

КЛИНИЧЕСКИЙ ПСИХОАНАЛИЗ

НЕЙРОПСИХОАНАЛИЗ

Сложная проблема сознания и принцип свободной энергии*

М. Солмс

*(Пер. с англ. Елена Калинина;
Науч. ред.: Екатерина Шаповалова)*

Солмс Марк – психоаналитик, нейропсихолог, профессор кафедры нейропсихологии в Университете Кейптауна и клиники Groote Schuur Hospital (отделения психологии и неврологии), президент Психоаналитической ассоциации Южной Африки. Сопредседатель Нейропсихоаналитической ассоциации (NPSA).

Данная статья применяет принцип свободной энергии к сложной проблеме сознания. Прояснив некоторые философские вопросы касательно функционализма, статья идентифицирует аффект как элементарную форму сознания и находит его физиологический механизм (расширенную форму гомеостаза) в верхней части ствола головного мозга. Затем этот механизм формализуется с точки зрения минимизации свободной энергии (в непредсказуемых контекстах), где уменьшение и увеличение ожидаемой неопределенности воспринимается как ощущение удовольствия и ощущение неудовольствия, соответственно. Акцент делается на причинах, по которым такие экзистенциальные императивы ощущаются как нечто по отношению к и для организма.

Ключевые слова: сложная проблема, сознание, свободная энергия, обработка предикций, аффект, Фрейд.

Недавно я опубликовал обширную статью на эту тему (*Solms and Friston, 2018*) – своего рода предварительное сообщение, на котором я хотел бы остановиться здесь перед тем, как будет опубликована книга под названием «Само Сознание» (*Solms*, в печати). Поскольку это психоаналитический журнал, я дополню свою аргументацию перекрестными ссылками на взгляды Фрейда по этим темам. Читателям с математическим образованием будет полезно внимательно прочитать *Solms and Friston (2018)* в сочетании с этой статьей, которая предназначена в первую очередь для аудитории читателей с психологическим образованием.

* Статья опубликована в журнале: *Frontiers in Psychology*, Гипотезы и теория, 30 января 2019, doi: 10.3389/fpsyg.2018.02714.

Мое рассуждение разворачивается в четырех разных по длительности частях. В первой рассматриваются некоторые философские вопросы, относящиеся к двухаспектному монизму в связи со сложной проблемой. Во второй пересматривается анатомическая локализация сознания (так называемый нейронный коррелят сознания, или НКС) в коре головного мозга. В результате переосмысливаются функциональные роли «уровня» и «содержания» сознания. Третья и самая важная часть объясняет двойственные аспекты сознания (его физиологические и психологические проявления) в формально-механистических терминах в связи с императивами минимизации свободной энергии. В четвертой части кратко приводятся некоторые выводы из этой формулировки для когнитивной нейронауки сознания в связи с консолидацией и реконсолидацией памяти.

Проблема со сложной проблемой

Является ли разум производным мозга?

Первоначально положение о сложной проблеме, сформулированное Дэвидом Чалмерсом, выглядит так:

Нельзя отрицать, что некоторые организмы являются субъектами опыта. Но вопрос о том, каким образом эти системы являются переживающими опыт субъектами, озадачивает. Почему, когда наши когнитивные системы участвуют в обработке визуальной и аудиальной информации, мы имеем визуальный или аудиальный опыт: качество темно-синего, ощущение от «до» первой октавы? Как мы можем объяснить, почему есть ощущение, каково это – представить себе психический образ или переживать эмоцию? Несмотря на то что широко распространено мнение, что опыт возникает из физической основы, у нас по-прежнему нет хорошего объяснения, почему и как он возникает. Почему физическая обработка вообще должна давать начало богатой внутренней жизни? Это кажется объективно неразумным, и все же это так (*Chalmers, 1995*).

Более короткое изложение проблемы звучит так: как и почему нейрофизиологическая деятельность *производит* «опыт сознания?» (*Chalmers, 1996*, выделено автором). Джон Сирл говорит нечто подобное: «Как именно нейробиологические процессы в мозге *вызывают* сознание?» (*Searle, 2017*, p. xiii, курсив).

Отправной точкой аргументации, которую я здесь изложу, является то, что мозг не «производит» и не «вызывает» сознание. Формулировка взаимосвязи между мозгом и разумом в причинно-следственных терминах делает сложную проблему сложнее, чем она должна быть. Мозг не производит сознание в том смысле, в каком печень производит желчь, и физиологические процессы не вызывают – или не становятся или не превращаются – в ментальные переживания посредством какой-то любопытной метафизической трансформации.

