

**КРИТИЧЕСКИЙ ПОТЕНЦИАЛ НЕЙРОФИЛОСОФИИ\****Ю.Ю. ПЕТРУНИН*

Термин «нейрофилософия», появившийся с легкой руки американской исследовательницы П.С. Чёрчленд почти 30 лет назад<sup>1</sup>, стал довольно популярным за рубежом. Хотя, по мнению специалистов, он до сих пор не находится в мейнстриме мировой философии<sup>2</sup>, однако с каждым годом приобретает все больше сторонников и занял достойное место в научных исследованиях.

В нашей стране этому термину повезло пока меньше. Даже в русскоязычной Википедии нет статьи с таким названием, хотя исследователям, занимающимся проблемами философии сознания, искусственного интеллекта, когнитивных наук, он хорошо известен. За последние два года стремительно стало возрастать количество научных статей и научных мероприятий, связанных с темой нейрофилософии<sup>3</sup>.

Развитие наук, изучающих человеческий разум, делает естественным появление понятия «нейрофилософия». Однако понимание того, что должна представлять собой нейрофилософия, может и отличаться от взглядов американского философа.

Как известно, П.С. Чёрчленд считает, что нейронауки могут разрешить, в той или иной степени, многие классические философские проблемы: психофизиологическую, проблему человеческого Я, этические проблемы. Думается, что в настоящее время эти упования звучат несколько странно. По мнению ученых, принадлежащих к этой области знания, нейронауки сейчас находятся в кризисе и ищут помощи у других областей знания, в особенности у психологии и философии, а не спешат спасать братьев по разуму (т.е. философов)<sup>4</sup>. Совершенно очевидно, что существуют значительные различия во взглядах коллег на достижения и проблемы друг друга.

Когда-то философия пыталась поучать и даже предписывать определенное поведение людям, придерживающимся других форм познания реальности, других правил интерпретации добытых потом и кровью результатов, способов аргументации выводов. Иногда, напротив: отличные от философии практики пытались навязать ей решение внутренних проблем и стандарты функционирования.

---

\* Работа выполнена в рамках проекта Российского гуманитарного научного фонда (РГНФ) «Постнеклассическая парадигма искусственного интеллекта», грант № 15-03-00519.

Если обратиться к позитивному опыту, то можно вспомнить ряд более или менее удачных примеров синтеза философии с другими науками или формами социальной практики: философия науки, философия права, философия искусства. Во всех этих примерах речь идет не о редукции одной формы дискурса к другой, а о попытках такого разграничения и взаимодополнения (взаимопомощи) исследований, которые были бы полезны для каждой остающейся относительно самостоятельной области познания.

Если посмотреть сегодня на панораму нейронаук, то перед нами раскроется грандиозная картина, включающая в себя и естественные науки (нейроматематика, нейробиология, нейрофизиология, нейромедицина и др.), и технические (нейрокибернетика, нейрокомпьютинг, квантовые вычисления и т.д.), и гуманитарные (нейроэкономика, нейромаркетинг, нейроменеджмент, и даже нейрогосударственное управление<sup>5</sup>). Невозможно определить роль и возможности нейрофилософии, не осознав всю область нейронаук в их полноте и связности. Собственно говоря, на наш взгляд, нейрофилософия и должна начинать с попытки решения этой традиционной для философии задачи — классификации конкретной области наук. Никакая отдельная нейронаука сделать это не в состоянии.

Следующий естественным образом возникающий вопрос, ответ на который могла бы дать нейрофилософия, можно обозначать по-разному — и как онтологический, и как эпистемологический. Если использовать традиционный советский лексикон, то этот вопрос можно назвать основным вопросом философии применительно к конкретной сфере познания. Чёрчленд занимает по этому вопросу последовательно материалистические (или даже вульгарно-материалистические) позиции, которые она вместе с мужем, канадским философом Полом Чёрчлендом, называет элиминативным материализмом. Элиминативизм был заложен еще в работах У. Куайна, П. Фейерабенда и Р. Рорти<sup>6</sup> и нашел свое более или менее современное выражение в работах супругов Чёрчленд<sup>7</sup>. Замена «идеалистических» понятий «верование», «надежда», «страх», «намерение», «предпочтение» «материалистическими» «нейрон», «нейронная сеть» и т.д., на наш взгляд, — не очень перспективная ни для философии, ни для нейронаук, позиция. Поскольку в данном случае существуют ограничения по объему, то, оставив более развернутую аргументацию для подходящих условий, приведем лишь некоторые исторические аналогии, проясняющие высказанную мысль.

