УДК: 165.62
DOI: 10.17726/phillT.2018.1.6

Моделирование интеллекта: возможна ли осознанность без репрезентаций?*

Михайлов Игорь Феликсович,
кандидат философских наук, старший научный сотрудник,
Институт философии РАН.
109240, г. Москва, ул. Гончарная, д. 12, стр. 1.

ifmikhailov@gmail.com

Аннотация. С помощью предлагаемого текста я пытаюсь разобраться в проблеме соотношения сознания, осознания и самоосознания с точки зрения возможности их моделирования в искусственных средах. Сознание понимается как *consciousness*, т.е. как совокупность всех тех свойств, которые мы теряем, когда «теряем сознание» Осознание – как *awareness,* т.е. как свойство некоторых когнитивных процессов быть предметом внутреннего контроля. Я показываю, что для осуществления такого контроля необходимы механизмы памяти и внимания. Однако, согласно весьма распространённому мнению, осознание неразрывно связано с самоосознанием – философские основы этого убеждения восходят к Декарту и Канту. Основываясь на концепции репрезентации как необходимо упрощённой формы данных, которыми обмениваются вычислительные – в т.ч. когнитивные – процессы, я показываю нерелевантность этой точки зрения.

Ключевые слова: сознание, осознание, репрезентация, вычисление, проприоцепция, осознание тела, когнитивная наука.

Modeling intelligence: is awareness without representations possible?

Mikhailov Igor F.
Candidate of Philosophy
Senior Researcher, Institute of Philosophy RAS.
109240, Moscow, ul. Goncharnaya, 12, building 1.

ifmikhailov@gmail.com

* Исследование выполнено за счет гранта Российского научного фонда (проект №16-18-10229).

Abstract. With the help of the foregoing text, I try to deal with the problem of the relationship between consciousness, awareness and self-awareness with the view of their modelling in artificial media. Consciousness is understood as the totality of all those properties that we lose, when we 'lose consciousness'. Awareness is the property of some cognitive processes to be the subject to internal control. I show that mechanisms of memory and attention are generally all that is needed for such monitoring. However, according to a very common opinion, awareness is inextricably linked with self-awareness, and the philosophical foundations of this belief date back to Descartes and Kant. Based on the concept of representation as a necessarily simplified form of data exchanged by computing (including cognitive) processes, I show the irrelevance of this point of view.

Keywords: consciousness, awareness, representation, computation, proprioception, bodily awareness, cognitive science.

Настоящая статья представляет собой продолжение концептуальных исследований сознания (consciousness) и осознания (awareness), предпринятых мною в некоторых предыдущих публикациях. Там мне удалось сформулировать подход, в соответствии с которым:

(1) осознанные состояния являются подмножеством сознательных состояний, если под последними понимаются состояния (биологических) систем, в которых они задействуют когнитивные регуляторы поведения, основывающиеся на эффективной обратной связи;

(2) осознанные состояния – это такие сознательные состояния, которые характеризуются существенным участием памяти и внимания;

(3) «я» как представление не является необходимым основанием осознания (в представлении «я мыслю», которое, согласно Канту, сопровождает любую нашу мысль, за собственно «я» не стоит ничего онтологически субстанциального);

(4) осознаваемые репрезентации данных возникают в усложняющихся системах, построенных на многофакторных взаимодействиях, где не оправдывает себя примитивное взаимодействие по схеме «стимул – реакция», и где данные, получаемые извне, должны быть сопоставлены с данными, хранящимися в памяти, и результаты этого сопоставления должны стать предметом внимания.

Осознание должно всегда рассматриваться как интенциональная категория (осознание чего-то), сознание, в отличие от него, – не обязательно.

Многие альтернативные – и, признаем, более популярные – подходы явно или неявно предполагают участие самосознания в осознаваемых процессах. Так, Робетр Арп утверждает, что термин «сознание» (consciousness) – который он понимает скорее как «осознание» (awareness) в моём смысле, – «относится к способности воспринимать себя как существо, тождественное в прошлом, настоящем и будущем, включая рефлексию себя как существа, реагирующего (aware) на окружающую его среду» [1, pp. 102-103]. Иными словами, если я делаю что-то осознанно, то я должен сознавать, что это делаю именно я.

Согласно Арпу, «осознание относится к процессингу, который совершается в результате взаимодействия нервной системы животного (включая сенсорные аппараты) со своей средой, в результате чего этот процессинг результируется в базисной способности животного реагировать на раздражители окружающей среды» [1, p. 103].

