would follow such a replacement.

Keywords: consciousness, person, personal identity, carriers of consciousness, David Chalmers, John Searle, Richard Swinburne.

DOI: 10.21146/0042-8744-2021-9-105-117

Citation: Vasilyev, Vadim V. (2021) "Human Mind and Its Carriers: Is It Possible to Transform Ourselves into Robots?", *Voprosi Filosofii*, Vol. 9 (2021), pp. 105–117.

Стремительный технический прогресс, свидетелями которого мы все являемся, постоянно расширяет возможности человека. Мы стали лучше видеть, быстрее передвигаться и комфортно общаться на большом расстоянии. Но далеко не все биологические ограничения наших тел уже преодолены нами. Одним из главных ограничений остается относительно небольшая продолжительность человеческой жизни. Неудивительно, что в последнее время люди стали все больше размышлять на эту

* The study was carried out with the financial support of a grant from the Ministry of Science and Higher Education of the Russian Federation (project "Latest Trends in the Development of Human and Social Sciences in the Context of Digitalization and New Social Problems and Threats: An Interdisciplinary Approach", Agreement No. 075-15-2020-798).

тему. Попытаться снять это ограничение можно, к примеру, изменив генетические механизмы, отвечающие за старение человеческого тела. Но есть и более заманчивые решения. Нельзя ли, к примеру, заменить биологическую основу человеческого сознания на другие, более надежные носители? Такая замена позволила бы продлевать существование личности на неопределенно долгое время. Возможность подобной замены широко обсуждается трансгуманистами и давно пустила корни в художественной литературе, кинематографе, видеоиграх и т.п. Обсуждали эту тему и академические философы при рассмотрении традиционной проблемы тождества личности. Некоторые авторы допускают возможность переноса человеческого сознания и личности на искусственные носители, способствуя дальнейшему росту влияния подобных идей. В этой статье я собираюсь взвесить философские аргументы в пользу возможности такого переноса и попробую показать, что оптимизм в данном отношении оказывается явно преувеличенным.

1.

Начнем обсуждение проблемы переноса человеческой личности на искусственные носители с уточнения основных понятий и терминов и с небольшого феноменологического упражнения, важность которого можно подчеркнуть, вспомнив знаменитое признание Д. Юма, что, обращая взор на свои внутренние переживания, он не может найти среди них впечатления тождественного Я [Юм 1996, 297–298]. Представьте, таким образом, что вы пробуждаетесь от глубокого сна без сновидений. Вы начинаете осознавать свое окружение, вспоминаете о задачах, которые стоят перед вами сегодня и которые вы решили накануне и т.п. Вы восстанавливаете свое прошлое и привязываете его к настоящему и к возможному будущему. В ходе этой реконструкции выстраивается цепь представлений, каждое из которых содержит комплекс перцепций: чувственных данных, образов, эмоций, убеждений и желаний. Вспоминаемый вами вечер предыдущего дня, к примеру, может быть сочетанием образов бесед с близкими людьми и переживаний от очередной серии любимого фильма. Воображая такой ряд, вы пребываете в уверенности, что его компоненты некогда были реальными. Можно утверждать, что именно эта уверенность позволяет конструировать фразы типа «я слышал это», «я переживал это» и т.п. Подобные высказывания могут быть соотнесены с любым звеном упомянутого ряда. Перцептивные декорации меняются, но в них всегда присутствует это я, причем в одном и том же аспекте. В целом возникает картина, где смена перцепций контрастирует с неким неизменным Я. Такое Я – сознание, увязанное с компонентами потока перцепций – обычно и именуют личностью. Его неизменность во времени обозначается как тождество личности. Сознание, обращенное не на перцептивное содержание того или иного отрезка времени, а на сам этот его устойчивый аспект, называется самосознанием. Личность, таким образом, открывается в самосознании. Впрочем личность – это не только тождественное самосознающее Я, или, как его иногда обозначают, чистое Я, но и совокупность относительно устойчивых черт характера человека. Каждая такая черта содержит некий ментальный комплекс, составными частями которого являются связки различных диспозиций и желаний, убеждений и эмоций. Уникальные сочетания таких черт или качеств продуцируют специфическое для данной личности неслучайное поведение, подлежащее моральным и иным оценкам. Хотя упомянутые качества личности и устойчивы, но все же – в отличие от чистого Я – они допускают изменение. Поэтому мы не удивляемся, слыша, к примеру, фразу «я стал совсем другим за десять лет». В ней можно усмотреть отсылку как к неизменному чистому Я, так и к изменившимся относительно устойчивым чертам характера.