Хорошо известно, что первые представления (модели) атома были не только у материалиста Демокрита, но и у его яркого оппонента — идеалиста Платона (хотя он и не употреблял этот термин). Будущее оказалось за «атомизмом» Платона<sup>8</sup>, в котором элементы бытия понимались как математические модели.

Подход Чёрчленд похож на демокритовский: современные (а иногда и не современные, а устаревшие) теоретические конструкции нейрофизиологии выдаются за реально существующие, не подвергаемые критике, фрагменты реальности. Сомнительно, что примитивные представления о «летающих закорючках» или «нервных клетках» смогут объяснить сложные явления реального человеческого мира. Нейронные сети, основанные на модели МакКаллока-Питса, являются одной из возможных идеализаций нейрофизиологических процессов в человеческом мозге. В то же время в современной нейронауке активно обсуждаются и другие, альтернативные модели нейрона и нейросетей в целом. В любом случае, речь пока идет о математических моделях, а не о реальном телесном субстрате.

Одним словом, в нейрофилософии, возможен и антиматериалистический подход. Может быть, он даже является более эффективным.

Следующий вопрос, который можно было бы отнести к компетенции нейрофилософии, это вопрос о методологии нейронаук. В чем специфика этой методологии? Какие методологические программы можно оценить как более перспективные? Могут ли извлечь что-то полезное из методологических подходов нейронаук другие науки?

Эвристическая ценность нейронных сетей состоит не только в том, что они могут объяснить — хотя бы частично! — функционирование мозга и нервной системы в целом. Даже упрощенные представления о нейронах и нейронных сетях могут использоваться как формальные модели при решении проблем, никак непосредственно не связанных ни с сознанием, ни даже с психикой: экономических, управленческих, инженерных.

Для прояснения этой мысли приведу два исторических примера.

Первый относится к Античности. Аполлоний Пергский, решая геометрическую задачу сечения конуса плоскостью под разными углами, описал несколько видов кривых: окружность, эллипс, параболу, гиперболу. Через полторы тысячи лет оказалось, что полученные кривые описывают движения физических тел в пространстве в зависимости от скорости. В свое время полученный

формализм (модель) стал мощным средством изучения явлений, не имеющих никакого отношения к первоначальной задаче.

Другой пример относится к Средним векам. Леонардо Пизанский (известный также как Фибоначчи) поставил задачу о скорости размножения кроликов. Казалось бы, это биологическая задача. Через семь веков выяснилось, что последовательность, показывающая количество животных в каждый последующий период времени, а также отношения их численности в разные отрезки времени (так называемые коэффициенты Фибоначчи) характеризует не только процессы размножения, но и многие совершенно далекие от этого природные и социальные явления, в частности, процессы на товарных и финансовых рынках, которыми занимается технический анализ. Более того, в рамках последнего высказываются идеи, что первоначальная модель описывает самые глубокие циклические процессы во Вселенной.

Очевидно, что и нейронные модели в настоящее время эвристически полноценно развиваются в далеких от «материнской» среды областях. Представим себе наиболее общее описание функционирования искусственного нейрона. Имеются несколько входов и один выход. При определенной комбинации сигналов (превышении порогового значения) образуется сигнал на выходе нейрона. Каждый вход имеет некоторый вес, т.е. степень влияния на срабатывание нейрона. Так или нет «работают» настоящие нейроны мы точно не знаем: ежечасно в научных исследованиях возникают все новые и новые уточнения.