Основываясь на здравом смысле, Арп утверждает, что существует иерархия сознательных способностей, начиная от более сознательных существ до менее сознательных или вообще не сознающих. С другой стороны, сознание – в том смысле, который ему придаёт Арп – есть результат сложного взаимодействия областей мультимодальных ассоциаций, по крайней мере, височно-теменной и лимбической коре головного мозга животного, и относится к способности ощущать себя как субъект, единый в прошлом, настоящем и будущем, включая осознание себя как существа, которое осознает свою окружающую среду. «Насколько существо способно инкорпорировать опознание самого себя в свете прошлого, настоящего и будущего, настолько оно может считаться скорее частично или скорее полностью сознательным. Чем больше опыта собственной самости имеет животное в сложных взаимодействиях с другими «я» и объектами в отношении прошлых, настоящих и будущих событий, ситуаций и сценариев, тем более сознательным оно является» [1, p. 103].

Согласно автору, психологический аспект следует рассматривать скорее как род, который включает в себя осознание и сознание как два разных вида, подобно тому, как род млекопитающих включает кошек и людей в качестве отдельных видов. [1, p. 104].

Вопрос, который возникает у меня при чтении Арпа, состоит в следующем: какую роль играют здесь временные измерения? Если мы имеем в виду память, мы имеем в виду ситуацию, когда кто-то может сказать: «Это именно я вчера ел яичницу на завтрак, и именно я завтра буду тоже есть яичницу. Но существенно здесь то,, что эта форма памяти идёт рука об руку с высказыванием – язык снова навязывает нам собственную грамматику, как сказал бы Витгенштейн. В такой ситуации мы описываем человека, который не только осознает окружающую среду, но и переживает «себя» во всех временных формах. Таким образом, мы вводим «я» как важную когнитивную категорию. Но представьте себе биоробота, который запрограммирован только на то, чтобы помнить, что он ел яичницу в прошлом и проецировать эту картину в будущее. Будет ли он существенно отличаться от настоящего человека, который мог бы использовать волшебное слово «я» в таком предложении, как «Вчера у меня была яичница на завтрак. Дорогая, может быть, сегодня у нас будет что-то другое?»

Биоробота, чьё поведение не должно отличаться от человеческого, нужно научить использовать «я» как абстрактный термин для самореференции во всех временных режимах. После этого он также сможет сказать: «У меня вчера была яичница на завтрак», и авторы, такие как доктор Арп, будут описывать это как «опыт собственного «я» в прошлом, настоящем и будущем». Хотя в том, что касается нейронной организации, такой робот может быть не сложнее червя. Червь как таковой не имеет временного опыта собственного «я», потому что он не владеет никаким языком индоевропейского происхождения, где он не мог бы не использовать «я» в разных грамматических временах. Но если вдруг ему понадобится вспомнить свою вчерашнюю еду, чтобы решить какую-либо текущую проблемы, он не будет отличаться от биоробота – а может быть даже, и от человека – с точки зрения структуры опыта. Нет особого опыта собственного «я», кроме лингвистического опыта использования личных местоимений. Ну и, конечно, проприоцепции и других форм восприятия собственного тела.

Статья Роберта Арпа опубликована как дискуссионный ответ на работу [2]. Согласно де Куинси, «сознание» относится к сфере онтологии, «осознание» – к сфере психологии. «Сознание» в этом смысле означает базисную, биологическую способность к чувствительности, переживанию, субъективности, самодеятельности,

интенции или знанию любого рода. Здесь «сознание» используется для обозначения того, что противостоит «бессознательному» – например, состояние бодрствования и бдительности в отличие от сна и сновидений [2, p. 9].

Здесь у меня возникает ряд вопросов. Например:

(1) Является ли осознание функциональным, а не (только) качественным, состоянием?

(2) Является ли осознание интенциональным?

(3) В чем разница между «Я знаю, что p» и «Я осознаю, что p» (или то же в отрицательной форме)?

(4) Действительно ли лунатик не знает, что делает, или мы имеем дело с блокировкой разнличных областей памяти?

(4) Существует ли зависимость между осознанием окружающей действительности и самоосознанием?

На первые два вопроса я ответил бы положительно, остальные требуют скрупулёзного концептуального анализа, для которого рамки настоящей работы слишком тесны. Ограничусь следующими замечаниями.

