

Байесовский разум и природа сознания*

© 2020 г. М.А. Сущин

*Институт научной информации по общественным наукам РАН,
Москва, 117997, Нахимовский пр., д. 51/21.*

E-mail: sushchin@bk.ru

Поступила 14.01.2020

В статье оценивается способность программы предсказывающей обработки и байесовских моделей познавательных процессов внести вклад в современные исследования сознания. Автор отталкивается от вопроса, рассмотренного в недавних работах философов Э. Кларка и Н. Блока, а именно: если механизмы восприятия, как предполагается сторонниками байесовских моделей познавательных процессов, являются вероятностными, то почему сознание не отражает в себе характер находящихся в его основании вероятностных перцептивных репрезентаций? При этом, в противоположность Кларку и Блоку, автор указывает на отсутствие оснований полагать, что сознание как высокоуровневый феномен в принципе должно отражать в себе характеристики каких-либо более базовых процессов и репрезентаций. Соответственно предпочтение отдается именно вопросу о возможностях и ограничениях байесовских моделей в когнитивной науке, психологии и нейронауке в деле объяснения человеческого опыта. С целью ответа на этот вопрос в работе в первую очередь исследуется статус самих байесовских моделей познавательных процессов, относительно которого среди исследователей существуют значительные разногласия. Утверждается, что байесовские модели в когнитивной науке, психологии и нейронауке могут формулироваться на всех трех уровнях анализа систем обработки информации по Д. Марру. При этом приоритетным для байесовских моделей является вычислительный уровень по Марру, предполагающий описание вычислительной задачи и доступной организму информации в среде. Тем не менее указывается, что к настоящему моменту применительно к исследованиям сознания преимущественное развитие получили модели, относящиеся к уровню системы репрезентации и алгоритмов по Марру. В заключение выдвигается предположение, что, чтобы охарактеризовать сознание с позиции вычислительного уровня по Марру, программе предсказывающей обработки, вероятно, потребуется интеграция с теорией глобального рабочего пространства.

Ключевые слова: восприятие, репрезентация, сознание, разум, мозг, байесовские модели в психологии и нейронауке, Дэвид Марр.

DOI: 10.21146/0042–8744–2020–6–69–80

Цитирование: Сущин М.А. Байесовский разум и природа сознания // Вопросы философии. 2020. № 6. С. 69–80.

* Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (РФФИ) в рамках научного проекта № 18–311–00138 «Репрезентационная перспектива в современной когнитивной науке и нейронауках».

The Bayesian Mind and the Nature of Consciousness*

© 2020 Mikhail A. Sushchin

*Institute of Scientific Information for Social Sciences of the Russian Academy of Sciences (INION RAN),
51/21, Nakhimovsky av., Moscow, 117997, Russian Federation.*

E-mail: sushchin@bk.ru

Received 14.01.2020

The article assesses the potential of the program of predictive coding and Bayesian models of cognitive processes to contribute to contemporary consciousness studies. The author starts from the question posed in the recent works of philosophers A. Clark and N. Block, namely, if the mechanisms of perception, as suggested by supporters of Bayesian models of cognitive processes, are probabilistic, then why does consciousness not reflect the nature of the probabilistic perceptual representations that underlie it? Contrary to Clark and Block, the author points out that there is no reason to believe that consciousness, as a high-level phenomenon, should in principle reflect the characteristics of some more basic processes and representations. Accordingly, preference is given to the question of the prospects and limitations of Bayesian models in cognitive science, psychology, and neuroscience to help shed light on the problem of human experience. In order to answer this question, the article primarily examines the status of the Bayesian models of cognitive processes, concerning which there is significant disagreement among researchers. It is argued that Bayesian models in cognitive science, psychology, and neuroscience can be formulated at all three of D. Marr's levels of analysis of information processing systems. At the same time, the most important level for Bayesian models is Marr's computational level which gives the description of the computational task and the information in the environment available to the organism. Nevertheless, it is pointed out that concerning consciousness studies models related to the Marr's algorithmic level have been predominantly developed. In conclusion, it is suggested that in order to characterize consciousness at Marr's computational level the program of predictive coding will probably need integration with the theory of global workspace.

Keywords: perception, representation, consciousness, mind, brain, Bayesian models in psychology and neuroscience, David Marr.

DOI: 10.21146/0042-8744-2020-6-69-80

Citation: Sushchin, Mikhail A. (2020) 'The Bayesian Mind and the Nature of Consciousness', *Voprosy Filosofii*, Vol. 6 (2020), pp. 69–80.

Постановка проблемы

Последнее десятилетие развития когнитивной науки было отмечено стремительным развитием байесовских моделей познавательных процессов и деятельности мозга. Сторонники этого направления исходят из того, что механизмы познания работают в примерном соответствии со схемой вероятностного байесовского вывода, а инфор-

* The reported study was funded by Russian Foundation for Basic Research (RFBR) according to the research project No. 18-311-00138 "A representational perspective in modern cognitive science and neurosciences".

мация (прежде всего, сенсорная) репрезентируется в мозге, как гласит известная гипотеза «байесовского кодирования», «в форме распределения вероятностей» [Knill, Pouget 2004, 712]. (Например, зрительная система, с этой точки зрения, репрезентирует расстояние до видимого объекта в форме функции плотности вероятности – то есть, что объект находится на расстоянии x_1 с вероятностью, скажем, 65%, на расстоянии x_2 – с вероятностью 20% и т. д. из ряда возможных расстояний x_n с учетом имеющихся сенсорных данных.)