Но посмотрим на модель с другой стороны. Представим себе начало июня 1941 г. Сталину приносят донесение советского разведчика Рихарда Зорге о том, что фашистская Германия нападет на СССР 22 июня. Советский вожь не доверяет этому донесению. Почему? Да потому, что есть еще десятки других донесений, т.е. «входов» — в терминологии нейрокомпьютинга, каждый со своим «весом». Есть функция активации, т.е. заданные условия того, когда сумма донесений сработает. Чем не идеализированный нейрон? Принятие решений — хорошая модель, иллюстрирующая то, как он работает. Ясно, что это аналогия, но весьма полезная.

Нейрокомпьютерные идеи могут использоваться в других науках совсем не обязательно «в лоб», а скорее как метод, модель, метафора. Если технические приложения нейрокомпьютинга хорошо известны, то использование его теоретических моделей в более широкой сфере науки и практики только намечается. При этом речь идет не только и не столько об использовании новых механизмов обработки

эмпирических данных или о новых возможностях для наблюдений человеческой реальности (например, построение вместо регрессионных моделей нейрокомпьютерных моделей кредитных рисков; классификация с помощью самообучающихся сетей вместо традиционного статистического кластерного анализа). Сама постановка вопроса, способ анализа материала, интерпретация полученных результатов могут быть навеяны нейрокомпьютерной метафорой. Не углубляясь в чрезвычайно сложный и дискуссионный вопрос об изоморфизме реальных «живых» нервных сетей и искусственных нейронных, можно отметить, что нейрокомпьютерная парадигма, рассматривающая информационные процессы под определенным углом зрения и через определенные формализмы, может служить плодотворным средством для решения многих проблем социально-гуманитарных наук: социологии, экономики, политологии, управления, культурологии<sup>9</sup>.

Примерами такого подхода являются исследования в области приложений нейронных сетей в экономике (искусственные сообщества, интерпретация некоторых парадигм нейронных сетей как моделей портфелей опционов), в менеджменте, социологии или политологии. Можно, допустим, рассматривать организацию как «нейронную сеть», а процесс принятия управленческих решений как процесс обучения нейронной сети, или социальные и политические сети как особую архитектуру нейронной сети, или научный симпозиум как самообучающуюся нейронную сеть.

На наш взгляд, классические исследования в области нейронных сетей и отчасти искусственного интеллекта уже несколько десятилетий как зашли в тупик. Тупик этот — исключительно философско-методологический. Еще одно усовершенствование представлений об аксонах, еще одна конструкция нейронных сетей, еще один алгоритм обучения не приближают нас ни на йоту к решению проблемы интеллекта — ни искусственного, ни естественного. «Трудную проблему сознания» не решить с помощью нейронных сетей. Однако это не означает, что нейронными сетями вообще не нужно заниматься.

Кем-то из классиков искусственного интеллекта сказано, что нейрофизиология и теория интеллекта идут навстречу друг другу, но никогда не встретятся, потому что их дороги находятся на разных континентах. Продолжая метафору, можно сказать, что Колумб искал дорогу в Индию, а открыл Америку. Наверное, пора это признать.

На наш взгляд, будущее нейронных сетей, по крайней мере, в социально-гуманитарных науках, — не в эссенциалистском подходе, а в

инструменталистском. Так же как генетические алгоритмы, основывавшиеся на модели эволюции, сегодня наиболее активно применяются в областях, никак не связанных со своими «прародителями» — прежде всего в задачах оптимизации, так и модели искусственных нейронных сетей ожидают перспективы там, где они используются вовсе не как пародия на происходящее в человеке. Аналогичным образом изучение сечения конуса Аполлонием Пергским привело впоследствии не к частному успеху в узко геометрических областях. Триумф ожидал модели древнегреческого ученого через полторы тысячи лет в области описания движения физических тел — явлений, никак не похожих ни на конус, ни на сечение его плоскостью. Будущее нейрокомпьютерных моделей лежит в конкретных прикладных областях деятельности, где они используются как чистый формализм, получающий новые интерпретации: в управлении знаниями, в маркетинговых коммуникациях, в теории принятия решений, в коммуникационном менеджменте. Философия может и должна сыграть в этом интеллектуальном движении свою весомую роль: критически переосмысливая основания нейронаук, выделяя внутри них эффективные методологические программы, упорядочивая различные подходы и дисциплины, изучающие нейронные сети, обогащая нейронауки колоссальным накопленным и систематизированным опытом когнитивных моделей, метафор, интуиций.

