Бруно Маршаль: Метафизика вычислений

Е. Б. Рудный, ©, 2023, blog.rudnyi.ru/ru

Читать онлайн: http://blog.rudnyi.ru/ru/2023/09/bruno-marchal-computationalism.html

Вычислительная теория сознания лежит в основе идей сильного искусственного интеллекта и загрузки сознания человека в компьютер. Вычислительная теория сознания также была популярна среди сторонников кибернетической парадигмы, хотя в настоящее время многие нейроученые занимают позицию углеродного шовинизма и поэтому отвергают возможность появления сознания в роботах.

В любом случае работы логика Бруно Маршаля показывают неожиданные следствия из вычислительной теории сознания. Он характеризует свою позицию как компьютационализм (computationalism). Важно отметить, что Бруно использует понятие вычисление исключительно в смысле теории вычислений и машины Тьюринга. Поэтому я охарактеризовал его взгляды как метафизика вычислений.

Ниже я опишу аргументацию Бруно из статьи ‘*Переформулировка проблемы сознание-тело в компьютационализме*‘ (все ссылки в конце заметки), но я проведу обсуждение в другой последовательности; начну с конца, с последнего шага, когда доказывается, что вычислительная теория сознания отвергает материализм. Лишний раз напомню, что вычисление ниже не является расплывчатой метафорой, это строго определенное понятие из теории вычислений.

Бруно позиционирует свою картину мира как нейтральный монизм: весь мир, как сознание, так и физика, состоит из вычислений. Важно отметить, что разделение вычисление-материя является дуализмом, подобным дуализму сознание-материя. Остановимся на этом моменте; в статье это не разбирается, там дается другая аргументация, которая будет рассмотрена после этого.

В общем случае вопрос стоит таким образом: когда физический процесс, например, переключение транзистора, считается вычислением, а когда нет. Поскольку не все физические процессы связаны с вычислением, требуется указать критерий, когда определенный физический процесс можно назвать вычислением. Проблема в том, что такой критерий отсутствует, то есть, физический процесс относят к вычислениям только с точки зрения человека.

Например, в компьютере человек отобрал такие физические процессы, результаты которых соответствуют точному результату элементарного вычисления. На этой основе связанные между собой физические процессы позволяют получить результаты сложных вычислений. Работа компьютера как вычислительного устройства связана с большим количеством стандартов, например, представление чисел, и эти стандарты не связаны напрямую с

законами физики. Дуализм в данном случае заключается в том, что человек использует законы физики в своих целях и эти цели не следует из законов физики; логика организации процессов в компьютере связана с теорией вычислений.

Дуализм «цели человека, связанные с алгоритмом» — «физический процесс» переносится в компьютер, где такая двойственность связывает алгоритм как таковой с протекающим физическим процессом. Во время работы компьютера нельзя выделить моменты превращения физического процесса в вычисление и влияние вычисления на последующие физические процессы, как показано ниже:

физический процесс -> вычисление -> физический процесс -> вычисление …

При рассмотрении на более детальном уровне у нас останется либо алгоритм в чистом виде, который не зависит от физики, либо физический процесс, который зависит от алгоритма лишь опосредованно, через граничные и начальные условия, созданные человеком.

Надеюсь, что вышесказанное облегчит понимание аргумента Бруно, к которому я теперь перехожу. Исходной точкой служит идея супервентности сознания по отношению к физическим процессам:

физические процессы -> сознание

Возбуждение нейронов никак не похоже на ментальные процессы, но предполагается, что для протекания определенного ментального процесса (например, боли) необходимо протекание соответствующих физических процессов. Концепция супервентности не говорит, что сознание эквивалентно возбуждению нейронов, но считается, что для появления боли требуется возбуждение нейронов.

Вычислительная теория сознания предполагает, что требуется включить в схему супервентности вычислительные состояния:

физические процессы -> вычислительные состояния -> сознание

Вычислительные состояния становятся посредниками между физическими процессами и сознанием. Другими словами, вычислительные состояния супервентны по отношению к физическим процессам, а сознание супервентно по отношению к вычислительным состояниям.

На этом этапе следует сказать, что такое вычислительное состояние. В теории вычислений вычисление связано с машиной Тьюринга, а вычислительное состояние ассоциируется с определенным состояниями ленты машины Тьюринга. Возможно, что это покажется странным, но другого пути нет. Вместо машины Тьюринга можно взять другие системы, но суть будет одна. В конце концов появится нечто, напоминающее ленту Тьюринга; другими словами, вычислительное состояние будет сводится к последовательности нулей и

единиц.