Хотя личность увязывается нами прежде всего с сознанием, а также с конгломератом приватных перцепций – эмоций, убеждений, желаний, ментальных образов и т.п., мы не можем отрицать ее связи и с перцепциями иного рода – зрительными, тактильными, слуховыми. Эти перцепции считаются нами публичными (непосредственно

доступными не только для нашего восприятия, но и для восприятия других людей), то есть трактуются нами как объекты и в совокупности образуют физическую реальность. Ее частью являются наши тела, от функционирования которых зависит не только состояние нашего сознания, но и характеристики наших приватных качеств. Большинство современных философов, впрочем, не ограничиваются тезисом о зависимости. Зависимость может быть слабой и не исключать самостоятельного существования зависимого. А они считают, что физическая основа необходима для существования личности. Это убеждение основано на их уверенности в неэффективности доводов в пользу субстанциальности сознающего субъекта, чистого Я, или совокупности ментальных качеств в целом. Соответствующие концепции стали терять авторитет уже в XVIII в., особенно после кантовской критики традиционной рациональной психологии, отстаивавшей такое видение ментальной жизни и допускавшей существование в ней субстанциального начала, души. Главным недостатком субстанциалистского подхода к ментальному можно признать онтологическую избыточность этого дуалистического учения, нарушение им оккамистского запрета на умножение сущностей без необходимости.

Отрицание ментальной субстанции, разумеется, не обязано приводить к отрицанию ментальной жизни. Но ее экономней рассматривать как порождение физической субстанции или ее модификаций, таких как человеческие тела или их органы. Так мы и приходим к представлению о необходимости физической основы ментальных качеств. И хотя британский философ Р. Суинберн и его сторонники пытаются в наши дни возродить субстанциальный дуализм [Swinburne 2009], им пока не удается убедить коллег в продуктивности своих усилий. Главный субстанциалистский аргумент Суинберна опирается на утверждение о несостоятельности альтернативной акциденталистской позиции. Суинберн считает, что без допущения субстанциальной души мы не сможем найти основание для решения вопроса о том, где оказалось бы наше Я после гипотетического рассечения мозга и распределения его полушарий по разным человеческим телам. При ближайшем рассмотрении, однако, выясняется, что в самом этом утверждении неявно допускается такая независимость ментального от физического, которая позволяет приписывать ментальному субстанциальность (Суинберн считает, что физические характеристики распределяемых по различным телам полушарий мозга не позволяют сказать, где окажется Я[1]). Если Суинберн предполагает то, что нужно доказать, то его аргумент не имеет реальной силы. Поэтому в дальнейшем я буду исходить из положения о необходимости наличия какого-то физического носителя ментальных состояний, сознания и человеческой личности в целом.

Отрицание субстанциальности чистого Я, между тем, подталкивало современных философов к поиску альтернативных объяснений сознания и его структур. Главной интуицией последних десятилетий стала компьютерная метафора ментального. Согласно этой метафоре, человеческий мозг можно рассматривать как аналог аппаратной части компьютера, а ментальное – как аналог ее программного наполнения. Аналогия ментального и программ объясняется функциональным сходством их ролей: как ментальное, так и программы опосредуют данные на входе соответствующих физических систем и данные на их выходе (к примеру, нажатие определенных клавиш на клавиатуре и появление определенных изображений на мониторе, с одной стороны, чувственные впечатления и поведенческие реакции – с другой).

Ментальное при таком функционалистском понимании естественно трактовать как набор функциональных схем, реализованных в структурах человеческого мозга. Самые фундаментальные из этих схем могут напрямую порождаться мозгом, другие же могут возникать при участии окружающей человека природной и культурной среды. Эта интуиция – отдающая, как мы видим, должное небезызвестному тезису об общественной или культурной природе ментального (или, в другой терминологии, идеального) – получила подкрепление в популярной ныне теории сознания как глобального рабочего пространства. Суть этой теории, предложенной когнитивным ученым Б. Баарсом [Baars 1997], в том, что сознаваемыми оказываются такие ментальные данности, которые, будучи продуктами специализированных модулей мозга, становятся

доступны для использования другими его модулями, словно бы попадая на доску, открытую для обозрения всем работникам офиса, или на сцену в театре. При таком подходе, кстати, явно или неявно различаются ментальные состояния как таковые и осознанные ментальные состояния. Это различение соотносится с другим известным различением, предложенным американским философом Н. Блоком, а именно различением феноменального сознания и сознания доступа [Block 1995]. Сознание доступа имеется именно тогда, когда некоторый ментальный контент становится доступен для широкого использования различными подсистемами организма.

Теория сознания как глобального рабочего пространства хорошо согласуется с обыденными представлениями о сознании. Скажем, если я сознаю какие-то из своих ощущений, то я могу рассказать об этом другим (то есть эти перцептивные данные доступны моему речевому модулю) или использовать информацию о них для планирования поведения. И данная теория имеет очевидно функционалистский характер (функция сознания – обеспечивать быструю передачу разного рода данных от одних модулей нашего мозга к другим). Подобные функции хотя бы в принципе доступны для программирования. Поскольку, таким образом, получается, что человеческое сознание (как в узком смысле сознания доступа, так и в широком смысле совокупности ментальных состояний) и впрямь имеет некое сродство с программами, то не выглядит нелепым утверждение, что оно – и увязанное с ним личностное начало – может быть, подобно программе, скопировано со своего естественного носителя, мозга, и перенесено на другой носитель.