Когда я просыпаюсь утром и ощущаю себя (мой разум) существующим, а затем подтверждаю в зеркале, что я (мое тело) действительно существую, я просто осознаю одно и то же с двух *разных перспектив*

наблюдения (от первого лица и от второго лица). Спрашивать, как мое тело производит мой психический опыт, – все равно что спрашивать, как молния вызывает гром.

Это двухаспектная монистическая позиция по проблеме разум/тело¹. Конечно, не может быть и речи о том, чтобы определить «правильную» метафизическую отправную точку, но двухаспектная монистическая позиция, которая является отправной точкой моей аргументации, поднимает интересный философский вопрос. Если тело и разум являются двумя проявлениями (асpekтами) одной и той же лежащей в основе вещи, то из какого материала состоит сама лежащая в основе вещь? Другими словами, используя аналогию грома и молнии, каков метапсихологический² эквивалент «электричества» (то есть того, что порождает гром и молнию одновременно)?

Этот вопрос требует прояснения того, что мы подразумеваем (онтологически) под такими терминами, как «физическая основа», «физическая обработка», «нейрофизиологическая деятельность» и «нейробиологические процессы», терминами, которые оказываются на удивление двусмысленными. Если физиологические явления, как и их психические корреляты, являются видимостью, то их основой должно быть нечто нефизиологическое.

Давайте подойдем к этому вопросу на примере. Если внутренний опыт наличия памяти и совокупность нейронов, воплощающих ту же самую память (например, изображенных внешне, с помощью оптогенетики), являются двумя реализациями одной лежащей в основе вещи, то из чего же состоит сама «память»? Ответ заключается в том, что память есть абстракция от обоих проявлений. Память – это не вещь, это *функция*. Мы описываем функции в терминах лежащей в их основе законной механики, а не их внешнего проявления³. Законы выводятся из наблюдаемых нами закономерностей; они *объясняют* внешние проявления.

Конечно, существуют как психологические, так и физиологические объяснения функций памяти; но механизм, который ищет двойственный монист, должен быть достаточно глубоким, чтобы в равной степени

¹ Фрейд был приверженцем двухаспектного монизма (см. *Solms*, 1997). Здесь я пренебрегаю клиническими сложностями, возникающими из-за достижения в развитии способности узнавать себя в зеркале.

² Когда Фрейд впервые ввел этот термин (Письмо Флиссу от 10 марта 1898 года; Фрейд, 1950 [1892–99]), он сказал, что он относится к уровню объяснения, который включает в себя как психологию, так и биологию. Таким образом, он стремился «превратить метафизику в метапсихологию» (Фрейд, 1901, с. 259).

³ Приоритет Фрейда в формулировке этой «функционалистской» позиции не признается: «[Мы] пытаемся сделать понятными сложности психического функционирования при помощи препарирования функции и приписывая ее различные составляющие различным составным частям аппарата. Насколько мне известно, до сих пор не проводился эксперимент с использованием этого метода препарирования для исследования того, как складывается психический инструмент, и я не вижу в этом никакого вреда» (Фрейд, 1900, с. 536, выделено автором).

объяснить оба его наблюдаемых проявления – психологическое и физиологическое. В приведенном выше примере: если мы объясним опыт запоминания с психологической точки зрения и активацию нейронной сборки (и связанных с ней клеточных процессов) с физиологической точки зрения, то наши функциональные выводы будут слишком поверхностными и между ними возникнет «разрыв в объяснении» (*Levine, 1983*). Соответственно, необходимо вывести законы, которые одинаково абстрагированы от двух феноменологических подходов, достаточно глубоко, чтобы обосновать психологические и физиологические составляющие⁴.

Это нетрудно сделать. Рассмотрим, например, кратковременную память (КВП). Закон Миллера утверждает, что люди способны удерживать в рабочей памяти семь плюс-минус две единицы информации в любой момент времени. Это абстракция, полученная как из (психологического) опыта попыток удержать в уме более семи вещей, так и из наблюдения коррелированной (физиологической) синаптической динамики следов КВП (*Mongillo et al., 2008*). То же самое относится и к закону Рибо, касающемуся временного градиента долговременной памяти (ДП), который лежит в основе как психологических, так и физиологических феноменов консолидации памяти во времени (*Kandel et al., 2012*). Эти законы касаются поведения абстрактной функции, которая (сама по себе) является как психологической, так и физиологической. В конечном счете во всех науках мы стремимся свести такие законы к формализованным алгоритмам – к математике – идеалу абстракции от третьего лица⁵.

Вот почему такие термины, как «физическая основа», «нейробиологические процессы» и т. д., удивительно неоднозначны по отношению к психическим функциям. Они предлагают асимметричные (то есть чрезмерно поверхностные) функциональные концепции, которые могут объяснить только неврологическую сторону нейро/психологического уравнения, тем самым оставляя разрыв в объяснении.