Важно отметить, что в отличие от физических процессов вычислительные состояния дискретны. Вернемся к компьютеру. Переключение транзистора является непрерывном процессом, но вычислительное состояние будет относиться только к определенным моментам времени физического компьютера. Нельзя забывать, что именно это обстоятельство позволяет аргументацию за сильный искусственный интеллект и загрузку сознания в компьютер. Непрерывные физические состояния в компьютере и мозге сильно отличаются друг от друга, но они приводят к одинаковым дискретным вычислительным состояниям.

Схема выше имеет смысл, когда изменение в состояниях физической системы по сложности сопоставимы с такими изменениями в сознании. Именно этот пункт подвергается атаке в аргументации против материализма. Идея заключается в том, что можно сжульничать — получить необходимую последовательность вычислительных состояния заранее, а затем просто ее быстро проиграть. Согласно исходным предположениям в этом случае должны возникнуть точно такие же состояния сознания, но сложность физических процессов на этом пути будет минимальна.

Бруно Маршаль предложил этот аргумент в 1988 году под названием аргумент фильмового графа (filmed graph argument или movie graph argument); название аргумента отражает идею выше. Философ Тим Модлин в 1989 году предложил независимо аналогичный аргумент, при этом в его изложении все выглядит более убедительно (Бруно признает это и всегда ссылается на статью Модлина).

Модлин анализирует схему выше и вводит условия достаточности и необходимости. Достаточность определяется тем, что определенное вычисление приводит к определенному состоянию сознания (например, определенное вычисление вызывает боль). Необходимость вводится требованием для физической системы поддерживать соответствующее вычисление, которое в свою очередь описывается машиной Тьюринга. Далее в статье Модлин показывает, что одновременно выполнить требования достаточности, необходимости и супервентности никак не удается.

В результате Модлин отказывается от вычислительной теории сознания в пользу материализма. Он считает, что следует считать сознание супервентным непосредственно на возбуждениях нейронов, рассматриваемых как физический процесс. Бруно же остается в рамках метафизики вычислений, а шаг рассмотрения связи сознания, вычислений и физики является для него обоснованием вычислительного монизма — существуют только вычисления, а выполнение алгоритма не связано с материальным носителем.

На седьмом шаге рассмотрения (напомню, что я рассматриваю аргументацию Бруно задом наперед) описывается мир вычислений. Для сравнения полезно начать с мира математических структур в книге Макса Тегмарка ‘*Наша*

математическая вселенная‘:

> ‘Это означает, что наш физический мир не только описывается математикой, но и является математической структурой, что делает нас самосознающими частями гигантского математического объекта.’
>
> ‘Иными словами, IV уровень параллельных вселенных, соответствующий различным математическим структурам, неизмеримо обширнее тех, с которыми мы до сих пор встречались.’
>
> ‘В этом можно усмотреть своего рода радикальный платонизм, согласно которому все математические структуры в платоновском царстве идей существуют где-то в физическом смысле.’

У Тегмарка сложно понять, что конкретно существует в Платонии, содержащией математические объекты. Он рассматривает вычислимые структуры, невычислимые, сложность представления структур, параллельные вселенные, соответствующие разным структурам. Также у него появление сознания в этих структурах описывается исключительно метафорически, например,

> ‘сознание – это способ, каким информация ощущает, что ее обрабатывают’

У Бруно же все достаточно конкретно. Согласно посылке сознание ограничено вычислимыми состояниями, поэтому невычислимые состояния следует отбросить; они не имеют отношение к делу. Далее в Платонии по сути дела остается только ряд натуральных чисел — утверждается, что этого достаточно для запуска машины Тьюринга. Последний шаг связан с введением особого алгоритма — universal dovetailer; мой перевод ниже универсальный переплетатель.

Вычислительное состояние связано с работой алгоритма; существует бесконечное, но счетное количество алгоритмов. В это счетное количество входит алгоритм универсальный переплетатель, который совместно исполняет все алгоритмы. Таким образом, в машине Тьюринга запускается универсальный переплетатель и это обеспечивает шаг за шагом появление всех возможных вычислительных состояний.

Хитрость алгоритма связана с тем, что некоторые алгоритмы зацикливаются, а также количество алгоритмов хотя и счетно, но бесконечно. Проблема решается тем, что универсальный переплетатель запускает конкретный алгоритмы только на определенное количество шагов, а алгоритмы последовательно подключаются в виде треугольника. Это обеспечивает, что в конечном счете все алгоритмы будут исполнены и таким образом решается задача по генерации всех возможных вычислительных состояний в рамках существующего в Платонии ряда натуральных чисел.