Важнейшая для этого направления посылка заключается в том, что восприятие и действие осуществляются в условиях существенной неопределенности. Несмотря на то что интроспективно восприятие может представляться весьма определенным и точным, отмечают когнитивные ученые Д. Нилл и А. Пуже, «многие факторы могут ограничивать надежность сенсорной информации о мире – отображение трехмерных объектов в двухмерное изображение, нейронный шум на ранних стадиях сенсорного кодирования» [Knill, Pouget 2004, 712] и т. д. Чтобы преодолеть сенсорную неопределенность мозг, с этой точки зрения, может полагаться на вероятностные байесовские вычисления.

В соответствии с теоремой Байеса, мы определяем вероятность некоторой гипотезы y при условии имеющихся сенсорных данных x (именуемую апостериорной вероятностью гипотезы и обозначаемую как $p(y|x)$ – например, вероятность гипотезы, что воспринимаемый объект является белым медведем), как произведение вероятности сенсорных данных при условии этой гипотезы $p(x|y)$ и вероятности самой гипотезы $p(y)$ (в формальном отношении этот компонент известен как априорная вероятность), деленное на вероятность сенсорных данных $p(x)$. В этой схеме наибольшее значение отводится именно вероятности самой гипотезы $p(y)$ – априорной вероятности, позволяющей отбрасывать заведомо неправдоподобные гипотезы и интерпретации воспринимаемого (скажем, что воспринимаемый объект является снежным человеком). Победившая гипотеза получает так называемый апостериорный максимум и в процессе байесовского вывода начинает выполнять функцию новой априорной вероятности, и далее весь цикл повторяется.

Синтез байесовского понимания восприятия с некоторыми другими предположениями лег в основу новой исследовательской программы в когнитивной науке – так называемой программы предсказывающей обработки / предсказывающего кодирования (predictive processing / predictive coding) [Clark 2013; Hohwy 2013]. Так, наиважнейшую роль для новой программы сыграло, во-первых, как нетрудно догадаться, предположение о том, что познание представляет собой в фундаментальном смысле предсказывающий феномен. А также, во-вторых, гипотеза, что познавательные процессы функционируют по большей мере без участия сознания (что актуализировало классические представления Ибн аль-Хайсама (Альхазена) и Г. фон Гельмгольца о «бессознательных умозаключениях»).

Данная программа и ее основные принципы уже рассматривались нами в некоторых других наших публикациях [Сущин 2017]. В общих чертах, разум, с этой точки зрения, представляет собой иерархически организованную статистическую машину, обладающую моделью мира (именуемой «генеративной моделью») и производящую на основе этой модели предсказания о том, с чем организм может иметь дело в следующий момент времени. Порождаемые генеративной моделью предсказания далее сопоставляются с действительными сенсорными сигналами, и, в случае расхождения между ними, имеющаяся разница (так называемая «ошибка в предсказании») поднимается вверх по иерархии для внесения соответствующих корректив в модель мира с целью порождения более точных сенсорных предсказаний.

К настоящему моменту байесовские модели и программа предсказывающей обработки (ППО) стали законодателями мод в когнитивной науке¹. Несмотря на целый ряд концептуальных и эмпирических проблем, с которыми сталкиваются ППО и байесовские модели в когнитивной науке, на их основе были предприняты попытки объяснения столь разнообразных когнитивных и ментальных феноменов, как моторный контроль, внимание, иллюзии, психические расстройства, эмоции, Я и сознание.

Тема отношения байесовских моделей к сознанию почти синхронно получила освещение в недавних работах двух известных современных философов – Энди Кларка [Clark 2018] и Неда Блока [Block 2018]. Кларк и Блок рассматривают следующую проблему: если механизмы восприятия работают в примерном соответствии со схемой вероятностного байесовского вывода, то почему такого рода вероятностный характер восприятия не отражается в характере субъективного опыта? Или, как гласит название работы Блока, если восприятие является вероятностным, почему субъективно оно не кажется вероятностным, то есть «почему нормальное сознаваемое восприятие не отражает полную функцию вероятности, от которой отталкивается вероятностная точка зрения?» [Ibid.]

Как отмечает Кларк, воспринимаемое представляется в нашем опыте ясным и недвусмысленным образом, а не в виде «мешанины вероятностей». Вслед за Ю. Лю, Т. Стаффордом и Ч. Фоксом Кларк характеризует перцептивный опыт как унитарно-целостный (unitary-coherent), что означает, что, во-первых, у нас может быть только одна перцептивная интерпретация сцены в каждый момент времени (к примеру, при восприятии «Вазы Рубина» мы попеременно видим либо лицо, либо вазу). И, во-вторых, что части воспринимаемых сцен типичным образом находятся в согласии друг с другом (так, в каждый момент мы видим либо лицо, либо вазу, но не некий совмещенный причудливый объект, являющийся наполовину лицом, наполовину вазой).

В двух словах предлагаемое Кларком решение состоит в том, что вызов действия в реальном мире в условиях ограниченного времени ставит перед системой необходимость останавливаться на какой-либо одной из множества соревнующихся интерпретаций воспринимаемых объектов и сцен. Аргумент Блока является более ухищренным и основывается на методологической концепции анализа сложных систем обработки информации пионера вычислительной нейронауки Д. Марра [Марр 1987, 36–44].