Мы тем самым подходим к обсуждению, пожалуй, самой заманчивой возможности сохранения человеческой личности и продления человеческого существования. Может быть, и правда это не только смелая фантазия, вроде путешествий во времени, а вполне реальная перспектива недалекого будущего? Нельзя ли и в самом деле переносить наши Я на внешние носители?

2.

Чтобы разобраться в вопросе о возможности сохранения личности при копировании функциональных схем, реализуемых в мозге, на внешние носители, надо найти какую-то прочную основу, какое-то исходное положение, которое может хотя бы претендовать на общее согласие. Еще Дж. Локк показывал, что для разных предметов должны использоваться разные критерии их сохранения или тождества [Локк 1985, 382]. Но логично предположить, что между ними должно быть нечто общее. И это общее проще всего нащупать в наших обыденных представлениях о тождественности окружающих нас физических объектов. Может, правда, показаться, что такие абстрактные вопросы вообще не попадают в сферу обыденных дискурсов, но это не так. В действительности мы постоянно выносим суждения о тождественности предметов, хотя можем и не рефлексировать на этот счет. Когда мы приходим в знакомые места, мы, как правило, констатируем, что там присутствуют какие-то предметы, которые были там и во время нашего предыдущего их посещения. Чайник, который я вижу сегодня на кухне, это тот самый чайник, который я видел там вчера, и т.п. Но при каких условиях мы приписываем подобное нумерическое тождество такого рода предметам? Нетрудно ведь вообразить, что в тот промежуток времени, в который мы не наблюдали за этими предметами, они были заменены на их точные копии. Такие копии были бы качественно, но не нумерически тождественны исходным предметам, а мы обычно говорим именно о нумерическом тождестве.

Ясно, что мы редко можем полностью исключить подмену знакомых предметов из нашего окружения. Но в большинстве случаев мы считаем ее маловероятной. Нас, впрочем, сейчас интересуют не механизмы оценки вероятности подмены предметов, а то, что отсутствие такой подмены является условием нашего убеждения в тождественности физических вещей. Но что мы вкладываем в понятие подмены? Подмена имела бы место тогда, когда исходный предмет в какой-то момент времени перемещался

из своего обычного места в другое место или уничтожался, а его место занимал бы некий сходный с ним предмет. Отсутствие подмены в таком случае означает, что существующий в момент времени tn и месте pn предмет так соотносится с существовавшим в момент времени t1 и месте p1 предметом, что в любой момент времени между t1 и tn их пространственные координаты совпадали.

Из этой дефиниции вытекает алгоритм, позволяющий удостоверяться в нумерическом тождестве предмета. Таким алгоритмом является непрерывное наблюдение за предметом. В самом деле, при подобном наблюдении нам известны его пространственные координаты во все моменты времени от t1 до tn. Допустим теперь, что предмет, наблюдаемый нами при этих условиях в момент времени tn, нетождествен предмету, который мы наблюдали в t1. Это значит, что пространственные координаты предмета, который мы наблюдали в t1, должны были бы не совпадать с координатами предмета, наблюдаемыми в tn. Но мы знаем, что они совпадают с последними.

Из приведенных формулировок также следует, что даже небольшой перерыв в существовании предмета уничтожает его нумерическое тождество (оно может уничтожаться также при сильном изменении предмета, который больше не будет подпадать под то понятие, под которым мы его идентифицируем, но этот случай сводим к тому, который мы сейчас разбираем). В самом деле, представим, что предмет А непрерывно существовал в промежуток времени от t1 до t2, затем исчез на несколько мгновений. В момент времени t4 в том же месте появился качественно идентичный с А предмет В. Можно ли считать его нумерически тождественным А? Судя по всему, нет, так как нельзя утверждать, что во все моменты времени с t1 до t4 их пространственные координаты совпадали: в момент времени t3 у них вообще не было пространственных координат.

Посмотрим теперь, как можно применить эту схему определения тождества предметов к личности. Ключевым здесь наверняка тоже должно быть непрерывное существование. Надо, правда, уточнить, о непрерывном существовании чего именно мы вправе говорить в данном случае. Одна из трудностей связана с тем, что личность выше была увязана с сознающим началом. Но сознание точно не существует непрерывно на протяжении нашей жизни. Оно может исчезать в глубоком сне и в других состояниях. Это, однако, не мешает нам говорить о своем тождестве.