Но прежде чем мы сможем определить функциональные законы, лежащие в основе закономерностей как сознательного опыта, так и его нейронных коррелятов, мы столкнемся с еще одним препятствием.

⁴ Фрейд сформулировал это так: «Мы должны представить себе инструмент, который выполняет наши психические функции, похожим на сложный микроскоп или фотографический аппарат или что-то в этом роде. Исходя из этого психическая локальность будет соответствовать точке внутри аппарата, в которой возникает одна из предварительных стадий образа. В микроскопе и телескопе, как мы знаем, они происходят в идеальных точках, областях, в которых не находится никакой осязаемой составляющей инструмента» (Фрейд, 1900, с. 536).

⁵ Такова была цель медицинской школы Гельмгольца: «Мы с Брюке дали торжественную клятву претворить в жизнь эту истину: никакие другие силы, кроме общих физических и химических, не действуют в организме. В тех случаях, которые в настоящее время не могут быть объяснены этими силами, нужно либо найти конкретный способ или форму их действия с помощью физико-математического метода, либо предположить новые силы, равные по достоинству» (*Du Bois-Reymond, 1842*; Письмо к Гальману, в *Du Bois-Reymond, 1918*). Молодой Фрейд был учеником школы Гельмгольца и описывал Брюке как одного из его формирующих образцов для подражания (Фрейд, 1925а).

Является ли сознание просто еще одной когнитивной функцией?

Чалмерс настаивает, что сознание не может быть объяснено с функциональной точки зрения. Он утверждает, что сведение сознания (как мы его воспринимаем) к функциональному механизму никогда не решит сложную проблему:

«Простые проблемы легки именно потому, что они касаются объяснения когнитивных способностей и функций. Чтобы объяснить когнитивную функцию, нам нужно лишь конкретизировать механизм, который может выполнять эту функцию. Методы когнитивной науки хорошо подходят для такого рода объяснений, так же как и для простых проблем сознания. Напротив, сложная проблема сложна именно потому, что она не связана с выполнением функций. Проблема сохраняется даже тогда, когда объясняется выполнение всех соответствующих функций... Что делает сложную проблему сложной и почти уникальной, так это то, что она выходит за рамки проблем, связанных с выполнением функций. Чтобы увидеть это, обратите внимание, что даже когда мы объяснили выполнение всех связанных с опытом когнитивных и поведенческих функций... перед нами все еще остается вопрос без ответа: *почему выполнение этих функций сопровождается опытом?* Простое объяснение функций оставляет этот вопрос открытым... Почему вся эта обработка информации не происходит "в темноте", без каких-либо внутренних ощущений?» (Chalmers, 1995).

В только что процитированном отрывке Чалмерс обращает внимание на то, что сознание – это не просто когнитивная функция. С ним легко согласиться. Все когнитивные функции (такие как память) не являются сознательными по своей сути. Не должно быть «чего-то, на что это похоже», чтобы помнить. Хорошо известно, что обучение и память могут оказывать свое воздействие без какого-либо «внутреннего ощущения»; то же самое относится и к восприятию. Отсюда и название знаменитой обзорной статьи (Kihlstrom, 1996): «Восприятие без осознания того, что воспринимается, обучение без осознания того, что изучается». Единственным исключением из этого правила является именно то, что должно быть объяснено, а именно *сознательная часть* познания – та часть, которая остается после объяснения выполнения всех соответствующих функций.

Почему опыт остается необъясненным, даже когда мы объяснили выполнение всех соответствующих когнитивных функций в его непосредственной близости? Некоторые философы утверждают, что это происходит потому, что «сознание имеет онтологию первого лица, или субъективную онтологию, и поэтому не может быть сведено ни к чему, что имеет онтологию третьего лица, или объективную онтологию» (Searle, 1997, с. 212). Сложная проблема была бы тривиальной, если бы все сводилось к тому, что ваш личный опыт здесь и сейчас не сводится к человеческому опыту вообще. Все, что нужно было бы сделать, чтобы решить проблему, – это взять опыт многих людей, усреднить их, найти общий знаменатель и объяснить *это* в функциональных терминах. Психологи постоянно

занимаются подобными вещами. Но Чалмерс не спрашивает о чем-то столь тривиальном. Он пишет:

«Почему, когда электромагнитные волны воздействуют на сетчатку и различаются и классифицируются визуальной системой, это различие и категоризация переживаются как ощущение ярко-красного? Мы знаем, что сознательный опыт *действительно* возникает, когда эти функции выполняются, но сам факт его возникновения является центральной тайной. Существует *разрыв в объяснении* (термин, предложенный Левином, 1983) между функциями и опытом, и нам нужен объяснительный мостик, чтобы преодолеть его. Простое описание функций остается на одной стороне разрыва, поэтому материалы для мостика должны быть найдены в другом месте» (*Chalmers, 1995*).