Схематично рассуждая, суть концепции Марра заключается в том, что любая система обработки информации может быть рассмотрена на трех уровнях: (1) на уровне вычислительной теории, определяющем общую цель вычислительного процесса, его целесообразность и логику стратегии, обеспечивающей его реализацию, (2) на уровне системы репрезентации входной и выходной информации и алгоритмов для реализации вычислительной цели, а также (3) на уровне физической реализации выбранной системы репрезентации и алгоритмов в мозге. (Для иллюстрации смысла своей концепции Марр использовал пример обычного кассового аппарата. Так, информационная теория нужна для того, чтобы прояснить, что делает кассовый аппарат (выполняет операцию сложения стоимости товаров, подчиняющуюся математически определенной операции сложения) и почему он выполняет именно эту операцию, а не, предположим, умножение (потому что правила объединения цен отдельных товаров подчиняются как раз математически определенной операции сложения). На следующем уровне необходимо выбрать систему репрезентации входной и выходной информации (ею может быть десятичная или двоичная система счисления) и алгоритм преобразования символов. На физическом уровне кассовый аппарат может представлять собой ЭВМ, однако те же операции можно выполнять и с помощью абака или карандаша и листа бумаги.)

Следуя распространенной интерпретации байесовских моделей, Блок утверждает, что они относятся к уровню вычислительной теории по Марру. Следовательно, они могут и не предполагать действительных байесовских алгоритмов в восприятии или действительных вероятностных перцептивных репрезентаций. А поскольку действительных вероятностных репрезентаций в восприятии может и не быть, то и означенная проблема в такой интерпретации может устраниться сама собой.

Между тем рассмотрение проблемы в таком ракурсе, по нашему мнению, ведет к следующему вопросу: почему вообще (перцептивное) сознание как высокоуровневый феномен должно отражать в себе артефакты каких-либо более базовых процессов? Ведь, как известно, любая достаточно сложная система, будь то компьютер или летательный аппарат, на высоких уровнях своей организации способна обретать свой-

ства, которые могут не иметь ничего общего со свойствами составляющих ее более базовых компонентов. Поэтому представляется, что рассматриваемая Кларком и Блоком проблема должна быть переформулирована. С нашей точки зрения, более интересным оказывается вопрос о том, какой вклад ППО и байесовские модели в когнитивной науке могут внести в объяснение сознания человека. В самом деле, каковы возможности и ограничения этой чрезвычайно влиятельной в настоящее время программы в деле объяснения человеческого опыта?

Однако для исследования возможностей ППО пролить свет на природу человеческого сознания необходимо располагать более основательным и ясным пониманием статуса и объяснительных возможностей байесовских моделей в когнитивной науке, психологии и нейронауке как таковых, поскольку между исследователями существуют значительные разногласия по поводу того, как понимать природу байесовского подхода к исследованию познавательных процессов. Таким образом, наша работа будет состоять из двух частей. В первой части мы постараемся внести ясность в вопрос о статусе байесовских моделей в когнитивной науке. Во второй части мы перейдем непосредственно к интересующему нас здесь вопросу о возможном вкладе ППО и байесовских моделей в объяснение перцептивного сознания человека.

Как понимать байесовские модели в когнитивной науке?

Рост популярности байесовских моделей в последнее десятилетие неизбежно повлек за собой дискуссии об их статусе и объяснительных возможностях [Jones, Love 2011; Chater et al. 2011; Bowers, Davis 2012^a; Bowers, Davis 2012^b; Griffiths et al. 2012]. Так, некоторое время назад ряд авторов подверг острой критике байесовские модели в когнитивной науке, психологии и нейронауке. Дискуссия была начата в работе психологов М. Джонса и Б. Лавы [Jones, Love 2011], которые связали значительную часть байесовских моделей в психологии и нейронауке с позицией, обозначенной ими «байесовским фундаментализмом». Так, «байесовский фундаментализм», по Джонсону и Лаву, отличает приверженность к объяснению интеллектуального поведения с позиции того же вычислительного уровня по Марру и методологии рационального анализа по Дж. Андерсону. С этой точки зрения, для объяснения познавательных способностей достаточно корректной идентификации вычислительной задачи, стоящей перед организмом, тогда как исследование конкретных психологических репрезентаций, алгоритмов и нейрофизиологических механизмов, реализующих вычислительную задачу, оказывается необязательным. «Байесовскому фундаментализму», по Джонсу и Лаву, противостоит позиция «байесовского просвещения», заключающаяся в признании необходимости и значимости изучения действительных психологических репрезентаций, алгоритмов и нейрофизиологических механизмов и процессов, учитывающая данные психологии развития и эволюционной психологии наряду с исследованием стоящей перед организмом вычислительной задачи. Между тем, как было указано в ответе ряда ведущих сторонников байесовских моделей в когнитивной науке, позиция «байесовского фундаментализма» является вымышленной и не имеющей своих приверженцев в реальных исследованиях в психологии и нейронауке [Chater et al. 2011].

