Философский журнал
2024. Т. 17. № 2. С. 92–109
УДК 165.12

The Philosophy Journal
2024, Vol. 17, No. 2, pp. 92–109
DOI 10.21146/2072-0726-2024-17-2-92-109

ДИСКУССИИ

Может ли сознание быть исчерпывающе исследовано методами эмпирических наук?

В начале 1990-х гг. Ф. Крик и К. Кох ввели понятие нейронных коррелятов сознания – комплексов нейронных контуров, минимально необходимых для появления тех или иных сознательных функций. Их исследование стало возможным благодаря быстрому развитию технологий нейровизуализации. Однако прогресс исследований тормозится в том числе различным пониманием предмета/предметов исследования. Идет ли речь о состояниях сознания или его содержании? Растут ли «сознание доступа» (Нед Блок) и феноменальное сознание из одного корня или это различные способности, случайно терминологически объединенные? Возможно ли исчерпывающее алгоритмическое описание сознательных функций? Каковы критерии релевантности научно-эмпирических теорий сознания? Дискуссию открывает статья И.Ф. Михайлова «Предметы и методы эмпирических исследований сознания», в которой дается обзор этих и некоторых других концептуальных проблем, а также актуальных и возможных подходов к их решению. В обсуждении участвуют П.Н. Барышников, С.Ю. Бородай, В.В. Васильев, М.А. Сущин.

И.Ф. Михайлов

ПРЕДМЕТЫ И МЕТОДЫ ЭМПИРИЧЕСКИХ ИССЛЕДОВАНИЙ СОЗНАНИЯ*

Михайлов Игорь Феликсович – доктор философских наук, ведущий научный сотрудник. Институт философии РАН. Российская Федерация, 109240, г. Москва, ул. Гончарная, д. 12, стр. 1; e-mail: ifmikhailov@gmail.com

Попытки создания эмпирически обоснованных теорий сознания сталкиваются с препятствиями двоякого рода. Во-первых, господствующая стратегия, основанная на поиске нейронных коррелятов сознания, не приносит успеха из-за отсутствия

* Исследование выполнено за счет гранта Российского научного фонда (проект № 24-28-00804), https://rscf.ru/project/24-28-00804/

работающих гипотез об их каузальной связи с сознательными состояниями. Во-вторых, препятствием выступает также и множественность explananda – отсутствие достаточных доводов в пользу убеждения, что все, что мы считаем феноменами сознания или сознательными состояниями, действительно обладает онтологическим единством. Возможно, эти проблемы порождены отсутствием в науках о сознании некоторого аналога радиоинженерного уровня, необходимого, помимо теоретической электродинамики, для того, чтобы разобраться с принципами функционирования радиоустройств. Этот уровень знания должен включать упрощенную онтологию предметной области, позволяющую вычленить принципиальные функциональные взаимосвязи на их алгоритмическом уровне. Учитывая историю наук о сознании и специфику их предмета, оптимальным кандидатом на роль инженерного уровня знания мог бы оказаться вычислительный подход, предполагающий описание предмета как комбинации вычислительных примитивов, которые позволяют реализовывать алгоритмы, порождающие сознательные состояния. Желательность такого подхода подтверждается тем фактом, что не-вычислительные теории сознания, основанные на традиционном естествознании, парадоксальным образом остаются де-факто метафизическими (спекулятивными). Помимо различных подходов к критериям «хорошей» эмпирической (не-спекулятивной) теории сознания, в статье дается обзор теории сознания, основанной на применении гипотезы активного вывода. На основании анализа этой теории высказывается предположение, что вычислительная модель, лежащая в основе хорошей теории сознания, должна быть не детерминированной, а вероятностной.

Ключевые слова: сознание, нейронные корреляты, интегрированная информация, предиктивный процессинг, активный вывод

Для цитирования: *Михайлов И.Ф.* Предметы и методы эмпирических исследований сознания // Философский журнал / Philosophy Journal. 2024. Т. 17. № 2. С. 92–109.

Постановка проблемы

Я хотел бы начать с терминологической проблемы. Когда мы читаем английские тексты на тему сознания и обозначаем интересующие нас сущности английскими терминами, у нас не возникает концептуальных проблем, потому что слову *Matter* мы противопоставляем слово *Mind*, а тому, что мы понимаем под этим словом, на мой взгляд, адекватного аналога в русском языке нет. Далее, мы можем делить понятие *Mind* по разным логическим основаниям, скажем, по функции – на когнитивное и аффективное, по содержанию – на интенциональное и квалитативное, по способу доступа – на сознательное и бессознательное. Но, как только мы переходим на русский язык, мы видим, что «материи» противопоставляется «сознание», и, говоря о когнитивном доступе, мы также по признаку его присутствия/отсутствия логически делим понятия на «сознательное» и «бессознательное». Получается, что сознание в широком смысле – там где по-английски используется слово *Mind* – включает в себя также и бессознательное, что, конечно, ведет к противоречию.

Теперь о том, что далее здесь будет пониматься под «эмпирическим». Под этим термином я подразумеваю весь комплекс теорий, которые эмпирически фальсифицируемы или, в более общем плане, где существуют тесные взаимосвязи или корреляция между теоретическими утверждениями и данными опыта. Под «концептуальным» я понимаю то, что присутствует в этих теориях как *схема* предметной области или *таксономия* объектов, полагаемых существующими, или как *модель* для интерпретации теории,

ее *онтология*. На мой взгляд, это все синонимы, по крайней мере частичные. В данном контексте под концептуальной стороной нашего предмета я понимаю онтологию сознания, к вопросам которой относятся, например, следующие: может ли сознание быть «подлинным» логическим субъектом в истинных высказываниях (*ουσία*, в терминологии Аристотеля) или оно только приписывается подлинным сущностям (*κατηγορία*[1])? В последнем случае высказывания типа «сознание конституирует предмет» содержит категориальную ошибку, если мы считаем, что оно не является «первой сущностью». Другим вопросом может быть: является ли сознание чем-то единым – сущностью, атрибутом, акциденцией – или мы одним словом обозначаем разные явления? Последняя точка зрения достаточно популярна. Еще один онтологический вопрос: достаточно ли для хорошей теории сознания онтологий естественных наук? Положительный ответ на него открывает дорогу редукционизму.

Для собственно эмпирических теорий характерны вопросы: какова каузальная структура сознания, каковы необходимые и достаточные условия его проявления, как можно его идентифицировать и измерить, с помощью каких единиц и способов измерений?