Для прояснения ситуации вернемся к нашему исходному примеру. Допустим, я сознаю, что какое-то время назад был свидетелем неких событий. Я убежден, что именно я воспринимал их, полагая тем самым, что мое Я сохранило свое тождество на этом временном отрезке. Но ведь мои воспоминания могут быть ложными. Более того, в принципе они могли бы в точности соответствовать реальным восприятиям другого человека. В таком случае говорить о тождестве моего Я на этом временном отрезке будет невозможно: восприятия тех событий не были моими. Этот сценарий будет исключен, если восприятия тех событий в соответствующий момент времени были составной частью единого потока перцепций, непрерывное течение которых продолжалось вплоть до моих нынешних восприятий. При ложных воспоминаниях последним не соответствовали реальные восприятия из данного конкретного потока перцепций, а поэтому здесь нет и тождественного Я в разные моменты времени. Осмысленность фразы о конкретном, уникальном потоке перцепций предполагает наличие в нем некого аналога места, где эти перцепции непрерывно сменяются. Если не принимать этого во внимание, то поток перцепций сведется исключительно к их временной или каузальной смежности, что неизбежно приведет к парадоксам, потому что тогда многие смежные по времени перцепции можно будет трактовать как части данного потока и личность будет допускать умножение, ср. [Parfit 1987].

Непрерывность потока перцепций (то есть наличие в каждый момент времени в данном «месте» каких-то перцепций) и удержание части появляющихся в этом потоке компонентов в последующие моменты времени создают, таким образом, объективные предпосылки для возникновения у нас идеи тождественного Я. Более того, она неизбежно возникает при наличии подобного потока перцепций и при доступности

воспоминаний о его фазах в смысле сознания доступа (чистое Я поэтому можно и не представлять в качестве какой-то особой инстанции или реальной сущности). И наоборот, хотя естественно возникающее при воспоминаниях о прежних событиях представление о тождестве Я во времени не гарантирует реального тождества нашего Я, или личности на данном временном отрезке, непрерывное существование внутренних перцепций, становящихся предметом подобных воспоминаний, или непрерывность самого потока, в котором встречались вспоминаемые нами перцепции, делает это представление истинным.

Подумаем теперь, при каких условиях существование этих внутренних перцепций, их комплексов или самого их потока будет непрерывным. Мы уже приняли, что они нуждаются в физическом носителе. Он не только порождает этот поток (как бы создавая ментальное место для смены приватных перцепций), структурирует и поддерживает его, но и индивидуализирует последний (конкретные потоки перцепций могут отличаться друг от друга именно в силу привязки к разным носителям). Поэтому залогом интересующей нас непрерывности оказывается непрерывное существование этого носителя. А к нему применимы обычные критерии тождества физических предметов. Разрушение этого носителя поэтому устраняет личность и ее тождество.

Важно, однако, не вкладывать в этот вывод больше того, что в нем содержится. А содержится в нем признание того, что человеческая личность должна утрачивать свое существование с разрушением тех частей ее тела, которым она была обязана поддержанием потока своих перцепций. Модальность этого утверждения, однако, не указывает на логическую необходимость подобного события. На этот момент иногда не обращают внимания и смешивают реальную, или физическую, необходимость исчезновения личности после распада тела или мозга (которую мы имеем право допустить) с логической необходимостью прекращения ее существования. Подобное смешение мешает адекватно интерпретировать то обстоятельство, что каждый из нас может отчетливо представить перемещение своего Я или личности в какое-то другое тело или мозг, не связанные с нынешними. Совпадение физической и логической необходимости означало бы, что сам факт подобных представлений опровергает тезис, что существование личности должно прекращаться с разрушением связанного с ней тела или мозга. Но поскольку мы можем или даже должны различать физическую необходимость (постоянство сочетаний событий, задаваемое контингентными законами природы) и логическую необходимость, предполагающую невозможность отчетливо мыслить положения дел, противоположные тем, в структуре которых мы ее допускаем, признание зависимости существования личности от непрерывного существования ее тела не противоречит нашей способности отчетливо представлять наше перемещение в другие тела. Хотя личность должна исчезать после разрушения тела в смысле физической необходимости, она совершенно не обязана исчезать после разрушения тела в смысле логической необходимости – именно потому, что мы можем представить ее сохранение.

Ситуация в данном случае аналогична той, когда мы размышляем о возможности людей летать в обычных земных условиях без использования каких-либо подручных средств. Мы можем отчетливо представлять, как мы летаем, но законы природы делают полеты невозможными. Мир мог бы оказаться таким, что мы могли бы летать. Так же и наша личность в иначе устроенном мире могла бы перемещаться в другие тела при разрушении изначальных тел. Но не в нашем.

Заметим также, что из сказанного следует, что в нашем мире личности не могут перемещаться в другие тела и при сохранении изначальных тел, ведь иначе мы столкнулись бы с противоречием: существование прежних личностей в новых телах зависело бы от того, сохраняются ли старые, которые, однако, уже не могли бы естественно поддерживать существование этих личностей.

В философской литературе по проблеме тождества личности идет давний спор о его критериях, биологические они или психологические, см. [Логинов и др. 2018]. Мы видим, однако, что их противопоставление необязательно. Главным критерием

является психическая непрерывность, непрерывность потока перцепций, но поскольку она зависит от непрерывности биологического субстрата, то правильный подход стоит называть скорее психо-биологическим или психо-субстратным, если биологический носитель личности может постепенно превращаться во что-то иное: попытки говорить о реальной возможности психической непрерывности при телепортации, когда создаются психо-физические клоны наших личностей [Parfit 1987], обречены на провал.