Продолжая начатую Джонсом и Лавом дискуссию, психологи Дж. Бауэрс и К. Дэвис в своей работе [Bowers, Davis 2012^a] подвергли всесторонней критике байесовский подход в когнитивной науке. С их точки зрения, в зависимости от понимания отношения между вычислительным и алгоритмическим уровнями по Марру байесовский подход в психологии и нейронауке может быть разделен на три группы. Первая группа, именуемая ими «радикальным байесовским подходом» и, по сути, совпадающая с вышеупомянутой позицией «байесовского фундаментализма», имеет чрезвычайно малое количество сторонников. Гораздо более распространенными, по Бауэрсу и Дэвису, являются методологический и теоретический байесовские подходы. Адепты методологического подхода выражают свою приверженность вычислительному уровню по Марру, используя байесовские модели как образец, представляющий меру опти-

мального байесовского поведения, с которой могут быть сопоставлены действительные психологические процессы. Методологический подход не предполагает действительных моделей психологических процессов на алгоритмическом уровне (будучи совместим и с небайесовскими алгоритмами), но и не отрицает необходимости исследований на уровнях алгоритмов и физической реализации. В дополнение к этому приверженцы теоретического байесовского подхода исходят из того, что процессы познания в действительности могут функционировать в примерном соответствии со схемой вероятностного байесовского вывода на алгоритмическом уровне.

Между тем, с точки зрения Бауэрса и Дэвиса, ценность методологического подхода оказывается сомнительной, поскольку находящаяся в его основании стратегия рационального анализа оставляет без внимания значительное количество свидетельств, получаемых в рамках эволюционной психологии, психологии развития и исследований конкретных алгоритмов и нейрофизиологических механизмов познания и разума, тогда как модели психологических процессов, развитые с позиций теоретического байесовского подхода, отличает чрезвычайно слабая эмпирическая поддержка. Причина этого, по мнению Бауэрса и Дэвиса, кроется в значительной гибкости байесовских моделей ввиду того, что ключевые компоненты теоремы Байеса (функция правдоподобия, априорная вероятность, а также добавляемая к ним так называемая функция полезности) могут быть произвольно подстроены под любые данные таким образом, что «если бы оказалось, что данные являются другими, была бы построена отличная байесовская модель, чтобы оправдать заключение, что работа человеческого познания является оптимальной» [Bowers, Davis 2012^a, 410]. (Бауэрс и Дэвис достаточно наглядно демонстрируют свое утверждение на примере байесовских моделей восприятия скорости, распознавания слов и высокоуровневого познания.) Как следствие, многие байесовские модели психологических процессов практически не поддаются фальсификации и надлежащей эмпирической проверке.

Резюмируя дискуссию сторонников и противников байесовских моделей в психологии и нейронауке, представляется возможным сделать несколько основных выводов. Прежде всего, (1) *нужно различать ППО как исследовательскую программу в когнитивной науке или байесовскую программу в психологии и нейронауке и конкретные байесовские модели когнитивных процессов*. Так, с одной стороны, ППО (или классический вычислительный когнитивизм, или коннекционизм, или воплощенное познание [Сушин 2019]) можно рассматривать как своего рода слабый аналог исследовательских программ зрелых естественных наук (прежде всего, конечно, физики), хорошо исследованных в философии науки XX в. [Лакатос 2008]. Как указывал Лакатос, ядро программы по методологическому решению ее сторонников может быть объявлено неопровержимым [Там же, 362]. С другой стороны, на основе исследовательской программы можно конструировать конкретные (байесовские, коннекционистские и т. д.) модели познавательных процессов, и эти модели, как указывали и сторонники байесовского подхода, отвечая своим критикам [Griffiths et al. 2012, 418], вполне могут быть сформулированы таким образом, чтобы быть пригодными для фальсификации.

Далее, мы полагаем, что (2) *приоритетным для байесовских моделей является вычислительный уровень по Марру, предполагающий описание вычислительной задачи и доступной организму информации в среде*. Исследование на этом уровне сосредоточено «на понимании природы среды (например, какая информация доступна организму) и природы задачи, подлежащей выполнению» [Bowers, Davis 2012^a, 390]. Тем не менее это не означает, что байесовские модели только этим уровнем и ограничиваются. Наоборот, (3) *байесовские модели в когнитивной науке, психологии и нейронауке могут развиваться на каждом из уровней по Марру*. Так, возможны специфические, относящиеся к алгоритмическому уровню по Марру, байесовские модели восприятия, моторного контроля, памяти, принятия решения и т. д. Как мы увидим далее, возможны и относящиеся к алгоритмическому уровню байесовские модели механизмов, находящихся в основании того или иного типа субъективного опыта. Равным образом

(несмотря на определенную запутанность этого вопроса [Bowers, Davis 2012^b, 425]), как мы полагаем, могут быть развиты (по крайней мере, принципиальным образом) и байесовские модели в нейронауке, относящиеся к уровню физической реализации в схеме Марра.

Вместе с тем, нужно повториться, для байесовского моделирования в когнитивной науке, психологии и нейронауке, выраженного в духе методологической концепции Марра, характерен акцент на приоритете вычислительного уровня и стратегии «от вычислительной цели к системе репрезентации и алгоритмов к физической реализации». (Как известно, сам Марр подчеркивал наибольшую значимость уровня вычислительной цели: «...алгоритм, – писал он, – вероятно легче понять, исследуя характер решаемой задачи, чем изучая устройство (и его аппаратную часть), в которой он реализуется» [Марр 1987, 43].) Таким образом, байесовские модели в когнитивной науке, психологии и нейронауке преимущественно ориентированы на отыскание универсальных принципов функционирования когнитивных систем или «универсальных законов познания – принципов, которые были в одинаковой степени верными для любых интеллектуальных организмов, где бы то ни было в универсуме» [Griffiths et al. 2012, 418].