Случай лунатизма показывает, что осознание подразумевает (и реализует), если можно так сказать, локализацию знаний: например, с точки зрения классической философии нельзя перестать знать что-либо – только опровергнуть или забыть. Если же мы вводим осознание как состояние локальной «подсвеченности» различных областей знания, то тогда прекращение знания становится грамматически возможным – но только с перспективы третьего лица. Таким образом, можно сказать, что существуют разные «поля» или «области» осознания.

Если классический когнитивизм претендует на описание «программной» части сознания, а коннекционизм – на описание «аппаратной» его части, то нужно начать с когнитивного исследования отдельных модулей или блоков, которые включаются / выключаются в разных состояниях осознания (таких как лунатизм, гипноз, раздвоение личности и т.д.). Тогда мы сможем проецировать его результаты на актуальные нейронные сети и таким образом воспроизвести эффект сознательных действий в искусственных устройствах.

Состояние осознания отличается тем, что есть «кто-то», который может изменить ход действий в любое время, кто имеет кон-

троль и может осознанно вмешаться. Но «кто-то» или «я» – это просто общая грамматическая форма для сложной системы с обратной связью и способностью менять свою программу на ходу. Таким образом, сознание проявляется в способности к саморегуляции поведения, а осознание – это состояние «подсвеченности» памятью и вниманием; состояние, которое не предполагает никакого особого «я», кроме грамматического. И де Куинси прав, утверждая, что осознание – это психологическая категория. Поэтому в русском языке мы должны различать то, что мы делаем бессознательно – например, дышим или ходим, – и то, что мы делаем неосознанно – например, рисуем, сидя на совещании. В первом случае имеется в виду неучастие сознания в соответствующих действиях. Во втором – отсутствие состояния осознанности.

Классическая философия основана на двух иллюзиях – чувственной «очевидности» и языке. Эмпирические (когнитивные) науки разоблачают их и показывают, что на самом деле многое обстоит не так, как представляют это нам наши самоочевидности.

* * *

Как правило, подобного рода проблемы имеют как концептуальное, так и эмпирическое измерение. Во втором случае последнее слово должно быть за психологическим экспериментом или нейрофизиологическим исследованием. Однако для их правильной постановки необходима корректная идентификация того, что измеряется или обнаруживается в экспериментальной ситуации.

Попытки локализовать самосознание в мозге имеют долгую историю. Одно время нейрофизиороги отдавали приоритет правому полушарию. По словам Алена Морена, сейчас среди них существует консенсус, что «в обработке самореферентной информации задействуются следующие регионы: корковые медиальные структуры, которые включают в себя вентромедиальную и дорсомедиальную префронтальную кору, предклинье, инсулу, заднюю кору головного мозга, левое и правое височно-теменное соединение и переднюю кору головного мозга» [2, p. 1007]. Тот же автор заключает, что «области мозга, которые поддерживают умозаключения относительно себя, не только расположены в обоих полушариях мозга (не только в правом полушарии) – они разбросаны по всему мозгу и активируются во время других, несаморефлексивных задач» [3, p. 1013].

Гарет Эванс утверждает, что, во-первых, «я» (the self) неуловимо в перцептивном опыте (здесь он апеллирует к Юму, с которым согласен лишь частично), а во-вторых, что осознание тела (bodily awareness) не играет роли в восприятии субъектом своего положения в мире [4].

Вместе с тем, понимание различия сознания и осознания приобретает всё более эмпирически обоснованный характер. Как свидетельствует Крисула Андреу, «неосознанный контроль, отнюдь не будучи оксюмороном, есть психологическая реальность» [5, p. 122].

Игнасио Авила, исследуя осознание тела, приводит интересные примеры: чувство равновесия, вызванное гравитацией, не позволяет рассматривать пол как верхнюю часть комнаты, даже если перевернуться вверх ногами. Аналогично, чувство положения тела даёт разные восприятия объекта (Букингемского дворца в примере, позаимствованном у Кристофера Пикока) в случае, если смотрящий повёрнут к нему всем телом, и в случае, когда тело ориентировано в другую сторону, и на объект смотрят, повернув голову вбок. Он объясняет: «Ясно, что разница в спецификации перцептивной перспективы не определяется разницей в том, как представлен дворец, поскольку в обоих случаях вид абсолютно одинаков» [6, p. 276]. Мысль очевидна: на перспективу оказывает влияние восприятие положения собственнного тела. Для сторонников сенсоримоторного варианта антирепрезентационализма (он же радикальный энактивизм) и положение тела, и самый вид Букингемского дворца – это наборы афордансов, ориентирующих моторную часть психики определённым образом. Для «сильных» репрезентационалистов то и другое – системы репрезентаций, взаимно влияющих (или не влияющих) друг на друга (сравнительный обзор обоих подходов см. в [7]).