В случае с этим уникальным предметом существенные эвристические преимущества принес бы еще один уровень, который в англоязычных публикациях традиционно называется *computational* – вычислительный[2]. В ведении этого уровня находятся вопросы: каковы формы и элементарные шаги – иногда называемые вычислительными примитивами – процессов, порождающих сознательные состояния; как структурировано пространство состояний сознательной системы; описываются ли указанные параметры рациональными – и тогда они дискретные – или действительными числами – тогда они континуальные; применяются ли детерминистские или вероятностные правила преобразования данных; как кодируются входные данные; как интерпретируются выходные данные и т.п.

Важно то, что вычислительный уровень, на мой взгляд, обнаруживает черты, которые роднят его как с концептуальным уровнем науки, так и с ее эмпирическим уровнем. С концептуальным уровнем его роднит тот факт, что он создает некую схематическую реальность, в каком-то смысле альтернативную перцептивной картине мира. Здесь напрашивается аналогия со схематизмом чистых рассудочных понятий у Канта[3]. С эмпирическим уровнем науки его роднит то обстоятельство, что вычислительные модели принципиально фальсифицируемы: вычислительная модель или работает, или не работает. А если работает, то или так, как предсказывала гипотеза, или не так.

Если вычислительное описание сознания в принципе возможно, то задача построения научной теории сознания значительно упрощается, а компьютерное моделирование сознательных процессов из имитации превращается в собственно моделирование. В литературе достаточно часто

---

1 Интересно, что исторически это слово означало ложное публичное обвинение, приписывание кому-либо чего-то такого, что ему на самом деле не свойственно (от κάτω – вниз и αγορά – народное собрание).

2 Хотя со словом «вычислительный» есть еще бóльшая этимологическая и концептуальная путаница, чем со словом «сознание». Подробнее об этом см.: *Михайлов И.Ф.* Социальные вычисления и происхождение моральных норм // Философский журнал / Philosophy Journal. 2022. Т. 15. № 1. P. 51–68.

3 *Кант И.* Собрание сочинений: в 8 т. Т. 3: Критика чистого разума. М., 1994. С. 156–162.

встречается слово «имитация», поскольку авторы стараются не делать каких-либо реалистических выводов. Моделирование в строгом смысле слова предполагает некий структурный и функциональный изоморфизм с оригиналом и обладает эвристическим потенциалом. Кроме того, абстрактная вычислительная модель как объяснительная схема может быть экстраполирована на другие предметные области. Наконец, делается следующий шаг в прогрессе технологий искусственного интеллекта, потому что появляется возможность создания сознательных устройств.

Инженерия сознания

Я бы уподобил соотношение теоретического и вычислительного в научных теориях, в частности в возможных теориях сознания, соотношению электродинамики и схемотехники: можно быть большим специалистом в области электродинамики, но не быть в состоянии построить радиоприемник или объяснить, как он работает. Для этого нужно овладеть еще одним языком – языком радиосхем: понимать, что такое резистор, конденсатор, где они используются, каковы их функции и т.п.

Наш бывший соотечественник Юрий Лазебник, который уже довольно давно является американским биологом, написал в 2002 г. очень остроумную статью «Может ли биолог починить радио»[4]. В статье представлены фотографии радиоприемника, как оказалось, сломанного, который его жена вывезла из Советского Союза. Когда Лазебник раскрыл его, он увидел там некие сгоревшие части и подумал, что, адресуясь к этой проблеме как биолог, он вряд ли может что-либо сделать с приемником: он видит перед собой какие-то группы элементов, окрашенных в различные цвета, но с непонятными каузальными связями. Рассуждая как биолог, он будет предполагать, что те из них, которые окрашены в один цвет, выполняют скорее всего одну определенную функцию, окрашенные в другой цвет выполняют другую, и, таким образом, ему придется очень долго искать какие-то возможные случайные корреляции, порождать гипотезы относительно этих корреляций, которые, вероятнее всего, не будут совпадать с реальностью. И поскольку статья была написана по поводу некоей дискуссионной теоретической проблемы в биологии, он делает вывод, что именно такого – схемотехнического или инженерного – уровня не хватает этой науке. Иначе говоря, недостает некоторого языка, описывающего примитивы или функциональные элементы, из которых состоят живые системы.

Аналогичным образом, мы можем сказать, что современные теоретические подходы к сознанию можно разделить по этому признаку на вычислительные и невычислительные. К невычислительным можно отнести популярные квантовые теории сознания, а также различные биологические теории, которые у нас достаточно популярны (мне случалось писать о так называемой теории биологического мозга или школы Коштоянца)[5]. К потенциально

---

[4] *Lazebnik Y.* Can a biologist fix a radio? – Or, what I learned while studying apoptosis // Cancer Cell. 2002. Vol. 2. No. 3. P. 179–182.

[5] *Михайлов И.Ф.* Человеческий мозг и сознание: биология или вычисления? // Философские проблемы информационных технологий и киберпространства. 2018. Т. 15. № 2. С. 92–110.

вычислительным теориям можно отнести рекурсивную теорию сознания[6], теорию сознания в рамках предиктивного процессинга (она же теория активного вывода), теорию интегрированной информации и некоторые другие. По моему экспертному мнению, которое может быть оспорено, невычислительные теории, даже сформулированные на языке естественных наук, вынужденно остаются де-факто метафизическими, поскольку не могут соотнести свои гипотезы с эмпирическими данными по каким-то более или менее понятным и фальсифицируемым правилам, которые по идее должны содержаться именно в этой схемотехнической или вычислительной подкладке.

Напротив, класс вычислительных теорий, на мой взгляд, уже сейчас предлагает эмпирически фальсифицируемые, т.е. научные, описания, которые, естественно, не идеальны и требуют различных улучшений, дополнений и т.д. Со многими из них можно поспорить, но тем не менее там идет движение в правильном направлении с точки зрения известных критериев научности.

Критерии не-спекулятивных теорий сознания

В 2021 г. вышла интересная статья трех авторов под названием «Жесткие критерии эмпирических теорий сознания»[7]. Авторы выдвигают четыре таких критерия. Первый: адресуется ли предполагаемая теория сознания парадигмальным случаям сознания или, проще говоря, включает ли она теоретические средства, позволяющие различить сознательные и бессознательные состояния. Потому что если теория сформулирована таким образом, что она одними и теми же средствами описывает сознательные и бессознательные состояния, то теорией сознания в строгом смысле она не является. Второй: свободна ли теория сознания от привязки каузальной структуре на уровне имплементации. Уточнение про уровень – мое дополнение, поскольку в статье это не совсем четко сформулировано. Я добавил его, потому что, как мы увидим далее, многие относительно успешные сегодняшние попытки создать теории сознания как раз уверенно настаивают на привязке к какой-то каузальности. Но каузальности именно на системотехническом (вычислительном в данном случае) уровне. Если же теория привязывается к каузальной структуре физической имплементации вычислительного процесса, т.е. к некоей материальной архитектуре, на которой осуществляется «сознательный» алгоритм, то возникают некоторые несуразности. Авторы приводят такой пример. Есть рекуррентная теория сознания, которая утверждает, что для появления сознания нужны рекуррентные сети. Но тем, кто в какой-то степени интересовался современными искусственными нейронными сетями, известно, что значительная часть их начиналась с сетей прямого распространения – так называемых *feed forward networks*. В таких сетях данные передаются строго от входного уровня через некоторое количество вложенных уровней на выходной уровень, а обратно распространяется только сообщение об ошибке. В рекуррентных же сетях ряд нейронов или все

---

6 *Peters F.* Consciousness as Recursive, Spatiotemporal Self-Location // Nature Precedings. 2008. DOI: 10.1038/npre.2008.2444.1.

