

КНИЖНАЯ ПОЛКА

А.А. Гусев

Рецензия на книгу П.Н. Барышникова «Вычислительные модели разума: От кода к смыслу» (М.: ЛЕНАНД, 2022. 320 с.)

Гусев Александр Андреевич – кандидат философских наук, младший научный сотрудник. Институт философии РАН. Российская Федерация, 109240, г. Москва, ул. Гончарная, д. 12, стр. 1; e-mail: sanya.cfg@yandex.ru

Когнитивная революция середины XX в. позволила ученым вернуть сознание в репертабельные научные исследования, тем самым минуя элиминативистские подходы бихевиоризма и физикализма. Машинный функционализм взял на вооружение компьютерную метафору и стал философским фундаментом для этого нового научного течения. Во второй половине XX в. усилиями когнитивных ученых было получено много значимых данных относительно того, как устроены психические процессы, но методологические границы вычислительных теорий и онтологический статус ментальных состояний так и остались серьезной проблемой. Именно данный сложный вопрос пристально анализирует П.Н. Барышников в своей книге «Вычислительные модели разума: От кода к смыслу». Двигаясь по исторической линии развития когнитивной науки и сопутствующих философских концепций, автор демонстрирует, что вычислительные теории не способны объяснить семантический аспект ментальных состояний, хотя они и обладают эвристическим потенциалом для создания искусственных интеллектуальных систем, реализующих некоторые когнитивные функции человеческого разума.

Ключевые слова: когнитивная наука, философия сознания, бихевиоризм, физикализм, функционализм, вычислительные теории сознания

Введение

Старая философская проблема о взаимоотношении мира ментального и мира физического является одной из главных тем для аналитической философии с середины XX в. Каков онтологический статус ментальных состояний

в целом и сознательных ментальных состояний в частности? Кажется, что с развитием эмпирических наук о психических явлениях данный вопрос не стал проще. Мы словно современники Рене Декарта – всё также склонны относить референты ментальных понятий к особому виду нефизических сущностей. Догма «призрака в машине» вынуждает нас рассматривать ментальные состояния как особые внутренние состояния субъектов, к которым имеет доступ только сам субъект. Как подмечал классик аналитической традиции Г. Райл: «[Н]еизбежной участью души является ее абсолютное одиночество. Встретиться могут только наши тела» [Райл, 1999, с. 25]. Сам же Райл полагал, что дуализм Декарта строится на категориальной ошибке и все ментальные термины могут быть проанализированы в терминах наблюдаемого внешнего поведения и диспозиций, что впоследствии стало называться логическим бихевиоризмом. Его идея шла в ногу с влиятельным в то время бихевиоризмом в психологии. Психологи этого толка строили объяснительные модели, которые не отсылали к «черному ящику» сознания, а концентрировались вокруг стимулов и реакций, вполне поддающихся интересубъективному описанию с позиции третьего лица. Это стало полезным инструментом в исследовании психики, позволившим найти альтернативу ненадежному интроспективному методу ранней научной психологии. Параллельно успехам в изучении работы мозга появлялись сторонники другого редуccionистского подхода – теории тождества ментального и физического. Идея заключалась в том, что ментальные состояния и есть состояния мозга, а не просто коррелируют с его активностью или находятся в причинных отношениях с ним. Для теоретиков тождества вопрос научного изучения ментальных состояний становится еще проще – изучайте физические свойства мозга и рано или поздно вы решите и все проблемы, связанные с психическим, ведь это одна та же вещь, просто данная нам в разных эпистемических модусах.

Затем в середине XX в. происходит бум в изучении искусственного интеллекта, появляется когнитивная наука – междисциплинарное направление, включающее в себя психологию, лингвистику, нейронауку, теорию искусственного интеллекта, антропологию и т.д. Вдохновленные идеями А. Тьюринга, ученые стали считать ментальные состояния реальными, заслуживающими полноценной роли в объяснительных моделях: «Модель компьютера как вычислительной системы, предназначенной для последовательного выполнения изменяемых операций, стала использоваться как центральная идея при объяснении когнитивных процессов» [Барышников, 2022, с. 74]. Компьютерная метафора стала не только полезной эвристикой в изучении естественного интеллекта и когнитивных процессов в целом, но и вариантом ответа на философский вопрос об онтологическом статусе ментального. Ментальное стало отождествляться с определенными абстрактными структурами, словно был найден компромисс между метафизикой дуалиста и мониста.

Анализу проблемы онтологического статуса ментальных состояний в вычислительной парадигме когнитивной науки и посвящена обсуждаемая нами книга П.Н. Барышникова, хотя, безусловно, в ней автор касается и других важных теоретических и методологических вопросов. В рамках рецензии мы сконцентрируемся лишь на теме онтологии ментального и релевантности вычислительных подходов для моделирования когнитивных процессов.