Программа предсказывающей обработки и исследования сознания

Вооружившись более ясным пониманием природы байесовских моделей в когнитивной науке, психологии и нейронауке, мы можем попытаться ответить на вопрос об их вероятном вкладе в исследования сознания человека. Первая и достаточно очевидная особенность моделей, отталкивающихся от стратегии «от вычислительной цели к системе репрезентации и алгоритмов, к физической реализации» и направленных на отыскание универсальных принципов функционирования когнитивных систем, состоит в том, что они едва ли способны предсказать конкретные характеристики опыта – скажем, конкретные характеристики осознаваемых зрительных репрезентаций человека или других агентов.

Так, из физиологии прекрасно известны особенности периферии нашего зрительного аппарата (неравномерное распределение на сетчатке фоточувствительных клеток, колбочек и палочек, почти непрерывно иницируемые движения глаз и т. д.), оказывающие влияние на характер непосредственных осознаваемых зрительных репрезентаций (скажем, в силу высокой концентрации колбочек в фовеальной области и их почти полного отсутствия на периферии сетчатки мы способны воспринимать мир в полной цветовой гамме и высоком разрешении только в пределах центральной области зрительного поля)². Совершенно ясно, что другие организмы в силу особенностей своей физиологии могут располагать иными характеристиками опыта³, и описание и исследование влияния такого рода особенностей на опыт конкретного типа организмов должны вестись в первую очередь с позиций физиологии и эволюционной психологии, относящихся к алгоритмическому уровню и уровню физической реализации в схеме Марра, тогда как ППО имеет дело с универсальной схемой, которая может относиться как к человеку, так и к осьминогу, муравьям или пчелам.

Далее, наибольший вклад в исследования сознания ППО и байесовские модели к настоящему моменту внесли на пути разработки моделей субличностных механизмов, находящихся в основании того или иного типа опыта (что соответствует алгоритмическому уровню в концепции Марра). Так, некоторое время назад было представлено несколько моделей, относящихся к этому уровню исследования.

Одна такая в достаточной степени спекулятивная модель была представлена в работе когнитивного ученого С. Куидера и его соавторов [Kouider et al. 2010] в контексте их более общей концепции сознания доступа (сознания, понимаемого как когнитивный феномен, о содержании которого можно вербально отчитаться благодаря его связи с высокоуровневыми когнитивными механизмами рабочей памяти, зрительного внимания, принятия решений, рационального контроля речи и т. д.). В рамках модели Куидера и др. сознание доступа предлагается понимать через иерархию различных

уровней репрезентации: от низких уровней (световой энергии, простых геометрических форм) к более высоким уровням (буквам, словам, значениям и т. д.). Осознание на каждом из этих уровней оказывается независимым от осознания на других уровнях: некто может обладать сознанием низкоуровневых элементов зрительной сцены (цвета, текстуры поверхностей), не обладая более высокоуровневым общим впечатлением об этой сцене и о том, что на ней представлено. При этом на каждом из уровней содержание сознаваемых репрезентаций представляет собой результат взаимодействия трех компонентов: идущих из внешнего мира сигналов, уверенности субъектов, что сигнал идет из внешнего мира, а также предшествующей информации (prior information, соответствующей априорной вероятности в теореме Байеса).

Куидер и др. предлагают свою типологию шести состояний сознания в зависимости от взаимодействия этих компонентов. Так, комбинация сильного сигнала, значительной уверенности в сигнале и слабого влияния предшествующей информации ведет к нормальному восприятию. Сочетание сильного сигнала, слабой уверенности и слабого влияния предшествующей информации может быть свидетельством так называемой «когнитивной слепоты» (также известной как конверсионное расстройство). Если имеется слабый сигнал, а уверенность оказывается сильной и влияние предшествующей информации возрастает, то результатом может быть перцептивная иллюзия. Если же при этом влияние предшествующей информации остается слабым, то имеет место подпороговое восприятие. В случае отсутствующего сигнала, но высокой уверенности в нем и сильного влияния предшествующей информации возникают когнитивные иллюзии (галлюцинации). Наконец, когда сигнал отсутствует, а уверенность и влияние предшествующей информации оказываются слабыми, имеет место восприятие пустоты (perception of absence).

Другой такого рода моделью стала модель весьма интересного и давно известного феномена бинокулярного соперничества, предложенная в работе философа Я. Хохви, антрополога А. Ропстерффа и нейрочеловека К. Фристон [Hohwy, Roepstorff, Friston 2008]. Данный феномен известен, по крайней мере, со времен итальянского врача и философа XVI–XVII вв. Джамбаттиста делла Порта и заключается в следующем. При одновременном предъявлении агенту с бинокулярным зрением двух в достаточной степени различающихся изображений (скажем, изображений лица и дома, либо представленных каждому глазу по отдельности при помощи стереоскопа, либо сделанных полупрозрачными и наложенных друг на друга), за исключением кратких мгновений, когда в сознании может наблюдаться своеобразное смешение двух стимулов, возникает чередование сознательных восприятий каждого из этих изображений в отдельности (то есть большую часть времени агент будет попеременно сознательно воспринимать либо изображение дома, либо изображение лица).