7 *Doerig A., Schurger A., Herzog M.H.* Hard criteria for empirical theories of consciousness // Cognitive Neuroscience. 2021. Vol. 12. No. 2. P. 41–62.

нейроны могут быть дополнительно замкнуты на самих себя, моделируя этим эффект кратковременной памяти. Сторонники этой теории утверждают, что сознание может возникнуть только в рекуррентных сетях, и соответственно стараются эмпирически обнаружить элементы этой архитектуры в мозге. Но, как пишут авторы статьи, все математики знают, что рекуррентные сети можно развернуть в сети прямого распространения, т.е. математически переписать одни в другие. Следовательно, рекурренция не является объяснительным принципом для сознания.

Третий критерий: свободна ли теория сознания от аргумента малых и больших сетей. Иными словами, если теория, подобно теории интегрированной информации Джулио Тонони, сформулирована таким образом, что даже небольшая несложная материальная структура может обладать каким-то показателем сознательности, то возникает опасность панпсихизма. Но это еще не самая большая опасность, поскольку сегодня достаточно много философов, придерживающихся современных версий панпсихизма. Но появляются проблемы единства сознания, потому что если, например, мозг является очень большой и очень сложной сетью с многочисленными нейронами и неразгаданными функциональными связями между ними, то возникает вопрос: чем объясняется то, что мы воспринимаем сознание как нечто единое? Может быть, это эпифеномен и вообще некая иллюзия, а на самом деле в нас существует много разных сознаний, соседствующих или следующих друг за другом.

Наконец, четвертый критерий: устойчива ли теория сознания к аргументу других систем. Авторы приводят интересный пример: есть, например, теория, которая утверждает, что сознательные явления возникают только при взаимодействии таламуса с корой, и для этого нужна определенная структура и определенный размер мозга. Однако они показывают (с разрешения соответствующих людей и органов) рентгеновский снимок реального человека, который родился с очень усеченным объемом мозга, в котором даже таламус не обнаруживается. Этот человек тем не менее прожил нормальную жизнь с абсолютно нормальными человеческими когнитивными способностями, которые ничем не отличают его от других. Значит, все-таки возможна реализация тех же функций на системах с упрощенной архитектурой.

Можно привести аналогичный пример из другой статьи, к сожалению, отозванной ее авторами. Там речь шла о некоторых видах птиц – врановых и некоторых разновидностях попугаев, – которые обнаруживают гораздо более интересные и развитые когнитивные способности, чем некоторые млекопитающие. Хотя, как известно, у птиц как у потомков динозавров серого вещества мало или оно отсутствует. Но это не мешает им достаточно эффективно решать всевозможные задачи на сообразительность. Поэтому теории сознания, которые привязываются к одной определенной архитектуре материального носителя, страдают этим недостатком и не проходят четвертый критерий.

В качестве наиболее популярных теорий они рассматривают теорию глобального рабочего пространства Баарса, теорию интегрированной информации Тонони и Коха, теорию мышления высшего порядка Розенталия и ряд других, не столь известных. В статье предлагается таблица, где по горизонтали отложены эти теории, объединенные в некие группы, а по вертикали расположены эти четыре критерия, плюс еще дополнительные обстоятельства: адресуются ли эти теории к состояниям сознания, или к его

содержанию, или к тому и другому; рассматривается ли сознание как градуированное или как бинарное (есть или нет); является ли оно единым, унитарным, или не унитарным, или ответ не определен; представляется ли оно континуальным, дискретным или ответ не определен, и какова судьба бессознательных элементов – ясна она или не ясна. Таблица наглядно показывает, что в каких-то отношениях выигрывает теория глобального рабочего пространства, а в других преимущество оказывается на стороне теории мышления (репрезентаций) более высокого порядка.

Не менее интересный подход к классификации и к критериям возможных эмпирических теорий сознания обнаружил К.В. Анохин. На мой взгляд, по состоянию на сегодня это лучший русскоязычный текст, который написан на тему методологии эмпирических исследований сознания: он очень обстоятелен и аналитичен, но при этом информативен. Анохин уверен, что свойства, которые мы приписываем сознанию, суть некое уточнение или спецификация общих биологических свойств. Специальная таблица в статье показывает взаимозависимость этих более абстрактных и более конкретных свойств, и выясняется, что наиболее важным является то из них, которое на уровне сознания возникает как качественность или квалитативность, а его прототипическим биологическим свойством оказывается специфичность. Эта же специфичность называется в качестве прародителя такого свойства, как интенциональность. Это интересное утверждение, потому что обычно в теориях сознания качественность и интенциональность рассматриваются как противоположные характеристики.

Интересно также его замечание по поводу так называемой трудной проблемы сознания. Как пишет Анохин, его подход, «безусловно, не снимает вопроса о специфике конкретных механизмов интеграции, дифференциации и других свойств в сознательной системе. Это также не означает, что все разнообразие субъективных феноменов, входящих в сложную структуру сознания, должно сводиться исключительно к квалиа. Однако отмеченное обстоятельство позволяет сфокусировать теорию на задаче объяснения, в первую очередь, именно основ квалитативности субъективного опыта. Если теория сможет сделать это в естественно-научных понятиях, то есть надежда, что и другие специфические характеристики сознания получат свое закономерное объяснение»[8]. Как можно понять, с его точки зрения, квалиа – качественная субъективная специфика сознания, или, что то же самое, феноменальное сознание – это ключ к разгадке других проблем. Это интересное и небесспорное заявление, которое, безусловно, нуждается в дальнейшем развитии.

Анохин предлагает четырех- или пятикомпонентную структуру исследования сознания[9]. В этой структуре категория «кто» обозначает когнитивного агента или ту самую сущность, чем бы она ни была, которой мы приписываем сознательное состояние. Категория «когда» обозначает порог, ниже которого располагаются бессознательные состояния, а выше которого – сознательные. Состояние интегрированности, при котором возникают

---

[8] *Анохин К.В.* Когнитом: в поисках фундаментальной нейронаучной теории сознания // Журнал высшей нервной деятельности им. И.П. Павлова. 2021. Т. 71. № 1. С. 44.