Машинный функционализм Х. Патнэма

Становление вычислительных теорий сознания напрямую связано с появлением функционализма в аналитической философии. Согласно одному из первых вариантов функционализма, машинному функционализму Х. Патнэма, ментальные состояния – это не физические состояния мозга или поведение, а скорее функциональные состояния сознательных систем по аналогии с машиной Тьюринга. Каждое ментальное состояние определяется взаимодействием с входными данными в виде стимулов внешней среды, выходными данными в виде реакций системы и другими ментальными состояниями. Но, как полагал Патнэм, если сущность ментального это его функциональная составляющая, то любая система, обладающая соответствующей функциональной организацией, будет обладать и соответствующим ментальным состоянием. Эта позиция стала называться тезисом множественной реализуемости. Функционализм бил напрямую по теории тождества, поскольку демонстрировал, что субстрат в виде мозга не является необходимым условием для ментального, а также косвенно по бихевиоризму, ведь в нем учитывалась роль ментальных состояний как именно внутренних свойств системы – одного наблюдаемого поведения недостаточно для индивидуации ментальных состояний. П.Н. Барышников выделяет следующие проблемные места машинного функционализма. Во-первых, существует проблема семантического характера. Для машинного функционализма Патнэма ментальные содержания являются результатом функциональных состояний, т.е. внутренними свойствами системы, но в то же время, как известно, Патнэм был одним из первых защитников экстернализма, согласно которому ментальное содержание определяется самими объектами внешней среды. Аргумент можно представить следующим образом:

- Если функционализм истинен, то содержание мышления должно быть обусловлено его функциональными свойствами.
- Машинная таблица (программа) определяет не экстенсионалы символов, а только команды операций над символами.
- Следовательно, машинный функционализм ложен, если содержание мышления не зависит от функциональных свойств [Барышников, 2022, с. 95].

Во-вторых, подобный функционализм в конечном итоге дрейфует либо к бихевиоризму, либо к физикализму. Бихевиористский вариант связан с тем, что позднее Патнэм дополнил свойства функциональной организации системы характеристиками поведенческих диспозиций [Там же, с. 96]. Физикалистская угроза исходит из возражений Дж. Кима и С. Сугдена о том, что если ментальные свойства могут реализоваться и в мозге, и полупроводниковых структурах, то это еще не означает, что данное состояние не может быть идентично какому-либо физическому свойству системы [Там же, с. 97]. Но почему тогда философская критика функционализма не сказалась на эмпирической методологии науки о сознании и когнитивной деятельности человека? По мнению Патнэма, это указывает на слабый объяснительный потенциал когнитивных наук в целом. На это можно посмотреть и под другим углом. Возможно, для такой молодой дисциплины не требовалась проработанная философская док-

трина, а был необходим лишь набор полезных эвристик для проведения экспериментов, построения моделей и т.д. Когнитивные исследования могут что-то сказать философам о том, как работают когнитивные системы, но они малоинформативны в вопросе об онтологии ментального.

Репрезентационная теория сознания Дж. Фодора

Следующей вехой в развитии философских оснований когнитивных наук была репрезентационная теория сознания Дж. Фодора. Фодор переключил внимание на сами символы и реанимировал старую философскую идею о языке мысли – ментализе. Он постулировал систему ментальных репрезентаций, включающую как примитивные базовые репрезентации, так и сложные репрезентации, образованные из примитивных. Например, примитивы ментализа в виде слов ДЖОН, МЭРИ и ЛЮБИТ могут быть объединены в ментализское предложение ДЖОН ЛЮБИТ МЭРИ. Ментализ обладает композиционной природой – значение сложного выражения является функцией от значений его частей и способа их объединения. Фодор полагал, что когнитивные процессы похожи на вычисления. Таким образом, символы из машин Тьюринга отождествляются с ментальными репрезентациями, а оперирование ими определяется правилами закодированного в мозгу универсального языка мысли [Лекторский, 2006, с. 16]. Хотя репрезентационализм и переключает свое внимание на семантический аспект, который был проблемой для машинного функционализма Патнэма, тем не менее это сталкивает Фодора с одной и из сложнейших философских проблем – проблемой значения языковых выражений [Барышников, 2022, с. 100].

Как подмечает П.Н. Барышников, сторонники вычислительной теории сознания, оставаясь на позициях Тьюринговской машинной метафоры, так или иначе поддерживают гипотезу физических символических систем, что накладывает ряд ограничений на их концепцию: «Ячейки памяти цифрового вычислительного устройства обладают лишь состояниями, обусловленными строго определенным синтаксисом. Содержательная сторона машинных состояний проявляется лишь посредством интерфейсов вывода, спроектированных под биологического пользователя, обладающего сознанием, – человека» [Там же, с. 102]. Таким образом, автор приходит к выводу, что данная версия вычислительной теории сознания также не способна продемонстрировать то, как семантические свойства ментального реализуются в функциональных состояниях вычислительной системы.