Как отмечают Хохви, Ропстерфф и Фристон, для объяснения феномена бинокулярного соперничества необходимо объяснить два его ключевых аспекта, а именно выбор и чередование. Во-первых, почему в каждый конкретный момент времени сознательно воспринимается только какое-либо одно из двух изображений (либо изображение лица, либо изображение дома) и не возникает (за исключением кратких мгновений) сознательного восприятия обоих объектов одновременно (скажем, смешения двух объектов в один)? Во-вторых, почему сознательное восприятие попеременно «переключается» с одного изображения на другое, а не застывает перманентно на каком-либо одном из них после того, как оно было выбрано?

Для объяснения феномена бинокулярного соперничества авторы предлагают развитую на основе ППО байесовскую модель. В соответствии с данной моделью, первый аспект феномена получает следующее довольно простое объяснение. Если исходить из того, что перцептивная гипотеза дома H и перцептивная гипотеза лица F одинаково соотносятся с имеющимися сенсорными данными (зарегистрированными изображениями дома и лица), то есть, по теореме Байеса, имеют одинаковое значение функции правдоподобия, то выбор между ними будет осуществляться на основе априорной вероятности гипотезы дома H и гипотезы лица F . Если по какой-либо причине ги-

гипотеза лица F имеет более высокое значение априорной вероятности, то именно эта гипотеза получит апостериорный максимум и будет определять сознательное восприятие субъекта. Таким же образом фактор априорной вероятности играет ключевую роль в том, что система останавливает свой выбор на одной из гипотез единичных объектов (гипотезе дома H или гипотезе лица F), а не на гипотезе смешанного объекта «дом-лицо» $H \text{ AND } F$ (даже с учетом того, что эта гипотеза лучше соотносится с имеющимися данными, соответствующими изображениям дома и лица). Причина того, что гипотеза смешанного объекта $H \text{ AND } F$ обладает более низкой априорной вероятностью, чем какая-либо из гипотез единичных объектов, полагают Хохви, Ропстерфф и Фристон, может заключаться в том, что «мозг выучил, что возможна только одна причина сенсорного входа в одно и то же время и в одном и том же месте. <...> Другими словами, априорная вероятность совместного пространственно-временного сосуществования дома и лица является чрезвычайно низкой, в такой степени, что мы едва ли в состоянии поддерживать эту репрезентацию или перцепт» [Hohwy, Roepstorff, Friston 2008, 691].

Второй ключевой аспект феномена бинокулярного соперничества заключается в том, что, после того как сознание остановилось на какой-либо гипотезе единичного объекта, через несколько секунд оно «переключается» на альтернативную гипотезу, а затем обратно и т. д. В рамках модели Хохви, Ропстерффа и Фристана характерное для бинокулярного соперничества чередование сознательных восприятий объясняется следующим образом. Дело в том, что господствующая в данный момент гипотеза может предсказать только часть поступающих на вход сигналов (соответствующую одному из двух изображений), тогда как связанные с другим изображением сигналы остаются незатронутыми ею. Неучтенные господствующей гипотезой сигналы формируют ошибку в предсказании, с целью минимизации которой мозг вынужден обращаться к альтернативной гипотезе. Однако вскоре схожий сценарий повторяется и для этой гипотезы, в результате чего мозг оказывается в своего рода бистабильном режиме, по очереди сознавая представленные изображения. (Глубинным основанием нестабильности ведущей гипотезы, по замечанию авторов, может быть «априорная вероятность высокого порядка ((hyper)-prior), что мир изменяется. Гипотеза статического мира быстро потеряет свою власть в изменяющемся мире» [Ibid., 692].)

Наконец, еще одной байесовской моделью, развитой на данном уровне исследования, стала модель так называемых «позитивных симптомов» шизофрении, предложенная в работе П. Флетчера и К. Фрита. Как отмечают эти авторы, двумя наиболее характерными особенностями шизофрении являются галлюцинации и бредовые представления (delusions), именуемые позитивными симптомами и противоположаемые негативным симптомам (отсутствию мотивации и алогии), которые связаны с отсутствием нормальных функций. При этом специфические именно для шизофрении галлюцинации и бредовые представления отличает особое чувство пассивности – чувство «нахождения под контролем некоей внешней силы или агента» [Fletcher, Frith 2009, 49].

Считалось, что источник галлюцинаций и бредовых представлений кроется в нарушении работы рассматриваемых в изоляции друг от друга систем восприятия и формирования понятий (beliefs). Соответственно, для объяснения позитивных симптомов привлекалась концепция двух нарушений (a two-deficit account). В противоположность изолированному рассмотрению систем восприятия и формирования понятий Флетчер и Фрит отталкиваются от схемы иерархической байесовской предсказывающей обработки, в рамках которой эти две системы понимаются как части единого механизма минимизации ошибки в предсказании. Таким образом, по их мнению, как галлюцинации, так и бредовые представления могут иметь единый общий источник, а именно нарушение в работе системы минимизации ошибки в предсказании.