[9] Четырех- или пятикомпонентную, потому что категория «кто» повторяется два раза: в начале и в конце. Мне представляется, что в конце стоило бы прибавить штрих, как в знаменитой схеме «деньги-товар-деньги´».

сознательные состояния, обозначаются категорией «что». Каузальная взаимосвязь с другими элементами когнитивной системы обозначается категорией «где». И, наконец, финальная категория «кто» – в моем предположении, «кто»-штрих – фиксирует изменение целостного состояния когнитивного агента.

«Кто», «что», «где» и «когда» – это вопросы, составляющие «универсальный сознательный эпизод», на которые должна ответить фундаментальная теория, чтобы объяснить его нейронаучную конкретику.

К.В. Анохин указывает на то, что невозможно по-настоящему понять сущность субъективного опыта, не понимая устройства его носителя. Невозможно также по-настоящему понять устройство разума, не понимая закона его формирования в процессах обучения. Но невозможно и по-настоящему понять процессы обучения, не понимая принципов развития нервной системы в онтогенезе. Конечно же, невозможно понять онтогенез без филогенеза и, наконец, невозможно по-настоящему понять закономерности эволюции нервной системы в филогенезе, не понимая роль нервных механизмов поведения и субъективного опыта. Этот замкнутый круг он называет «циркулярной ловушкой» сознания.

Здесь мы подступаем к критериям хорошей научной теории сознания по Анохину. Для начала он формулирует три требования ко всякой фундаментальной теории. Первый: теория должна обладать «широким кругозором»: в случае с теорией сознания это означает уметь объяснить ментальные явления, наблюдаемые от первого лица, их же, наблюдаемые от третьего лица, и нервные явления, наблюдаемые от третьего лица. Второй и третий критерии можно объединить следующим образом: фундаментальная теория должна быть «компактной», что означает ее простоту на входе и всеохватность на выходе. Иначе говоря, минимум простых принципов должен порождать удовлетворительные и универсальные объяснения для максимума явлений. Это соответствует критериям компактности и компрессии. На мой взгляд, эти требования можно применить не только к научной, но и ко всякой фундаментальной теории. Парадоксальным образом, этим требованиям может соответствовать и гегелевская «наука логики». Недостает требования фальсифицируемости.

Наконец, собственно критерии хорошей теории сознания. Первый: она должна быть научной в вышеописанном смысле, т.е. фундаментальной. Второй: она должна ответить на ряд сформулированных в статье вопросов «как» и «почему» в отношении отличительных свойств ключевых ингредиентов субъективного опыта. Третий: ответы на эти вопросы должны быть совместимы с требованием выхода из циркулярной ловушки сознания. Четвертый: фундаментальная теория сознания должна охватывать своим объяснением максимальное количество феноменов из целевой предметности, исходя при этом из минимального числа первых принципов. Пятый принцип: фундаментальная теория сознания должна основываться на понятиях биологического уровня, из которых должно закономерно выводиться возникновение свойств когнитивного уровня.

Последний пункт мне представляется наиболее дискуссионным, прежде всего потому, что из него не совсем понятно, что значит «основывается» и что значит «закономерно выводиться». Если мы вспомним критерии из предыдущей статьи, то теория, которая будет целиком и полностью основываться на биологической архитектуре, скорее всего не пройдет четвертый критерий.

Проблема множественности предметов объяснения

В 1990 г. Крик и Кох выпустили статью, на которую до сих пор массово ссылаются публикации по эмпирическим исследованиям сознания[10]. Там они впервые сформулировали понятие нейронного коррелята сознания. А через десять лет Чалмерс переопределил это понятие как минимум нейронных механизмов, совместно необходимых для того, чтобы дать возможность любому конкретному сознательному опыту[11]. В статье группы авторов[12] ставится проблема множественности explananda – отсутствие единого понимания того, что именно должна объяснять наука о сознании и отсутствие консенсуса относительно того, как сознание должно быть операционализировано и измерено. То есть дело не только во множественности предмета или множественности явлений сознания, за которыми не всегда просматривается единая онтологическая основа, но еще и во множестве онтологических и методологических подходов. В статье В. Визе[13] сознание описано как набор характеристик, каждая из которых подвергается анализу с помощью различных исследовательских программ, в которых пока не достигнуто согласие относительно таксономии и экспериментальных парадигм, предназначенных к использованию.

Если попытаться внести некоторый логический порядок в эту множественность предмета эмпирического исследования сознания, мы увидим, что можно говорить о сознательных состояниях и о содержании сознания. При этом разные теории обращаются к разной предметности, и это разделение отчасти соответствует нашим онтологическим представлениям о сознании как о предмете, который в принципе разделяется на феноменальное и интенциональное. По этому вопросу имеют место философские дискуссии: являются ли эти два аспекта одним и тем же или это совершенно разные явления. Исследования состояний сознания скорее имеют отношение к тому, что философы называют феноменальным сознанием, и наоборот, исследования содержания сознания скорее имеют отношение к тому, что философы называют интенциональным. С другой стороны, некоторые философы считают, что феноменальное сознание также репрезентативно. Это направление известно как феноменальный репрезентационализм. Его сторонники считают, что качественные состояния сознания тоже репрезентациональны, т.е. содержательны, и наоборот, интенциональное полагается существенно квалитативным.

В эмпирических исследованиях, которые адресуются прежде всего состояниям сознания, в основном применяются такие эмпирические методы, как исследования контрастных паттернов мозговой активности между бодрствованием и сном, между сновидениями и сном без сновидений, а также между бодрствованием и анестезией, измеренными с помощью функциональной магнитно-резонансной томографии (фМРТ), магнито/электроэн-

---

10 *Crick F., Koch C.* Towards a neurobiological theory of consciousness // Seminars in the Neurosciences. 1990. Vol. 2. P. 263–275.

11 *Chalmers D.J.* What Is a Neural Correlate of Consciousness? // *Chalmers D.J.* The Character of Consciousness. New York, 2010. P. 59–90.

12 *Vilas M.G., Auksztulewicz R., Melloni L.* Active Inference as a Computational Framework for Consciousness // Review of Philosophy and Psychology. 2022. Vol. 13. No. 4. P. 859–878.

13 *Wiese W.* Toward a mature science of consciousness // Frontiers in Psychology. 2018. Vol. 9. DOI: 10.3389/fpsyg.2018.00693.