В то же время состояние сознание, согласно посылке вычислительной теории

сознания, непосредственно связано с определенными вычислительными состояниями. Более того, у Бруно нет параллельных вселенных, также у него не физика определяет сознание, а сознание, если так можно сказать, исследует физику вычислительных состояний, получающихся универсальным переплетателем. В этом смысле метафизика вычислений оказывается прозрачной и она согласуется с принятыми постулатами вычислительной теории сознания.

Теперь перейду к рассмотрению первых шести шагов аргументации Бруно. Они связаны между собой, поэтому я рассмотрю их вместе. Все шаги связаны с возможностью телепортации, которая представляется как непосредственное следствие вычислительной теории сознания. Если сознание есть последовательность вычислительных состояний, то их можно скопировать и это приведет к появлению аналогичного сознания.

Рассуждение начинается в рамках обыденной жизни, когда человеку предлагается воспользоваться машиной для телепортации. Уверенность в счастливом исходе эксперимента соответствует согласию с вычислительной теорией сознания и именуется «Да, доктор.» Представим себе, что человек сказал «Да, доктор» и его телепортация из одного города в другой прошла успешно. Человек не почувствовал никакой разницы с обычным путешествием, за исключением того, что он заснул в одном городе, а очнулся в другом.

Далее ученые предлагают ему принять участие в следующем эксперименте, когда человека восстановят в двух разных городах одновременно. Такое, конечно, выходит за рамки человеческого общежития, однако технически в рамках принятых гипотез вполне возможно. Человек привержен научному духу и соглашается принять участие в таком странном эксперименте на благо науки.

Для рассмотрения результатов эксперимента необходимо отделить точку зрения от третьего лица (повествование ученых, проводящих эксперимент) от точки зрения от первого лица (повествование участника эксперимента). С точки зрения от третьего лица один человек исчезнет в одном городе, а два совершенно одинаковых человека появятся в городах А и Б. Один из них, оглядевшись по сторонам, скажет: «*Я нахожусь в городе А, но не в городе Б*» , а другой: «*Я нахожусь в городе Б, но не в городе А.*«

В рамках точки зрения от третьего лица не происходит ничего необычного. Парадокс возникает в рамках точки зрения от первого лица — вспомним, что сознание связано именно с этой точкой зрения. Участнику эксперимента задают вопрос: «*В каком городе вы очнетесь после проведения эксперимента?*» Что он может ответить перед проведением эксперимента? Важно отметить, что ответ должен быть дан с точки зрения от первого лица. Следует исключить ситуацию, когда человек посмотрит на будущую ситуацию как бы со стороны и заявит, что оба полученных человека будут точными копиями и поэтому он появится одновременно в обоих городах.

Поэтому для определенности скажем, что в городе А стены комнаты просыпания человека будут покрашены в красный цвет, а в городе Б — в зеленый. Поэтому вопрос человеку будет звучать так: «*Какой цвет стен вы увидите, когда вы проснетесь?*» На этот вопрос уже невозможно отметить таким образом: «*Я очнусь завтра и одновременно увижу красные и зеленые стены.*» Если человек даст такой ответ, то самым разумным будет снятие его с проведения эксперимента и отправление на консультацию к психиатру.

Видны два разумных ответа на вопрос выше: «*Я очнусь завтра в городе А и увижу красные стены*» и «*Я очнусь завтра в городе Б и увижу зеленые стены*» . Парадокс заключается в том, что с точки зрения первого лица невозможно выбрать между двумя альтернативами; все что остается, так это сказать: «*Я не знаю. С вероятностью одна вторая я увижу красные или зеленые стены.*«

Данную ситуацию Бруно называет неопределенностью точки зрения от первого лица (first person indeterminacy) при проведении телепортации. Он отмечает, что неопределенность появляется в рамках полностью детерминированного эксперимента. Далее в следующих шагах Бруно развивает эту ситуацию, когда просыпание в разных городах происходит в разные моменты времени и тому подобное. Он доказывает, что неопределенность точки зрения от первого лица остается во всех рассмотренных случаях.