Проблема осознания тела и проприоцепции концептуально важна с двух точек зрения. С одной стороны, речь всё же идёт об одной из разновидностей *awareness*, и авторы, пишущие на эту тему специально различают влияние осознанных ощущений собственного тела и его неосознанного восприятия. С другой стороны, проприоцепция и весь набор внутренних телесных ощущений многими рассматривается как часть «самоосознания» (self-awareness), и те или иные психологические и нейрофизиологические данные на этот счёт способны оказать влияние на концептуальные подхо-

ды к самой природе осознанных состояний, а именно – должны ли они с необходимостью включать самоосознание.

* * *

Чтобы оценить сравнительную адекватность конкурирующих когнитивных подходов, нужно сказать несколько слов о проблеме репрезентаций. С моей точки зрения, репрезентации являются необходимым элементом взаимосвязанных вычислительных процессов (ВП), если результаты одних ВП должны быть использованы в качестве данных в других ВП. Вычислительные системы – такие природные или искусственные системы, которые используют естественные процессы на нижних уровнях в качестве атомарных операций для алгоритмических процессов на высших уровнях. Абстракция вычислительного процесса удобна тем, что он может быть описан некоторым алгоритмом, где каждый шаг может быть представлен как подчиняющийся постоянно действующим правилам. Это означает надёжную функциональную зависимость между параметрами «входа» и «выхода», как и между промежуточными состояниями. И если различные ВП (или даже их отдельные этапы) обмениваются репрезентациями своих данных, то мы можем быть уверены, что функциональная взаимосвязь репрезентации и репрезентируемого инвариантна. Таким образом, физическая или феноменальная определённость репрезентации считывается процессом, которому она адресована, как данные, важные в его дальнейших вычислениях.

Когнитивный аппарат человека, животного или робота, согласно наиболее распространённому представлению, представляет собой некоторую иерархию процессов обработки данных, где обработка чаще всего концептуализируется в терминах вычислений, а данные часто называются информацией. В этой системе данные проходят через многие этапы и иерархические уровни обработки, и большая их часть, участвуя так или иначе в работе того, что мы называем сознанием, вполне обходится без осознанных репрезентаций, хотя обрабатывающие их процессы, конечно же, в той или иной форме репрезентируют их друг другу.

Термином «репрезентации» часто обозначают как единичные образы или заместители объектов, так и постоянно действующие системы, обеспечивающие и обслуживающие способы представленности объектов. Так, говоря о репрезентации, лежащей в основе осознания собственного тела (bodily awareness), Фредерик де Виньь-

мон пишет: «Согласно минимальному определению понятия репрезентации, репрезентация тела представляет собой внутреннюю структуру, которая имеет функцию отслеживания и кодирования состояния тела, и которая может исказить его или быть отделена от него» [7]. Во втором случае речь идёт о репрезентативной системе, которая ответственна за формы и способы репрезентаций.

Чтобы объяснить, почему когнитивная наука нуждается в репрезентациях, можно привести «говорящий» пример. Если шахматную доску отобразить на прямой в виде поименованных отрезков (вариант: на ленту машины Тьюринга в виде поименованных клеток), то компьютеру играть будет так же легко или даже легче, а человеку – труднее. Объяснение состоит в том, что вычисления, лежащие в основе (пространственных) восприятий происходят за пределами сознания, а сознанию явлены в качестве «репрезентации», т.е. экономной формы уведомления. Так каждый высший уровень системы освобождается от обработки данных низшего с целью экономии энергетических затрат.

Отношение репрезентации предполагает функциональную связь (иногда сложную и опосредованную) репрезентируемого с репрезентующим. По сути дела, эта связь должна быть достаточно устойчивой, чтобы сделать возможным отношение управления. Я бы предложил понимать управление как воздействие на более сложную систему средствами более простой через абстрагирование, формализацию, репрезентацию и построение каузальных связей между управляющей и управляемой системами. Управление – разновидность многоуровневого системного взаимодействия. Физиологический механизм сокращения мышцы основан на физико-химических процессах, энергия которых адекватна энергии, необходимой для поднятия гирей. Однако выработка мозгом решения о поднятии гири и трансляция этого решения мышцам в виде электрического импульса, идущего по нервным волокнам, энергетически независимы от самого действия.