цефалографии (М/ЭЭГ), электрокортикографии (ЭКоГ). Что касается исследований феноменального содержания, их предметами являются (1) универсальные качества опыта, (2) особенные качества конкретного опыта с помощью таких методов, как качественные исследования и отчеты от первого лица. В качестве методологической основы используется то, что Ф. Варела в свое время обозначил как нейрофеноменологию, а также теория интегрированной информации Дж. Тонони. В исследованиях содержания доступа, во-первых, исследуются (1) процессы, которые позволяют делать конкретную информацию доступной для использования когнитивными процессами более высокого уровня, (2) метакогниции и метаосознание. В качестве методов исследования используются (1) бинокулярное соперничество, (2) количественные параметры стимула (например, длительность). Сканируются нейронные возбуждения, соответствующие разным отчетам от первого лица.

Согласно относительно недавнему исследованию Дж. Тонони[14], который утверждает, что пространство – это также квале, т.е. качественное состояние сознания, лучшим кандидатом на роль нейронного коррелята феноменального пространства является решетчатая структура нейронов задней коры. Как известно, одна из аксиом теории интегрированной информации состоит в том, что структура феноменального восприятия должна быть основана на структуре своего нейронного коррелята. Соответственно, если пространство воспринимается как некая решетка, состоящая из пятен, точек, каких-то вложенных друг друга структур и отношений между ними, то это именно потому, что так же организована решетчатая структура нейронов задней коры. Второй вывод его эмпирического исследования состоит в том, что система в некотором состоянии должна задавать максимально нередуцируемую, специфическую, композиционную, внутреннюю причинно-следственную структуру, которая состоит из различий и их отношений.

Насколько можно понять, его знаменитая степень сознательности Ф, которая вычисляется по определенной формуле, в качестве своего физического смысла имеет эту, как он ее называет, каузальную нередуцируемость. Если система достаточно сложна для того, чтобы ее внутренние каузальные отношения невозможно было редуцировать к более простым, то ее степень сложности – это и есть величина ***Ф***, которая эквивалентна степени сознательности.

Активный вывод как методология теории сознания

Наконец, теория, которую в контексте исследований сознания называют теорией активного вывода (active inference), но которая ранее и в других контекстах была известна как теория предиктивного процессинга (ПП). Прежде чем сформулировать концепцию сознания, предложенную в рамках общей предиктивной гипотезы, нужно разобраться в довольно сложной терминологии, которую используют адепты этого направления, к которым относится Карл Фристон, Крис Фрит, Якоб Хохви, Энди Кларк и многие другие, включая Томаса Метцингера и уже цитированного ранее Ваню Визе. Главный принцип, который лежит в основе этого подхода, – это *принцип*

[14] *Haun A., Tononi G.* Why does space feel the way it does? Towards a principled account of spatial experience // Entropy. 2019. Vol. 21. No. 12. DOI: 10.3390/e21121160.

минимизации свободной энергии, который, по признанию авторов теории, не является фальсифицируемым. Однако из этого не следует ненаучность теории, поскольку в основе классического естествознания также лежат нефальсифицируемые принципы – например, принцип причинности или сохранения энергии. Напротив, законы классического естествознания фальсифицируемы, потому что из них выводятся определенные эмпирические события: падающий мяч можно в конечном счете объяснить законом всемирного тяготения. Таким образом, это потенциально эмпирически фальсифицируемый и потому в полном смысле слова эмпирически верифицированный закон: он мог бы оказаться ложным в некотором возможном мире, поэтому в актуальном мире он считается эмпирически обоснованным и истинным в том смысле, в котором вообще может быть истинен научный закон. А принципы вроде принципа сохранения энергии используются в науке как некие ограничительные нормативные требования, которые сами по себе не являются фальсифицируемыми: т.е. невозможно представить себе такой опыт, который опроверг бы их.

Согласно принципу минимизации свободной энергии, все адаптивные биологические агенты стремятся свести к минимуму долговременную неожиданность, т.е. энтропию, которая понимается здесь в смысле Шеннона. Общий подход, определяемый как предиктивный процессинг («предсказательная обработка» в некоторых вариантах перевода), – это набор гипотез, предполагающих, что мультистабильная аутопойетическая система (например, мозг) стремится уменьшить «удивление» (*surprisal*)[15] – термин, обозначающий расхождение актуальных перцептивных данных с теми, которые предсказаны ее внутренней генеративной моделью. Согласно ПП, мозг стремится уменьшить «удивление», делая выводы о скрытых состояниях мира или «причинах», порождающих наш сенсорный опыт, с использованием байесовских механизмов вывода.

Когда сторонники ПП говорят, что мозг, или вообще когнитивная система, или даже вообще биологические системы целом делают какие-то «выводы» о скрытых «причинах», их можно заподозрить в панпсихизме: будто бы все подобные системы на самом деле являются природными философами, которые постоянно размышляют о скрытых причинах. Но на самом деле, если переводить этот специфический жаргон на научный язык, речь идет о том, что в сложных мультистабильных – например, когнитивных – системах действуют некие аттракторы, порождающие априорные распределения вероятностей на «верхнем» системном уровне. Их значения транслируются вниз по байесовой иерархии, а наверх передается только разница между предсказанными и полученными данными. То есть, это не локковская модель, в рамках которой наша душа – это «чистая доска», где отпечатываются подлинные образы внешних объектов: никаких «подлинных» образов внешних объектов у нас нет. У нас есть предсказанные параметры перцептивных данных и параметры актуально полученных сигналов, разница между которыми провоцирует многочисленные итерации обновления генеративной

[15] Этот термин, как многие другие в этом семействе теорий, заимствован из байесовской статистики, где, как известно, используется специфический жаргон, который звучит очень психологично, например «убеждения» или «политика». Несмотря на то, что эта статистическая теория применяется в том числе к системам, которые не предполагают какой бы то ни было интенциональности или психологичности, терминология сохраняется.

модели, порождающей эти предсказания, ведущие к минимизации свободной энергии – «удивления». В результате итоговая репрезентация внешних «причин» – гипотез, из которых вероятностно выводятся ожидаемые перцептивные данные, – более или менее соответствует получаемой информации. Эта порождающая модель считается иерархической, где каждый уровень иерархии кодирует состояния во вложенных временных масштабах и каждый уровень принимает в качестве наблюдений скрытые состояния более низких уровней. Из устных разъяснений Фристона на его московских лекциях 2019 г. можно было заключить, что генеративная модель состоит из аттракторов, которые представляют собой некие циклические процессы. Если развернуть их по оси абсцисс, они превращается в синусоиду определенной формы, и когда они накладываются друг на друга, результирующая форма становится довольно сложной. Он предполагает, что в качестве аттракторов выступают волны мозговой активности, хотя, наверное, возможна и другая нейрофизиологическая интерпретация этой абстрактной вычислительной схемы. Мозг обладает достаточно сложной структурой, как, собственно, и любая живая материя, чтобы допускать множественные каузальные интерпретации.