Для уточнения специфики этой позиции сравним ее со сходной концепцией М.А. Секацкой [Секацкая 2018]. Секацкая рисует красивую картину: проблемы биологического и психологического подходов к решению проблемы тождества личности заставили философов искать экстравагантные альтернативы, такие как нарративная теория личности [Dennett 1991; Волков 2018], но альтернатива, с ее точки зрения, может быть и более реалистичной. Речь идет о «психофизической» теории, признающей необходимость как физической, так и психической непрерывности. А их сочетания достаточно для тождества личности во времени; ср. также [Garrett 1998, 56]. Проблема этой концепции не только в том, что сочетания указанных факторов, по-видимому, все же недостаточно для тождества личности (для этого к непрерывности существования мозга и потока перцепций надо добавить еще и сознательную рефлексию над этим потоком; без такой рефлексии, то есть без самосознания мы были бы не личностями, а некими «мы», к которым, возможно, лучше было бы применять другие критерии тождества, ср. [Olson 1997]), но и в том, что здесь тоже не дифференцируются логические и физические условия. Однако без такой дифференциации можно будет, к примеру, утверждать, что физическая непрерывность не необходима для тождества личности, так как можно отчетливо помыслить продолжение существования наших личностей в других мозгах, которые никак не связаны с их изначальными мозгами и телами, а значит, и допустить его возможность. В частности, я могу представить, как мои ощущения постепенно трансформируются в ощущения другого человека с другой локализацией (в стандартных мысленных экспериментах на эту тему испытуемый засыпает и просыпается в другом теле, но допущение перерыва в восприятиях необязательно и снижает очевидность выводов). Чтобы избежать подобных затруднений, надо непременно добавлять, что эта возможность чисто логическая. Впрочем, даже такой возможности хватает для признания логической необходимости психической непрерывности для тождества личности. Но она должна быть дополнена физической необходимостью биологической или субстратной непрерывности.

3.

Итак, перенос человеческой личности на другие носители физически невозможно осуществить путем копирования функциональных схем мозга и инсталляции их на другие устройства. Даже если допустить техническую возможность такой процедуры, на новом устройстве будет воспроизведена лишь копия: оригинальная личность останется на прежнем носителе. Однако у нас остается еще одна возможность замены носителя: постепенная трансформация нашего организма, при которой его биологические компоненты будут замещаться чем-то другим.

Заметим, что сама по себе трансформация носителя человеческой личности точно не препятствует сохранению личности: такая трансформация постоянно происходит на клеточном и молекулярном уровнях существования мозга и организма в целом. Но нас интересует случай замены биологических компонентов на их искусственные аналоги. Нет ли здесь какой-то специфики?

В последние десятилетия философы многократно обсуждали подобные сценарии. По итогам этих обсуждений можно зафиксировать две принципиальные позиции. Одна из них связана с аргументами американского философа Дж. Серла, другая – с доводами австралийца Д. Чалмерса.

Суть первой позиции в том, что постепенная замена компонентов человеческого мозга на искусственные аналоги приведет к угасанию сознания и в конечном счете

к исчезновению личности. Сторонники второй позиции, напротив, доказывают, что ничего подобного не случится.

Главным аргументом в пользу первой позиции являются соображения, основанные на знаменитом мысленном эксперименте Дж. Серла «Китайская комната» [Searle 1980]. С его помощью Серл пытался решить вопрос, обладают ли компьютеры, симулирующие человеческое поведение, специфическими модусами человеческого сознания, такими как понимание. Для разрешения этой загадки Серл предлагал встать на место компьютера, в котором инсталлированы соответствующие программы, и посмотреть, как это будет выглядеть изнутри. К примеру, мы видим компьютер, умеющий осмысленно отвечать на вопросы на китайском языке. Внешним наблюдателям кажется, что он понимает по-китайски. Чтобы проверить это, выучим его программу, то есть сами станем такими компьютерами. Серл был убежден, что если мы не знали китайский и если программа сводится к правилам формальных преобразований символов, то даже после изучения этой программы мы не будем понимать по-китайски, хотя наблюдателям со стороны будет казаться, что мы понимаем этот язык, так как даем осмысленные ответы на их вопросы.

Если мы, став компьютерами, не обретаем понимания, то его – а также других специфически человеческих аспектов сознания – наверняка лишены и сами компьютеры. Но почему у нас есть сознание, субъективные ментальные состояния, а у компьютеров – нет? Чтобы ответить на этот вопрос, мы должны задуматься о том, что, собственно, отличает нас от компьютеров. Поведенческие схемы, как мы предположили, могут совпадать. Остаются лишь материальные отличия (которые, правда, отсутствуют в «Китайской комнате», где мы сами становимся компьютерами, что делает всю эту линию рассуждений несколько парадоксальной). Наши организмы имеют белковую природу, а компьютеры состоят из других веществ. Значит, именно белковая природа наших нейронных компьютеров отвечает за наше сознание.