Описание управленческих действий в терминах формальных операций над «символами», репрезентирующими материальные взаимодействия, и есть то, что лежит в основе метафоры «вычислений» или «процессинга». Динамические системы описывают взаимодействия в рамках закона сохранения энергии. Информационные системы описывают многоуровневые функциональные взаимодействия.

Содержание когнитивной репрезентации – результат многочисленных итераций обучения: действие – входящие данные – реакция – обратная связь – коррекция реакции – запись в память и т.д. В результате (феноменальная) репрезентация входящих данных воспринимается (сознанием) как репрезентация «предмета», который сам по себе есть комплексная репрезентация цели, способов её достижения (афордансов) и препятствий.

Репрезентация в когнитивных актах может присутствовать в трёх смыслах: (1) материальный транспорт феноменального образа, (2) сам феноменальный образ и (2) модуль-зависимый способ представления (цветовой, геометрический, символьный, тональный и т.п.).

Репрезентация – результат, а не предмет вычислений. Температура – репрезентация броуновского движения молекул. Но чтобы она состоялась, необходимы вычисления и *coarse-graining* – буквально «крупнозернение», предельное упрощение и «символизация» репрезентируемой структуры.. *Coarse-graining* как особый ВП имеет целью создание грубых репрезентаций сложных сущностей, чтобы снизить термодинамическую затратность вычислений.

В свете сказанного о репрезентациях нельзя не отметить, что понимание осознания как само-осознания чревато парадоксами. Так, если самоосознание представляет собой осознанную (доступную для внутренних систем контроля) репрезентацию, то оно также с необходимостью должно предполагать самоосознание в качестве своей основы. В результате *explanandum* содержится в *explanans*.

Этот парадокс полностью аналогичен парадоксу гомункула или, в терминологии Дэниела Деннета, юмову парадоксу, который я формулирую в терминах когнитивного классицизма: чтобы «вычислять» символы в соответствии с их семантикой, когнитивная система должна «знать» это семантическое отношение. Но знание есть репрезентация. Т. о., любая репрезентация в рамках семантически-зависимой компьютации нуждается в поддерживающей её репрезентации, и так до бесконечности.

Но представленное здесь понимание репрезентаций приоткрывает ещё один парадокс, который носит скорее не концептуальный, а эмпирический характер.

Компьютер есть многоуровневый механизм, где на нижних уровнях процессор вычисляет, в какие ячейки памяти записать ин-

формацию, а на верхних происходит репрезентация текста и изображений. Но последние адресуются когнитивной системе человека, которая также представляет собой многоуровневый механизм, в котором языковой и визуальный процессоры надстраиваются над многочисленными уровнями более элементарных операций, которые в конце концов сводятся к атомарным – нейронным спайкам. В этом симбиозе очевидно есть лишние звенья, и они так или иначе будут преодолены в ходе технического прогресса.

* * *

Итак, возможна ли осознанность без репрезентаций? Мой ответ – нет, если и поскольку осознанность и осознание есть эффект сложных вычислительных процессов. ВП не обходятся без репрезентаций. Другое дело, что нет никаких концептуальных или эмпирических аргументов в пользу представления, что осознание необходимо предполагает *само*репрезентацию субъекта в том или ином виде.

Литература:

1. Arp, R. Consciousness and Awareness. Switched-On Rheostats: A Response to de Quincey // Journal of Consciousness Studies, 14, No. 3, 2007
2. De Quinsey, C. Switched-on Consciousness. Clarifying What It Means // Journal of Consciousness Studies, 13, No. 4, 2006, pp. 7-12
3. Morin, A.. Social and Personality Psychology Compass 5/12 (2011): 1004-1017, DOI: 10.1111/j.1751-9004.2011.00410.x
4. Evans, G. (1982), The Varieties of Reference. Oxford: Oxford University Press
5. Andreou, C. (2013), AGENCY AND AWARENESS. Ratio, 26: 117-133. doi:10.1111/j.1467-9329.2012.00539.x
6. Ignacio Ávila. Evans on Bodily Awareness and Perceptual Self-Location. DOI: 10.1111/j.1468-0378.2012.00525.x
7. de Vignemont, Frédérique, «Bodily Awareness», The Stanford Encyclopedia of Philosophy (Spring 2018 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2018/entries/bodily-awareness/>.