Флетчер и Фрит полагают, что корнем проблемы может быть нарушение в работе системы выработки нейромедиатора дофамин, действие которого они связывают с оценкой системой точности сигналов об ошибке в предсказании. По их предположению, нарушения выработки нейромедиатора дофамин приводят к тому, что сигналы

об ошибке не оцениваются должным образом (незначительные сигналы могут рассматриваться системой как важные, тогда как действительно значимые сигналы могут игнорироваться). В результате внимание начинает перенаправляться со значимых на несущественные события и вещи. Так, «стимулы могут казаться необычными, значимыми и заметными, приводя к трудностям в уделении внимания соответствующим аспектам среды. Это может переживаться как усиление фоновых звуков или видений, которые нужно игнорировать... сбой в выработке ошибки в предсказании, сопровождающей собственные действия, может вести к тому, что эти действия переживаются как странные и порождаемые извне... стимулы, сопровождаемые большой ошибкой в предсказании, легче ассоциируются между собой, возможно, объясняя странные, но убедительные совпадения, о которых зачастую рассказывают пациенты» [Fletcher, Frith 2009, 55].

В свою очередь, распространение несущественных сигналов об ошибке приводит к искажению модели мира агентов, которая начинает порождать неадекватные выводы и предсказания о том, что происходит в мире. «В конце концов, пациенту с шизофренией потребуется сформировать комплекс убеждений, который должен объяснить значительную часть странных и временами противоречивых данных» [Ibid., 56]. (К примеру, очень часто у больных шизофренией развивается мания преследования, за которой следуют другие драматичные изменения в их взглядах на мир.)

Заключение

Итак, отправной точкой для данной работы послужил вопрос, рассмотренный в недавних работах философов Кларка и Блока. В их формулировке вопрос этот звучит следующим образом: если механизмы восприятия, как гласят чрезвычайно влиятельные в настоящее время байесовские модели познавательных процессов, являются вероятностными, то почему сознание не отражает в себе характер находящихся в его основании вероятностных перцептивных репрезентаций?

Между тем, как мы утверждали, этот вопрос необходимо переформулировать, поскольку нет никаких оснований полагать, что сознание как высокоуровневый феномен в принципе должно отражать в себе характеристики каких-либо более базовых процессов и репрезентаций. Таким образом мы пришли к другому, более интересному, с нашей точки зрения, вопросу о возможном вкладе байесовских моделей познавательных процессов в современные исследования сознания.

Для рассмотрения этого вопроса нам потребовалось прояснить непосредственно характер самих байесовских моделей познавательных процессов, относительно которого среди исследователей существуют значительные разногласия. Так, нами было высказано утверждение, что байесовские модели в когнитивной науке, психологии и нейронауке могут формулироваться на всех трех уровнях анализа систем обработки информации по Марру, с акцентом на приоритет верхнего уровня вычислительной цели и стратегии «от вычислительной цели к системе репрезентации и алгоритмов, к физической реализации». Тем не менее к настоящему моменту применительно к исследованиям сознания преимущественное развитие получили модели, относящиеся к уровню системы репрезентации и алгоритмов по Марру. Данные модели направлены на прояснение характера механизмов, находящихся в основании того или иного типа опыта.

Можно ли охарактеризовать сознание с позиции уровня вычислительной цели по Марру? В соответствии с ведущей современной научной теорией сознания, теорией глобального рабочего пространства (ГРП), сознание представляет собой механизм глобальной трансляции информации для множества работающих без того изолированно когнитивных подсистем мозга (систем речи, принятия решений, планирования и т. д.) [Dehaene 2014]. Поэтому, для того чтобы охарактеризовать сознание с позиции вычислительного уровня по Марру, ППО, вероятно, потребуется интеграция с теорией ГРП.

В настоящее время некоторые исследователи предпринимают попытки интеграции ППО и теории ГРП [Whyte 2019].

Примечания

¹ Здесь необходимо внести ясность: в этой статье речь идет как о распространенных в психологии и нейронауке байесовских моделях, так и о ППО в когнитивной науке. Последнюю мы понимаем, прежде всего, как исследовательскую программу в когнитивной науке. Между тем возможна также байесовская программа в психологии и нейронауке, на основе которой могут быть представлены частные байесовские модели психофизиологических феноменов. С нашей точки зрения, в терминах диаграмм Венна между этими программами возможно отношение пересечения, поскольку не все психологические процессы являются когнитивными (например, эмоции) и не все проблемы когнитивной науки, очевидно, находятся в ведении психологии.

² Подробнее об особенностях периферии зрительного аппарата у человека и возникающих в связи с этим проблемах: [O'Regan 2011, 3–22].

³ К примеру, известная проблема слепого пятна может быть поставлена только в отношении хордовых, у которых фоточувствительные рецепторы находятся за несколькими слоями нервных клеток, для передачи сигналов в мозг собирающихся в пучок и проходящих сквозь сетчатку, благодаря чему и образуется слепое пятно; у головоногих же фоторецепторы ориентированы в сторону источника света, не закрываясь другими типами клеток, поэтому и слепого пятна у них нет.

Источники и переводы – Primary Sources and Russian Translations

Лакатос 2008 – Лакатос И. Фальсификация и методология научно-исследовательских программ // Лакатос И. Избранные произведения по философии и методологии науки. М.: Академический Проект; Трикста, 2008. С. 281–475 (Lakatos, Imre, Falsification and the Methodology of Scientific Research Programmes, Russian Translation 2008).