Коннекционизм

Следующим этапом в развитии когнитивной науки было появление коннекционизма в 80-х гг. XX в. как противника классического компьютерационализма. Как подмечает Д.В. Иванов: «Основной недостаток компьютерной метафоры, на которую опирается классическая когнитивная наука, заключается в том, что наш мозг во многих существенных аспектах не подобен компьютерам, основанным на архитектуре фон Неймана... Мозг представляет собой не набор дискретных элементов, последовательно выполняющих специфические

функции (центральный процессор, память и т.д.), а сеть нейронов, характеризующуюся параллельной обработкой информации, при которой в возбуждение приходит вся сеть целиком» [Иванов, 2016, с. 93–94]. Тем не менее в данном подходе не ставится вопрос об онтологии сознания, поскольку подразумевается, что сознание лишь эффект от вычислений сложной биологической материи. Помимо функционального сходства с работой биологических нейронов, у коннекционизма есть и другие сильные стороны. *Мягкие ограничения*, что позволяют адаптировать модель под реальные условия внешней среды. *Амортизация отказов системы*, повторяющая способность биологического мозга работать в условиях информационной избыточности или конфликтующих запросов. Но, на наш взгляд, самая важная черта, выгодно отличающая коннекционистские модели от классических, это способность сетей к *самообучению* посредством корректировки весов связей. У коннекционизма есть и значительные недостатки, вытекающие из его же преимуществ – когнитивные функции не сводятся лишь к распознаванию образов, категоризации и других схожих процессов, это еще и способность построения суждений, но мягкие ограничения коннекционистских моделей не справляются с жестким логико-синтаксическим каркасом пропозициональных отношений. В конечном итоге нейросеть остается неэффективным инструментом при моделировании языка, систем принятия решений, представления знания, поскольку репрезентационалистские установки не реализуются в распределенных вычислениях.

Заключение

Как было продемонстрировано в книге П.Н. Барышникова, революционные прорывы в когнитивной науке не должны вводить нас в заблуждение относительно такого сложного фундаментального вопроса, как природа сознания и психического в целом. Хотя последние полвека ученые когнитивисты довольно успешно моделируют различные когнитивные процессы, сводя их к глубинным алгоритмическим структурам, но именно семантический аспект всё еще остается серьезной проблемой для многих вычислительных теорий.

Список литературы

- Барышников, 2022 – *Барышников П.Н.* Вычислительные модели разума: От кода к смыслу. М.: ЛЕНАНД, 2022. 320 с.
- Иванов, 2016 – *Иванов Д.В.* Энактивизм и проблема сознания // Эпистемология и философия науки. 2016. Т. 49. № 3. С. 88–104.
- Лекторский, 2006 – *Лекторский В.А.* Философия, искусственный интеллект и когнитивная наука // Искусственный интеллект: междисциплинарный подход / Под ред. Д.И. Дубровского и В.А. Лекторского. М.: ИИнтелЛ, 2006. 448 с.
- Райл, 1999 – *Райл Г.* Понятие сознания. М.: Идея-Пресс: Дом интеллектуальной книги, 1999. 408 с.

Book review: P.N. Baryshnikov
“Computational models of the mind: From code to meaning”
(Moscow: LENAND, 2022. 320 pp.)

Alexander A. Gusev

Institute of Philosophy, Russian Academy of Sciences. 12/1 Gonchamaya Str., Moscow 109240, Russian Federation; e-mail: ddi29@mail.ru

The cognitive revolution of the mid-20th century allowed scientists to return consciousness into respectable scientific research, thereby bypassing the eliminativist approaches of behaviorism and physicalism. Machine functionalism adopted the computer metaphor and became the philosophical foundation for this new scientific trend. In the second half of the 20th century, many significant data were obtained by the efforts of cognitive scientists on how mental processes are arranged, but the methodological boundaries of computational theories and the ontological status of mental states remained a serious problem. This complex issue is closely analyzed by P.N. Baryshnikov in his book “Computational models of the mind: From code to meaning”. Moving along the historical line of development of cognitive science and related philosophical conceptions, the author demonstrates that computational theories are not able to explain the semantic aspect of mental states, although they have heuristic potential for creating artificial intelligent systems that implement some of the cognitive functions of the human mind.

Keywords: cognitive science, philosophy of mind, behaviorism, physicalism, functionalism, computational theories of mind

References

- Baryshnikov, P.N. *Vychislitelnyye modeli razuma: Ot koda k smyslu* [Computational Models of the Mind: From Code to Meaning], Moscow: LENAND Publ., 2022. 320 pp. (In Russian)
- Ivanov, D.V. *Enaktivizm i problema soznaniya* [Enactivism and the problem of consciousness], *Epistemologiya i filosofiya nauki*, 2016. vol. 49. No. 3, pp. 88–104. (In Russian)
- Lektorsky, V.A. *Filosofiya. Iskusstvennyy intellekt i kognitivnaya nauka* [Philosophy, artificial intelligence and cognitive science], *Iskusstvennyy intellekt: mezhdistsiplinarnyy podkhod* [Artificial intelligence: an interdisciplinary approach], ed. by D.I. Dubrovskiy, V.A. Lektorskiy. Moscow: IinteLL Publ., 2006. 448 pp. (In Russian)
- Ryle, G. *Ponyatiye soznaniya* [The concept of mind], Moscow: Ideya-Press. Dom intellektualnoy knigi Publ., 1999. 408 pp. (In Russian)