Эта концепция была позже экстраполирована на живые клетки и в пределе – на все биосистемы.

Теперь о концепции «активного вывода». Если цель нейронной системы состоит в том, чтобы свести к минимуму ошибки предсказания, то этого можно достичь двумя способами: или делая выводы о скрытых состояниях мира, что предполагает, что в мире действуют такие же аттракторы и такие же генеративные модели, как в воспринимающей системе, которые максимизируют вероятность перцептивных данных; или путем *активного вывода*, т.е. смены выборки данных, получаемых системой, чтобы увеличить вероятность исполнения этих прогнозов. Философски говоря, эта концепция объясняет как познание, так и деятельность. Согласно Фристону, в каких-то случаях информация о рассогласовании предсказанных и полученных данных не приводит к обновлению генеративной модели, а передается сразу на рефлекторные дуги, и тогда организм начинает действовать, чтобы, не обновляя генеративную модель, получить другие данные, которые больше соответствуют предсказаниям генеративной модели.

Это похоже на общий закон бизнеса: чтобы максимизировать прибыль, нужно или сократить издержки, или увеличить доход. Перед похожей дилеммой стоит всякий организм. Модели объяснения сознания в рамках ПП называются поэтому моделями активного вывода (*active inference models (AIM)*).

Еще один важный термин для этого подхода – «точность предсказаний». Согласно теории, организм оценивает точность своих убеждений. Эти так называемые *оценки точности* могут взвешивать влияние ошибок предсказания (функция усиления), и когда они развертываются нисходящим (сверху вниз) образом в качестве ожиданий, они рассматриваются как механизмы внимания.

В совокупности эти вычислительные принципы составляют архитектуру активного вывода, которую можно использовать для построения теорий процессов конкретных когнитивных явлений. Но описанный теоретический подход не составляет специфицированную теорию сознания и не проходит первый критерий интернационального коллектива авторов: он не является

собственно теорией сознания, но скорее общей теорией когнитивных, а может быть, даже и вообще биологических процессов. Следовательно, это универсальная объяснительная схема. В каком случае она может стать или попытаться стать теорией сознания?

Существует не так много публикаций, написанных адептами ПП и, прежде всего, самим Фристоном, где этот подход специфицировался бы в отношении интересующей нас проблемы. Я опираюсь на две релевантные публикации: на его авторскую статью[16] и на статью, написанную в соавторстве[17].

Он вводит еще два дополнительных понятия, которые превращают теорию предиктивного процессинга / активного вывода в своего рода теорию сознания. Это понятие *утолщения времени*, которое он заимствует у Мерло-Понти, и понятие *контрфактуальной глубины*. Утолщение времени и контрфактуральной глубины в его случае рассматриваются как градуированные характеристики генеративных моделей, которые позволяют делать выводы о состояниях, находящихся дальше во времени, и сравнивать большее количество политик[18]. Поскольку предполагается, что эти два свойства лежат в основе явления сознания, сторонники активного вывода утверждают, что сознание, следовательно, должно быть градуированным явлением. Я интуитивно склонен считать так же и полагаю, что сознание определяется не булевыми значениями «есть/нет», а по принципу реостата: «больше/меньше». Насколько можно судить, утолщение времени – это своего рода аппроксимация перцептивных данных по времени. Темпорально-утолщенная генеративная модель способна к предикции и ретродикции, но она также может предсказывать «удивление», ожидаемое как результат возможного действия – активного вывода. Это обстоятельство добавляет поведению такую характеристику, как целесообразность. И оно же понятным образом связано с контрфактуальной глубиной – способностью предсказывать последствия альтернативных действий.

Каким образом с помощью этих теоретических дополнений намечается решение традиционных вопросов научных теорий сознания? Содержание сознания с точки зрения теории активного вывода формируется на средних уровнях иерархии, и там же формируются феноменологические свойства сознания: «…видение красного цвета и ощущение боли… сами по себе являются предполагаемыми причинами, *сконструированными* (курсив мой. – *И.М.*), чтобы освоить (т.е. наилучшим образом объяснить) сырой сенсорный поток – и иерархические махинации, которые они вызывают»[19]. И там же: «Квалиа – точно так же, как собаки и кошки – являются частью предполагаемого набора скрытых причин (т.е. эмпирических гипотез), которые лучше всего объясняют и предсказывают развивающийся поток энергий через наши сенсорные поверхности»[20]. То есть, с точки зрения Фристона и соавторов, квалиа сами по себе – боль или цвет

---

16 *Friston K.* Am I Self-Conscious? (Or Does Self-Organization Entail Self-Consciousness?) // Frontiers in Psychology. 2018. Vol. 9. DOI: 10.3389/fpsyg.2018.00579.

17 *Clark A., Friston K., Wilkinson S.* Bayesing qualia: Consciousness as inference, not raw datum // Journal of Consciousness Studies. 2019. Vol. 26. No. 9–10. P. 19–33.

18 «Политика» – тоже байесовский термин, означающий ту или иную последовательность или правила забора данных, которые тоже могут меняться и модулироваться в зависимости от общего поведения системы.

19 Ibid. P. 21.

20 Ibid. P. 22.

как парадигмальные примеры – функционально ничем не отличаются от воспринимаемых образов целостных объектов.

Интересно, что в знаменитой кантовской концепции схематизма чистых понятий рассудка также речь идет о роли времени, поскольку Кант задается вопросом: каким образом такие разнородные вещи, как категории рассудка и чувственные образы, могут сложиться вместе, сформировав то, что он называет чистой схемой предмета? Как мы помним, она имеет по сути дела перцептивную природу и через *время* как априорную форму чувственности привязывается к рассудочным понятиям. Эта реминисценция, возможно, требует дополнительного изучения.

С моей точки зрения, утолщение времени, конечно, вряд ли решает трудную проблему сознания, а именно, каково это – видеть красный помидор. Но оно объясняет, почему мы видим его постоянно красным, хотя очевидно, что частоты и фазы отражаемых его поверхностью электромагнитных излучений не постоянны. Наличие утолщения времени на более высоких уровнях иерархии, которые делают вероятностные выводы относительно временной последовательности предсказанных восприятий, – весьма правдоподобная гипотеза. Таким образом, мать-природа экономит наши вычислительные и энергетические мощности, поставляя только необходимую для выживания информацию. Эта мысль высказывается также Дональдом Хоффманом, который выступил с весьма оригинальной «интерфейсной теорией восприятия»[21]. Он – также с математическим аппаратом в руках – объясняет, почему мы не просто не видим мир «как он есть», но мы не можем видеть его таким никогда в силу эволюционных причин.