Марр 1987 – Марр Д. Зрение. Информационный подход к изучению представления и обработки зрительных образов. М.: Радио и связь, 1987 (Marr, David, Vision, Russian Translation 1987).

Ссылки – References in Russian

Сущин 2019 – Сущин М.А. Ситуативное и воплощенное познание как исследовательская программа в когнитивной науке // Наукоедческие исследования. Сб. науч. тр. / Отв. ред. Е.Г. Гребенщикова. М.: ИНИОН, 2019. С. 158–178.

Сущин 2017 – Сущин М.А. Байесовский разум: новая перспектива в когнитивной науке // Вопросы философии. 2017. № 3. С. 74–87.

References

Block, Ned (2018) 'If Perception Is Probabilistic, Why Does It Not Seem Probabilistic?', *Philosophical Transactions of the Royal Society B: Biological Sciences*, Vol. 373, no. 1755, pp. 20170341.

Bowers, Jeffrey S., Davis, Colin J. (2012^a) 'Bayesian Just-So Stories in Psychology and Neuroscience', *Psychological bulletin*, Vol. 138, no. 3, pp. 389–414.

Bowers, Jeffery S., Davis, Colin J. (2012^b) 'Is That What Bayesians Believe? Reply to Griffiths, Chater, Norris, and Pouget', *Psychological Bulletin*, Vol. 138, no. 3, pp. 423–426.

Chater, Nick, Goodman, Noah, Griffiths, Thomas L., Kemp, Charles, Oaksford, Mike, Tenenbaum, Joshua B. (2011) 'The Imaginary Fundamentalists: The Unshocking Truth about Bayesian Cognitive Science', *Behavioral and Brain Sciences*, Vol. 34, no. 4, pp. 194–196.

Clark, Andy (2018) 'Beyond the "Bayesian Blur": Predictive Processing and the Nature of Subjective Experience', *Journal of Consciousness Studies*, Vol. 25, no. 3–4, pp. 71–87.

Clark, Andy (2013) 'Whatever Next? Predictive Brains, Situated Agents, and the Future of Cognitive Science', *Behavioral and Brain Sciences*, Vol. 36, no. 3, pp. 181–204.

Dehaene, Stanislas (2014) *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts*, Viking Press, New York.

Fletcher, Paul C., Frith, Chris D. (2009) 'Perceiving Is Believing: a Bayesian Approach to Explaining the Positive Symptoms of Schizophrenia', *Nature Reviews Neuroscience*, Vol. 10, no. 1, pp. 48–58.

Griffiths, Thomas L., Chater, Nick, Norris, Dennis, Pouget, Alexandre (2012) 'How the Bayesians Got Their Beliefs (and What Those Beliefs Actually Are): Comment on Bowers and Davis', *Psychological Bulletin*, Vol. 138, no. 3, pp. 415–422.

Hohwy, Jakob (2013) *The Predictive Mind*, Oxford, Oxford University Press.

Howhy, Jakob, Roepstorff, Andreas, Friston, Karl (2008) "Predictive Coding Explains Binocular Rivalry: An Epistemological Review", *Cognition*, Vol. 108, no. 3, pp. 687–701.

Jones, Matt, Love, Bradley C. (2011) "Bayesian Fundamentalism or Enlightenment? On the Explanatory Status and Theoretical Contributions of Bayesian Models of Cognition", *Behavioral and Brain Sciences*, Vol. 34, no. 4, pp. 169–188.

Knill, David C., Pouget, Alexandre (2004) 'The Bayesian Brain: the Role of Uncertainty in Neural Coding and Computation', *Trends in Neurosciences*, Vol. 27, no. 12, pp. 712–719.

Kouider, Sid, de Gardelle, Vincent, Sackur, Jérôme, Dupoux, Emmanuel (2010) 'How Rich Is Consciousness? The Partial Awareness Hypothesis', *Trends in cognitive sciences*, Vol. 14, no. 7, pp. 301–307.

O'Regan, Kevin (2011) *Why Red Doesn't Sound Like a Bell: Understanding the Feel of Consciousness*, Oxford University Press, New York.

Sushchin, Mikhail A. (2019) 'Situated and Embodied Cognition as a Research Program in Cognitive Science', E.G. Grebenshchikova (Ed.), *Naukovedcheskiye issledovaniya* (2019), Institute of Scientific Information for Social Sciences, Russian Academy of Sciences, Moscow.

Sushchin, Mikhail A. (2017) 'The Bayesian Mind: A New Theory in Cognitive Investigations', *Voprosy Filosofii*, Vol. 3 (2017), pp. 74–87.

Whyte, Christopher J. (2019) 'Integrating the Global Neuronal Workspace into the Framework of Predictive Processing: Towards a Working Hypothesis', *Consciousness and Cognition*, Vol. 73, pp. 102763.

Сведения об авторе

СУЩИН Михаил Александрович – кандидат философских наук, старший научный сотрудник Центра научно-информационных исследований по науке, образованию и технологиям ИНИОН РАН.

Author's Information

SUSHCHIN Mikhail A. – CSc in Philosophy, Senior Research Fellow at the Center of Scientific Information Studies in Science, Education and Technologies of the Institute of Scientific Information for Social Sciences, Russian Academy of Sciences.