Теперь ясно, почему аргументы Серла можно использовать для обоснования тезиса о том, что постепенная замена биологических компонентов нашего мозга на искусственные аналоги должна приводить к угасанию сознания. Ведь именно эти компоненты отвечают за его присутствие в нас.

«Китайская комната» Серла уже более сорока лет вызывает оживленные споры. Многих убеждали его рассуждения, но, конечно, далеко не всех. На мой взгляд [Васильев 2014, 128–134], проще всего критиковать этот мысленный эксперимент, указав на то, что в описанной Серлом ситуации человек-компьютер в реальности не всегда сможет давать осмысленные ответы на обычные вопросы, хотя это предполагается в данном эксперименте. Ведь программисты не смогли бы заранее исключительно формальными, синтаксическими средствами прописать ответы на фактические вопросы, которые мы можем задавать человеку, взявшему на себя роль компьютера. Представим, что мы спрашиваем его, что находится перед ним или сколько сейчас времени. Чтобы осмысленно ответить на такие вопросы, участник эксперимента должен будет соотносить те или иные китайские иероглифы с предметами или процессами из своего окружения. Скажем, если он видит на терминале поступивший извне на непонятных для него иероглифах вопрос о том, что находится перед ним на столе, то в программе должно быть прописано, что, увидев подобные иероглифы, он должен посмотреть на стол, и если, к примеру, там стоит банка кока-колы, то он должен выдать в качестве ответа какие-то другие определенные иероглифы. Но, делая это, он начнет понимать значение данных иероглифов, понимать по-китайски в специфически человеческом семантическом смысле. И это разрушает логику мысленного эксперимента Серла и лишает его эффективности.

Если эта или иная критика «Китайской комнаты» достигает своей цели, то вопрос о сохранении сознания и личности при постепенной замене биологических компонентов на их искусственные аналоги остается открытым. Впрочем, Д. Чалмерс попытался закрыть его с другой стороны. Он выдвинул аргумент, призванный доказать, что указанная замена не только не приведет к угасанию сознания, но и не повлечет за собой изменений в самих сознательных состояниях.

Аргумент Чалмерса, как и аргумент Серла, базировался на мысленных экспериментах [Чалмерс 2013, 309–342]. В частности, он предложил представить, что замена биологических компонентов мозга на их искусственные функциональные изоморфы происходит таким образом, что при желании мы можем вернуть старые нейронные компоненты – и можем переключаться между ними. И Чалмерс утверждал, что если предположить, что при таких переключениях возникнет эффект «скачущих квалиа», то есть субъективных переживаний, связанных с физическими процессами в мозге, (а этот эффект должен возникать, если исходить из того, что рассматриваемая нами замена ослабляет или меняет сознательные состояния), то он будет сильнейшим образом контрастировать с неспособностью испытуемого заметить эти изменения и рассказать о них. В самом деле, поведенческие схемы при подобной замене, согласно предположению, остаются неизменными, а если бы субъект заметил скачки квалиа и рассказал о них, это было бы не так.

Кажущаяся нелепость описанной ситуации заставляет, по Чалмерсу, отвергнуть ее реальность и признать наиболее вероятным другой сценарий, при котором в наших субъективных состояниях не будет происходить изменений при замене биологических компонентов мозга на их искусственные аналоги. Этот аргумент был выдвинут Чалмерсом уже около четверти века назад, но, судя по рукописи его будущей книги, доступной автору, он сохранил веру в него и в наши дни.

Если Чалмерс прав, то перспектива постепенной замены нашего мозга на его искусственный аналог при сохранении исходного сознания и личности выглядит вполне реальной. Если он прав, мы действительно можем постепенно превратить себя в роботов без утраты самих себя.

4.

Оценивая аргумент Чалмерса, стоит обратить внимание на один неявный момент. Чалмерс рассуждает о замене биологических компонентов мозга, то есть прежде всего нейронов на их искусственные изоморфы. При этом предполагается, что эти изоморфы функционально идентичны нейронам. По итогам этой постепенной замены искусственный мозг будет функционально идентичен изначальному биологическому мозгу и продуцировать такое же поведение субъекта, как и тот. Задумаемся теперь о том, как можно добиваться такой функциональной идентичности. И прежде всего надо ответить на один важный вопрос: вносит ли вклад в поведение человека его сознание?

Если сознание вносит вклад в поведение, добавочный к вкладу сугубо физических факторов, то поведенческие функциональные схемы человека определяются в том числе некими ментальными параметрами. Должны ли мы учитывать эти параметры при замене биологических компонентов мозга на их искусственные аналоги? Проще всего было бы учесть их. Но если допустить, что они будут реализованы на физическом уровне, то не станут ли сами ментальные параметры и сознание избыточными? А если они избыточны, убедительными ли будут рассуждения об их существовании в новой искусственной системе? Похоже, проще предположить, что в ней вообще не будет сознания. Конечно, искусственный мозг будет порождать прежнее поведение, и новый субъект будет, как и раньше, рассуждать о наличии у него внутреннего опыта. Но это не должно вызывать удивления, поскольку мы сами встали на путь зомбификации нашего испытуемого: философские зомби должны вести себя именно так, поэтому тут не будет ничего неожиданного.