Правда, из AIM-теории сознания следует, что квалитативные восприятия могут быть только сознательными, и это неожиданный вывод также и для меня. Данный вывод говорит в пользу онтологического единства феноменального сознания и сознания доступа. Ранее я склонялся к мысли, что это все-таки разные вещи, но если принять схему активного вывода, то получается, что необходимые нейронные или, лучше сказать в данном случае, вычислительные корреляты феноменального сознания и сознания доступа одни и те же.

Воплощенное познание, мозг в бочке и летучая мышь

И последнее, по поводу популярной ныне гипотезы *4E (embodied, embedded, enactive & extended) cognition*. В соответствии с нею, разум (mind) не находится исключительно в мозге, но распространяется на тело и даже физический мир. Части тела и объекты внешней среды, как считают ее адепты, могут реализовывать когнитивные процессы и, таким образом, функционировать как расширения самого разума (*mind*). Суть и цель разума (*mind*) они видят в обслуживании деятельности. Разум (*mind*) охватывает, с их точки зрения, все когнитивные уровни, включая физический[22].

---

[21] *Hoffman D.D., Singh M., Prakash C.* The Interface Theory of Perception // Psychonomic Bulletin & Review. 2015. Vol. 22. No. 6. P. 1480–1506.

[22] См.: *Varela F.J., Rosch E., Thompson E.* The Embodied Mind: Cognitive Science and Human Experience. Cambridge (Mass.), 1992; *Pritchard D.* Cognitive ability and the extended cognition thesis // Synthese. 2010. Vol. 175. P. 133–151.

Для того чтобы разобраться с этой проблематикой, я бы предложил объединить вместе несколько известных мысленных экспериментов и задаться вопросом, является ли тело и окружающая среда необходимыми с каузальной точки зрения элементами когнитивной системы. Я имею в виду прежде всего т.н. «мозг в бочке» – мысленный эксперимент, приписываемый Гилберту Харману. Его сюжет состоит в том, что если мы разгадываем перцептивные коды мозга, т.е. коды электрических импульсов, которые подаются в его перцептивные отделы, то мы можем заставить мозг, отделенный от тела и находящийся в питательном растворе, воспринимать мир, как нам хочется. С другой стороны сцены появляется Джордж Мур со своим знаменитым опровержением идеализма, который говорит: для того чтобы опровергнуть (берклианский) идеализм, достаточно хотя бы одного высказывания, которое было бы несомненно истинным, но истинность его зависела бы от существования внешнего мира как необходимого условия. С этой целью он поднимает руку, выступая с лекцией, и произносит такое несомненно истинное предложение: «Это моя рука». Оно истинно, по мнению Мура, но истинно только при условии существования внешнего мира. Однако суперученые, управляющие мозгом в бочке, могут заставить этот мозг считать, что у него есть рука и что он может поднять ее. И тогда, если это мозг философа и ему сообщили реальное положение вещей, он непременно спросит: «А как же быть с моим воплощенным познанием?»

И в этот момент у нас появляется третье действующее лицо, а именно Томас Нагель со своей знаменитой летучей мышью. Что если попытаться подать на этот мозг перцептивные коды летучей мыши с ее знаменитым сонаром или ультразвуковым эхолотом – возникнет ли что-либо новое в его феноменальной сфере? Этому, скорее всего, помешает или разница в физической архитектуре мозга летучей мыши и человека, или их принципы кодировки перцептивной информации, которые также могут оказаться разными. Эти коды, или, что то же самое, механизмы репрезентации, развиваются эволюционно и в корреляции с развитием телесной организации. И тогда получается, что то, что мы называем «воплощенным познанием», – на самом деле просто определенные коды мозга, которые, действительно, находятся в функциональной связи с реальными анатомическими органами, но они тем не менее суть не более чем видоспецифичные коды репрезентаций, используемых в когнитивных вычислениях.

Выводы

И, наконец, финальные соображения:

(1) Метод прямого сопоставления феноменальных (интроспективных) представлений с нейрофизиологическими данными пока успеха не принес.

(2) Причина – «починка радио биологом»: слепой поиск корреляций между нечеткими и неверифицируемыми концептами и неструктурированным потоком эмпирических данных.

(3) Недостающее звено – «инженерный уровень»: функциональные схемы, составленные из конечного множества типовых элементов и интерпретируемые как в терминах теоретических положений («законов»), так и в терминах эмпирических описаний.

(4) Кандидат на это место, достойный рассмотрения, – вычислительные модели как по-оккамовски простые описания сложных мультистабильных саморегулирующихся систем.

(5) Учитывая общепризнанную теперь необходимость учета обучения и эволюции, вычислительная модель, лежащая в основе хорошей теории сознания, должна быть не детерминированной, а вероятностной.

Список литературы

Анохин К.В. Когнитом: в поисках фундаментальной нейронаучной теории сознания // Журнал высшей нервной деятельности им. И.П. Павлова. 2021. Т. 71. № 1. С. 39–71.

Кант И. Собрание сочинений: в 8 т. Т. 3: Критика чистого разума / Пер. с нем. М.: ЧОРО, 1994.

Михайлов И.Ф. Социальные вычисления и происхождение моральных норм // Философский журнал / Philosophy Journal. 2022. Т. 15. № 1. С. 51–68.

Михайлов И.Ф. Человеческий мозг и сознание: биология или вычисления? // Философские проблемы информационных технологий и киберпространства. 2018. Т. 15. № 2. С. 92–110.

Chalmers D.J. What Is a Neural Correlate of Consciousness? // *Chalmers D.J.* The Character of Consciousness. New York: Oxford University Press, 2010. P. 59–90.

Clark A., Friston K., Wilkinson S. Bayesing qualia: Consciousness as inference, not raw datum // Journal of Consciousness Studies. 2019. Vol. 26. No. 9–10. P. 19–33.

Crick F., Koch C. Towards a neurobiological theory of consciousness // Seminars in the Neurosciences. 1990. Vol. 2. P. 263–275.

Doerig A., Schurger A., Herzog M.H. Hard criteria for empirical theories of consciousness // Cognitive Neuroscience. 2021. Vol. 12. No. 2. P. 41–62.

Friston K. Am I Self-Conscious? (Or Does Self-Organization Entail Self-Consciousness?) // Frontiers in Psychology. 2018. Vol. 9. DOI: 10.3389/fpsyg.2018.00579.

Haun A., Tononi G. Why does space feel the way it does? Towards a principled account of spatial experience // Entropy. 2019. Vol. 21. No. 12. DOI: 10.3390/e21121160.

Hoffman D.D., Singh M., Prakash C. The Interface Theory of Perception // Psychonomic Bulletin & Review. 2015. Vol. 22. No. 6. P. 1480–1506.