Неопределенность точки зрения от первого лица приводит при просыпании после телепортрации к появлению дополнительного неалгоритмического бита информации с точки зрения от первого лица — невозможно предсказать, что увидит человек в следующий момент времени. Поскольку сознание связано с точкой зрения от первого лица, то отсюда следует вывод, что при ближайшем рассмотрении телепортация в рамках вычислительной теории сознания невозможна. Параллельные вселенные в такой метафизике также исключаются.

Такой вывод на первый взгляд выглядит странным — начали с одного, пришли к другому. Тем не менее следует вспомнить, что это достаточно типично для научного исследования. Самый распространенный пример связан с вращением Земли вокруг Солнца. При начале наблюдений видно, что Солнце вращается вокруг Земли, но далее после проведенного анализа ученые приходят к противоположному выводу.

Заключительный этап связан с рассмотрением отношений между сознанием в рамках точки зрения от первого лица и вычислениями — каким образом в такой метафизике человек ощущает себя в физическом мире. Отмечу, что это рассмотрение опирается на изощренные методы математической логики, что выходит далеко за рамки моих знаний. Поэтому могу только сказать, что получается в конце концов.

Бруно как логика не особо интересуют проблемы рождения, воспитания и смерти человека, хотя он кратко рассматривает квалиа и у него есть интересные

алгоритмы, связанные с копированием самого себя (амёба) и ростом организма (планария). В любом случае для него самое важное мышление и именно этому вопросу уделяется основное внимание.

Сознание в рамках представленной метафизики можно назвать машиной (последовательность вычислений). Мышление машины связано с математической логикой и наталкивается на предел — теорему Гёделя. Бруно, как и многие другие математики, в отличие от Роджера Пенроуза считает, что теорему Гёделя нельзя использовать для доказательства отличия человека от машины. Теорема Гёделя формальна, поэтому машина может доказать теорему Гёделя и понять ее смысл.

В рамках метафизики вычислений смысл теоремы Гёделя заключается в том, что машина не может доказать, является ли она машиной или нет. Это является интересной чертой представленной картины мира. Бруно утверждает, что можно строго доказать, что в таких рамках эпистемологически нельзя строго доказать верность представленной картины мира.

Таким образом для машины остается возможность занять одну из возможных позиций (Я — машина или Я — не машина), другими словами, делать ставки. Бруно на этом этапе заявляет о машинной теологии (машине не дано знать, является ли она машиной или нет). Бруно тем не менее считает, что теологию можно рассматривать как научную дисциплину и он указывает, что в рамках математической логики и метафизики вычислений наилучшим выбором является картина мира Плотина.

Бруно утверждает, что существуют логики, которое прекрасно соответствуют метафизике Плотина — Единое, Ум и душа. Это должно было бы заинтересовать знатоков Плотина, но, к сожалению, для понимания идей Бруно требуется крайне высокий уровень знания математической логики.

Информация

Bruno Marchal, *The computationalist reformulation of the mind-body problem*, Progress in Biophysics and Molecular Biology, Volume 113, Issue 1, September 2013, Pages 127–140

Трансгуманизм: Парадокс копирования

Tim Maudlin, *Computation and consciousness*, The Journal of Philosophy, v 86, N 8 (1989): 407-432.

Тим Модлин: Вычисления и сознание

Bruno Marchal, *The Amoeba's Secret*, 2014.

Книга о взглядах Бруно, написанная в жанре автобиографии.

Бруно Маршаль: Бессметрна ли амёба?

Обсуждение позиции сторонника сильного ИИ в рамках метафизики Бруно — в конце заметки:

> А. С. Потапов: Искусственный интеллект и универсальное мышление

Статьи Бруно:

> https://www.researchgate.net/profile/Bruno_Marchal3/publications

Обсуждение

https://evgeniirudnyi.livejournal.com/333173.html

19.10.2024 **Математическая структура сознательного опыта**

Увидел статью с выразительным названием ‘*Что такое математическая структура сознательного опыта?*‘:

> ‘(MSC) Математическая структура S является математической структурой сознательного опыта тогда и только тогда, когда выполняются два следующих два условия:
> (S1) Домены A_i из S являются подмножествами A.
> (S2) Для каждого S_j существует S_j-аспект в A.
> Здесь A обозначает совокупность всех аспектов опыта в E; формально A = ∪_e∈E A(e), A_i обозначают области структуры S, а S_j -аспекты определены ниже.’

Johannes Kleiner and Tim Ludwig. *What is a mathematical structure of conscious experience?* Synthese 203, no. 3 (2024): 89.

Интересно, насколько поможет математика включению сознания в объект научного исследования.

https://evgeniirudnyi.livejournal.com/382743.html