#### ПРИМЕЧАНИЯ

<sup>1</sup> См.: Churchland P. Neurophilosophy: Towards a Unified Understanding of the Mind-Brain. — Cambridge (MA): MIT Press, 1986.

<sup>2</sup> См.: Block N. Neurophilosophy or Philoneuroscience // Science. 5 September 2003. Vol. 301. P. 1328–1329.

<sup>3</sup> См.: Алексеев А.Ю., Кузнецов В.Г., Петрунин Ю.Ю., Савельев А.В., Янковская Е.А. Актуальные вопросы нейрофилософии // Нейрокомпьютеры: разработка, применение. 2015. № 4. С. 9–11; Алексеев А.Ю., Кузнецов В.Г., Петрунин Ю.Ю., Савельев А.В., Янковская Е.А. Нейрофилософия как концептуальная основа нейрокомпьютинга // Нейрокомпьютеры: разработка, применение. 2015. №5. С. 69–77; Кузнецов В.Г. Вопрос обоснования нейрофилософии // Нейрокомпьютеры: разработка, применение. 2015. № 4. С. 51–52; Петрунин Ю.Ю. Нейрофилософия в системе нейронаук // Нейрокомпьютеры: разработка, применение. 2015. № 4. С. 63–64; Петрунин Ю.Ю. Нейрокомпьютинг: между наукой и лженаукой // Нейрокомпьютеры: разработка, применение. 2015. № 8. С. 52–64; Савельев А.В. Нейрофилософия или нейрософия — что является более общим? // Нейрокомпьютеры: разработка, применение. 2014. № 4. С. 71–72; Чечкин А.В., Алексеев А.Ю., Петрунин Ю.Ю., Савельев А.В., Янковская Е.А. Нейрофилософия, впервые в России // Нейрокомпьютеры: разработка и применение. 2014. № 10. С. 58–69.

На XIII Всероссийской конференции в марте 2014 г. впервые в России была создана секция по нейрофилософии. С октября того же года на философском факультете МГУ им. М.В. Ломоносова работает научный семинар по нейрофилософии.

<sup>4</sup>См. обзорную статью по поиску новых моделей для нейроисследований в гуманитарных науках: *Савельев А.В.* Нейрокомпьютеры и общество // *Нейрокомпьютеры: разработка, применение.* 2015. № 7. С. 35–43.

<sup>5</sup>См.: *Farmer D.J.* Change the course, neurons // *Administrative Theory and Praxis.* 2006. Vol. 29. № 1. P. 182–192.; *Farmer D.J.* Neuro-Gov: Neuroscience and governance // *Administrative Theory and Praxis.* 2006. Vol. 28. № 4. P. 653–662.

<sup>6</sup>См.: *Quine W.* Word and Object. – Cambridge (MA), 1960 (в русском переводе: *Квайн У.* Слово и объект. – М. 2000); *Feyerabend P.* Mental Events and the Brain // *Journal of Philosophy.* 1963. Vol. 60; *Rorty R.* Philosophy and the Mirror of Nature. – Princeton, 1979 (в русском переводе: *Рорти Р.* Философия и зеркало природы. – Новосибирск, 1997).

<sup>7</sup>*Churchland P.M.* Matter and Consciousness. – Cambridge (MA), 1984; *Churchland P.S.* Neurophilosophy: Towards a Unified Understanding of the Mind-Brain. – Cambridge (MA): MIT Press, 1986; *Churchland P.M.* The Engine of Reason, the Seat of the Soul. – Cambridge (MA), 1995.

<sup>8</sup>См.: *Гейзенберг В.* Физика и философия. Часть и целое. – М.: Наука, 1989.

<sup>9</sup>См.: Нейрокомпьютерная парадигма и общество / под ред. Ю.Ю. Петрунина. – М.: Изд-во Моск. ун-та, 2012.