Lazebnik Y. Can a biologist fix a radio? – Or, what I learned while studying apoptosis // Cancer Cell. 2002. Vol. 2. No. 3. P. 179–182.

Peters F. Consciousness as Recursive, Spatiotemporal Self-Location // Nature Precedings. 2008. DOI: 10.1038/npre.2008.2444.1.

Pritchard D. Cognitive ability and the extended cognition thesis // Synthese. 2010. Vol. 175. P. 133–151.

Varela F.J., Rosch E., Thompson E. The Embodied Mind: Cognitive Science and Human Experience. Cambridge (Mass.): MIT Press, 1992.

Vilas M.G., Auksztulewicz R., Melloni L. Active Inference as a Computational Framework for Consciousness // Review of Philosophy and Psychology. 2022. Vol. 13. No. 4. P. 859–878.

Wiese W. Toward a mature science of consciousness // Frontiers in Psychology. 2018. Vol. 9. DOI: 10.3389/fpsyg.2018.00693.

Subjects and methods of empirical studies of consciousness*

Igor F. Mikhailov

Institute of Philosophy, Russian Academy of Sciences. 12/1 Goncharnaya Str., Moscow, 109240, Russian Federation; e-mail: ifmikhailov@gmail.com

Attempts to create empirically based theories of consciousness face two kinds of obstacles. First, the dominant strategy of searching for the neural correlates of consciousness has been unsuccessful due to the lack of working hypotheses about their causal connection with conscious states. The second obstacle is multiplicity of explananda – the lack of sufficient evidence for the belief that everything that we consider to be phenomena of consciousness or conscious states is ontologically unified. Perhaps, these issues are caused by the fact that the sciences of consciousness are devoid of an analog to the radio engineering level, which, in addition to theoretical electrodynamics, is essential for understanding the principles of radio devices' functioning. This level of knowledge should include a simplified ontology of the subject area, allowing one to isolate fundamental functional relationships at its algorithmic level. Considering the history of the sciences of consciousness and the specifics of their subject, the optimal candidate for the role of the engineering level of knowledge could be a computational approach, which would involve describing the subject as a combination of computational primitives that allow for the implementation of algorithms generating conscious states. Such an approach looks even more promising as non-computational theories of consciousness based on traditional natural science paradoxically remain de facto metaphysical (speculative). In addition to different approaches to the criteria for a "good" empirical (non-speculative) theory of consciousness, the paper provides an overview of a theory of consciousness based on the active inference hypothesis. The analysis of this theory suggests that a computational model underlying a good theory of consciousness should not be deterministic, but probabilistic.

Keywords: consciousness, neural correlates, integrated information, predictive processing, active inference

For citation: Mikhailov, I.F. "Predmety i metody empiricheskikh issledovanii soznaniya" [Subjects and methods of empirical studies of consciousness], *Filosofskii zhurnal / Philosophy Journal*, 2024, Vol. 17, No. 2, pp. 92–109. (In Russian)

References

Anokhin, K.V. "Kognitom: v poiskakh fundamental'noi neironauchnoi teorii soznaniya" [Cognitome: In Search of Fundamental Neuroscience Theory of Consciousness], *Zhurnal vysshei nervnoi deyatel'nosti im. I.P. Pavlova*, 2021, Vol. 71, No. 1, pp. 39–71. (In Russian)

Chalmers, D.J. "What Is a Neural Correlate of Consciousness?", in: D.J. Chalmers, *The Character of Consciousness*. New York: Oxford University Press, 2010, pp. 59–90.

Clark, A., Friston, K. & Wilkinson, S. "Bayesing qualia: Consciousness as inference, not raw datum", *Journal of Consciousness Studies*, 2019, Vol. 26, No. 9–10, pp. 19–33.

Crick, F. & Koch, C. "Towards a neurobiological theory of consciousness", *Seminars in the Neurosciences*, 1990, Vol. 2, pp. 263–275.

Doerig, A., Schurger, A. & Herzog, M.H. "Hard criteria for empirical theories of consciousness", *Cognitive Neuroscience*, 2021, Vol. 12, No. 2, pp. 41–62.

Friston, K. "Am I Self-Conscious? (Or Does Self-Organization Entail Self-Consciousness?)", *Frontiers in Psychology*, 2018, Vol. 9, DOI: 10.3389/fpsyg.2018.00579.

* The research was supported by Russian Science Foundation (project No. 24–28–00804), https://rscf.ru/project/24-28-00804/

Haun, A. & Tononi, G. “Why does space feel the way it does? Towards a principled account of spatial experience”, *Entropy*, 2019, Vol. 21, No. 12, DOI: 10.3390/e21121160.

Hoffman, D.D., Singh, M. & Prakash, C. “The Interface Theory of Perception”, *Psychonomic Bulletin & Review*, 2015, Vol. 22, No. 6, pp. 1480–1506.

Kant, I. *Sobranie sochinenij, Vol. 3: Kritika chistogo razuma* [Collected Works, Vol. 3: Critique of Pure Reason]. Moscow: ChORO Publ., 1994. (In Russian)

Lazebnik, Y. “Can a biologist fix a radio? – Or, what I learned while studying apoptosis”, *Cancer Cell*, 2002, Vol. 2, No. 3, pp. 179–182.

Mikhajlov, I.F. “Chelovecheskij mozg i soznanie: biologija ili vychislenija?” [The Human Brain and Consciousness: Biology or Computation?], *Filosofskie problemy informatsionnykh tehnologij i kiberprostranstva*, 2018, Vol. 15, No. 2, pp. 92–110. (In Russian)

Mikhajlov, I.F. “Social’nye vychislenija i proishozhdenie moral’nykh norm” [Social computations and the origin of moral norms], *Filosofskij zhurnal / Philosophy Journal*, 2022, Vol. 15, No. 1, pp. 51–68. (In Russian)

Peters, F. “Consciousness as Recursive, Spatiotemporal Self-Location”, *Nature Precedings*, 2008, DOI: 10.1038/npre.2008.2444.1.

Pritchard, D. “Cognitive ability and the extended cognition thesis”, *Synthese*, 2010, Vol. 175, pp. 133–151.

Varela, F.J., Rosch, E. & Thompson, E. *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, Mass.: MIT Press, 1992.

Vilas, M.G., Auksztulewicz, R. & Melloni, L. “Active Inference as a Computational Framework for Consciousness”, *Review of Philosophy and Psychology*, 2022, Vol. 13, No. 4, pp. 859–878.

Wiese, W. “Toward a mature science of consciousness”, *Frontiers in Psychology*, 2018, Vol. 9, DOI: 10.3389/fpsyg.2018.00693.