Казалось бы, можно вычесть каузальный вклад ментальных состояний и ограничиться созданием таких искусственных аналогов нейронных процессов, которые будут воспроизводить функциональные схемы полученного в результате этого вычитания физического остатка. Но откуда мы знаем, что такие искусственные заменители будут генерировать ментальные состояния, нужные для возникновения функциональных схем, идентичных изначальным? Ведь ментальные состояния не существуют сами по себе, и мы исходим из того, что они порождаются мозгом. Как доказать, что искус-

ственный мозг будет порождать такие состояния, порождать сознание? Важно, что при только что уточненных нами предпосылках в пользу такого порождения нельзя аргументировать прежним чалмерсовским способом. Нельзя, к примеру, утверждать, что в случае фрагментарного отсутствия сознания при замене каких-то частей мозга их искусственными аналогами возникала бы нелепая ситуация, когда человек вел бы себя так, будто никаких изменений в сознании у него не произошло. Ведь в новых условиях не предполагается, что искусственные заместители нейронов сами по себе полностью дублируют прежние поведенческие схемы испытуемого. Поэтому при отсутствии или изменении сознания такой человек мог бы вести себя совершенно иначе.

Мы видим, таким образом, что если сознание влияет на поведение, то, похоже, нет гарантии, что замена нейронов на искусственные компоненты не приводила бы к значительным изменениям в человеческом сознании или даже к исчезновению сознания и самой человеческой личности.

Ситуация выглядела бы более благополучной, если бы сознание не влияло на поведение или если бы оно влияло на поведение просто в силу онтологической тождественности сознания и физических процессов в мозге. Тогда поведенческие схемы могли быть без проблем продублированы искусственными аналогами (в случае тождества такое дублирование было бы возможным в силу известного принципа множественной реализации: функциональные схемы, тождественные каким-то структурам в биологическом мозге, могли бы быть адекватно реализованы и на каком-то другом материале) и аргумент Чалмерса выглядел бы более правдоподобным. Но можно ли принять тезис тождества ментального и физического или эпифеноменалистский тезис о бездеятельности сознания? В философской карьере Чалмерса были моменты, когда он сближался с эпифеноменализмом. Чаще он, впрочем, искал пути его преодоления. И в самом деле, каузальную значимость сознания можно отстоять, даже отрицая тезис о тождественности ментального и физического и признавая, что функциональные схемы нашего поведения включают лишь физические компоненты: для этого надо показать, что сознание является условием их существования. Главной опцией при таком подходе оказываются разновидности современного панпсихизма. Сторонники этой позиции обычно говорят, что физическая реальность мыслится нами как система отношений, но отношения предполагают то, что соотносится, а поскольку помимо физического нам известно только ментальное, то оно вполне может быть своего рода субстратом физической реальности.

Обратим внимание, к какому странному выводу мы пришли: если постепенная замена биологических компонентов нашего мозга на их искусственные аналоги не приводит к драматическим изменениям нашего сознания и личности, то теория тождества ментального и физического, эпифеноменализм или панпсихизм должны быть признаны самыми перспективными теориями сознания. Думаю, однако, что признать их таковыми очень непросто. Каждая из них имеет большие проблемы, и эти проблемы в тех или иных вариациях хорошо известны современным философам. 1) Теория тождества ментального и физического сталкивается с проблемой осмысленности ее центрального тезиса. Обычно, когда мы говорим о тождестве, мы четко понимаем, что имеем в виду и как убедиться в истинности сказанного. Но совершенно непонятно, как убедиться в тождестве ментального и физического, если только, конечно, не смешивать тождество с корреляцией[2]. 2) Эпифеноменализм кажется неэлегантной концепцией, плохо сочетающейся с нашей уверенностью в существовании сознаний у других людей: если бы эпифеноменализм был верен, не было бы никакой необходимости допускать реальность таких сознаний: они были бы образцовыми избыточными сущностями. 3) Что же касается панпсихизма, то он не только еще хуже, чем эпифеноменализм, согласуется со здравым смыслом, но и плохо стыкуется с общепринятым положением о том, что сознание является порождением мозга. Добавлю, что сами тяготеющие к панпсихизму философы считают очень опасной для него так называемую «проблему комбинации»: «…как переживания, присущие таким фундаментальным физическим сущностям, как кварки и фотоны, комбинируются для порождения той привычной разновидности человеческого сознания, которую мы знаем и любим?» [Chalmers 2017, 179].

Проблемы теории тождества, эпифеноменализма и панпсихизма настолько серьезны, что едва ли можно опираться на эти концепции при рассуждениях о личности. Но вместе с ними исчезают и надежды на то, что постепенная замена биологических компонентов мозга на их искусственные аналоги не будет приводить к исчезновению или к резкой трансформации сознания и личности. Чтобы таких драматических изменений не происходило, искусственные аналоги должны обладать порождающими возможностями изначальных биологических компонентов мозга. В общем, личность будет наверняка сохраняться, если новые компоненты будут неотличимы от старых.