#### REFERENCES

Alexeev A.Yu., Kuznetsov V.G., Petrunin Yu.Yu., Savelyev A.V., Yankovskaya E.A. Neurophilosophy as conceptual neurocomputing. In: *Neurocomputers: Design and Applications.* 2015. Vol. 5, pp. 69-77 (in Russian).

Alexeev A.Yu., Kuznetsov V.G., Petrunin Yu.Yu., Savelyev A.V., Yankovskaya E.A. Current issues of Neurophilosophy. In: *Neurocomputers: Design and Applications.* 2015. Vol. 4, pp. 9-11 (in Russian).

Block N. Neurophilosophy or Philoneuroscience. In: *Science.* 5 September 2003. Vol. 301, pp. 1328-1329.

Chechkin A.V., Alexeev A.Yu., Petrunin Yu.Yu., Savelyev A.V., Yankovskaya E.A. Neurophilosophy. In: *Neurocomputers: Design and Applications.* 2014. Vol. 10, pp. 58-69 (in Russian).

Churchland P. *Neurophilosophy: Towards a Unified Understanding of the Mind-Brain.* Cambridge (MA), MIT Press, 1986.

Churchland P.M. *Matter and Consciousness.* Cambridge (MA), 1984.

Churchland P.M. *The Engine of Reason, the Seat of the Soul.* Cambridge (MA), 1995.

Farmer D.J. Change the course, neurons. In: *Administrative Theory and Praxis.* 29 (1): 182-192. 2006.

Farmer D.J. Neuro-Gov: Neuroscience and governance. In: *Administrative Theory and Praxis.* 28 (4): 653-662. 2006.



Feyerabend P. Mental Events and the Brain. In: *Journal of Philosophy*. 1963. Vol. 60.

Heisenberg W. *Physics and philosophy*. Moscow, Science Press, 1989 (Russian trans.).

Kuznetsov V.G. The justification question of Neurophilosophy. In: *Neurocomputers: Design and Applications*. 2015. Vd. 4, pp. 51-52 (in Russian).

Petrinin Yu.Yu. (ed.) *Neurocomputing Paradigm and Society*. Moscow: Moscow State University Press, 2012 (in Russian).

Petrinin Yu.Yu. Neurocomputing between sciences and pseudoscience. In: *Neurocomputers: Design and Applications*. 2015. Vd. 8, pp. 52-64 (in Russian).

Petrinin Yu.Yu. Neurophilosophy in system neuroscience. In: *Neurocomputers: Design and Applications*. 2015. Vd. 4, pp. 63-64 (in Russian).

Quine W. *Word and Object*. Cambr. (Mass.), 1960.

Savelyev A.V. Neurophilosophy vs Neurosophy – what is more general? In: *Neurocomputers: Design and Applications*. 2014. Vd. 4, pp. 73-74 (in Russian).

Rorty R. *Philosophy and the Mirror of Nature*. Princeton, 1979.

Savelyev A.V. Neurocomputing and Socium. In: *Neurocomputers: Design and Applications*. 2015. Vd. 7, pp. 35-43 (in Russian).

### Аннотация

В статье раскрывается понимание нейрофилософии как онтологии, эпистемологии и методологии нейронаук. Определены перспективы интерпретации искусственных нейронных сетей в качестве формальной модели для исследования общества и человеческого поведения. При таком понимании нейронаук философская программа Платона представляется наиболее адекватной для решения задач нейрофилософии.

**Ключевые слова:** нейрофилософия, нейронауки, Чёрчленд, Платон, нейрон, нейросеть, нейрокомпьютинг, искусственный интеллект, методология нейрокомпьютинга.

### Summary

The article reveals the understanding of neurophilosophy as the philosophy and methodology of neuroscience and highlights its ontological, epistemological and methodological problems. The author proposes a methodological program of Plato for the development of neuroscience and defines the prospects for interpretation of artificial neural networks as formal models for the study of society and human behavior.

**Keywords:** neurophilosophy, neuroscience, Churchland, Plato, neuron, neural network, neurocomputing, artificial intelligence, neurocomputing methodology.