Можно сказать, что мы знали это с самого начала, но задачей данной статьи было показать, что мы пока не можем далеко уйти от этого нехитрого изначального знания. Даже если постепенная трансформация нашего мозга в электронное устройство и нас самих в роботов не уничтожит нашу личность, она может изменить ее так, что мы с трудом будем узнавать в этой личности самих себя.

Примечания

[1] В реальности такое предсказание вполне можно сделать: Я продолжит существование в том теле, в которое попадет левое полушарие, обычно отвечающее за язык и рефлексию. А если допустить гипотетическую ситуацию с одинаковостью полушарий, то и Я изначально было бы не одно, а два.

[2] Смешение тождества с корреляцией, впрочем, допускал один из создателей теории тождества, У. Плейс [Place 1956]. Этим обстоятельством отчасти можно объяснить само возникновение данной теории.

Источники и переводы – Primary Sources and Translations

Локк 1985 – *Локк Дж.* Опыт о человеческом разумении. Кн. 1–3 // Сочинения. В 3 т. Т. 1. М.: Мысль, 1985. С. 77–582 (Locke, John, *An Essay Concerning Human Understanding*, Russian Translation).

Юм 1996 – *Юм Д.* Трактат о человеческой природе // Сочинения. В 2 т. Т. 1. М.: Мысль, 1996. С. 53–655 (Hume, David, *A Treatise of Human Nature*, Russian Translation).

Primary Sources

Dennett, Daniel C. (1991) *Consciousness Explained*, Back Bay Books, New York.

Parfit, Derek (1987) *Reasons and Persons,* Clarendon Press, Oxford.

Place, Ullin T. (1956) “Is Consciousness a Brain Process?”, *British Journal of Psychology*, Vol. 47, pp. 44–50.

Searle, John R. (1980) “Minds, Brains, and Programs”, *Behavioral and Brain Sciences*, Vol. 3, pp. 417–457.

Ссылки – References in Russian

Васильев 2014 – *Васильев В.В.* Сознание и вещи: Очерк феноменалистической онтологии. М.: URSS, 2014.

Волков 2018 – *Волков Д.Б.* Свобода воли: иллюзия или возможность. М.: Карьера Пресс, 2018.

Логинов и др. 2018 – *Логинов Е.В., Мерцалов А.В., Салин А.С., Чугайнова Ю.И., Юнусов А.Т.* Пролегомены к проблеме тождества личности // Финиковый компот. 2018. № 13. С. 6–40.

Секацкая 2018 – *Секацкая М.А.* Необходимые и достаточные критерии тождества личности // Вопросы философии. 2018. № 5. С. 125–133.

Чалмерс 2013 – *Чалмерс Д.* Сознающий ум: В поисках фундаментальной теории. М.: УРСС, 2013.

References

Baars, Bernard (1997) *In the Theater of Consciousness: The Workspace of the Mind*, Oxford University Press, New York.

Block, Ned (1995) “On a Confusion about a Function of Consciousness”, *Behavioral and Brain Sciences*, Vol. 18, pp. 227–287.

Chalmers, David J. (1996) *The Conscious Mind: In Search of a Fundamental Theory*, Oxford University Press, Oxford (Russian Translation).

Chalmers, David J. (2017) "The Combination Problem for Panpsychism", Brüntrup, Godehard, Jaskolla, Ludwig, eds., *Panpsychism: Contemporary Perspectives*, Oxford University Press, Oxford, pp. 179–214.

Garrett, Brian (1998) *Personal Identity and Self-Consciousness*, Routledge, London.

Loginov, Evgeny V., et al. (2018) "Prolegomena to the Problem of Personal Identity", *Finikovyj Kompot*, Vol. 13, pp. 6–40 (in Russian).

Olson, Eric T. (1997) *The Human Animal: Personal Identity Without Psychology*, Oxford University Press, Oxford.

Sekatskaya, Maria A. (2018) "Necessary and Sufficient Criteria of Personal Identity", *Voprosy Filosofii*, Vol. 5, pp. 125–133 (in Russian).

Swinburne, Richard (2009) "Substance Dualism", *Faith and Philosophy*, Vol. 5, pp. 501–513.

Vasilyev, Vadim V. (2014) *Consciousness and Things,* URSS, Moscow (in Russian).

Volkov, Dmitry B. (2018) *Free Will: Illusion or Possibility*, Kar'era Press, Moscow (in Russian).

Сведения об авторе

ВАСИЛЬЕВ Вадим Валерьевич – член-корреспондент Российской академии наук, доктор философских наук, заведующий кафедрой истории зарубежной философии философского факультета МГУ им. Ломоносова.

Author's Information

VASILYEV Vadim V. – Correspondent member of the Russian Academy of Sciences, DSc. in Philosophy, Head of the Department of the History of World Philosophy, Lomonosov Moscow State University.