Оставляя в стороне кажущееся смешение двух различных видов разрыва в объяснении (между опытом и физиологией, с одной стороны, и опытом и функцией – с другой), теперь становится ясно, почему Чалмерс считает, что даже последний разрыв непреодолим. *Он фокусируется на неправильной функции*. Объяснение опыта никогда не будет найдено в функции зрения или памяти, если уж на то пошло, или в любой другой функции, которая не является по своей сути эмпирической.

Функция опыта не может быть выведена из восприятия и памяти, но она *может* быть выведена из чувства. Не обязательно восприятие или познание сопровождаются опытом «каково это воспринимать и познавать нечто», но кто когда-либо слышал о неосознаваемом чувстве – чувстве, которое вы не можете почувствовать?⁶ Если мы хотим определить механизм, объясняющий феномены сознания (как в его психологическом, так и в физиологическом аспектах), мы должны сосредоточиться на функции чувства, техническим термином для которого является «аффект». Вот почему легко согласиться с тем, что сознание – это не просто еще одна когнитивная функция. Познание издавна отличалось от аффекта, и не без оснований⁷.

Такое внимание к аффекту далеко не случайно.

⁶ Фрейд всегда настаивал на том, что «бессознательный аффект» – это оксюморон: «Несомненно, сущность эмоции состоит в том, что мы должны осознавать ее, т. е. это должно стать известным сознанию. Таким образом, возможность признака бессознательности была бы полностью исключена в том, что касается эмоций, чувств и аффектов» (Фрейд, 1915а, с. 177). Он объясняет: «Вся разница заключается в том, что идеи являются катексисами – в основном следами памяти, – тогда как аффекты и эмоции соответствуют процессам разрядки, конечные проявления которых воспринимаются как чувства. При нынешнем состоянии нашего понимания аффектов и эмоций мы не можем выразить это различие более ясно» (там же, с. 178).

⁷ Стрейчи назвал различие Фрейда между «квантом аффекта» и «следами памяти идей» («memory traces of ideas») «самой фундаментальной из всех его гипотез» (*Strachey, 1962, с. 63*).

В начале был аффект

Является ли сознание функцией коры головного мозга?

Массовые попытки в последнее время идентифицировать нейрокорреляты сознания (НКС) – *Научный Поиск Души*, как памятно называл это Фрэнсис Крик (1994), – использовали зрение в качестве модельного примера. Это было обосновано тем, что детали визуальной обработки лучше понятны, чем таковые для любой другой модальности сознания.

Стратегия Крика состояла в том, что НКС для зрения должен быть обобщен для других форм сознания. Его рассуждения были просты: должно быть возможно выделить что-то, происходящее где-то в зрительном мозге, когда вы видите сознательно, что отсутствует, когда вы видите бессознательно, и это НКС для зрения. Более тщательное изучение этого НКС (чем бы он ни оказался: активацией определенного типа нейрона, или определенной нейронной сети, или определенной полосы частот и т. д.) должно в конечном итоге выявить, как и почему возникает визуальное сознание.

По мнению Чалмерса, стратегия Крика способна решить только легкую (корреляционную) часть проблемы сознания/тела; она не может решить сложную (причинную) часть. Есть по крайней мере еще три проблемы со стратегией Крика.

Во-первых, не может быть никаких объектов сознания без *субъекта* сознания. Вы не можете переживать объекты (визуально или иначе), если вы не находитесь там, чтобы переживать их. Это ставит под сомнение, находится ли сущность сознательного опыта в какой-либо перцептивной модальности. Что, если НКС находится в том, что связывает объекты сознательного восприятия, – в том, кто воспринимает, нежели в восприятиях?

Эта проблема не должна быть фатальной для Крика, если окажется, что переживание возникает из некоторой совокупности или некоторого взаимодействия между различными типами восприятия, как утверждают некоторые теоретики. Переживающий субъект не обязательно должен принимать форму гомункула; он может быть распределен по коре головного мозга и возникать через механизм сродни транскорковой «ассоциации». Именно так видели это немецкие анатомы XIX века, когда они впервые сформулировали кортико-центрическую концепцию сознания по образцу британской эмпирической философии сознания XVII и XVIII веков (см. *Meynert*, 1884).

Это приводит ко второй проблеме стратегии Крика. Когда Мунк (1878, 1881) определил затылочную кору как локус психического аспекта зрения (который, что важно, он, как Мейнерт и британские эмпирики, отождествлял со способностью формировать зрительные «образы памяти», или «идеи», а не просто зрительные ощущения), казалось достаточно разумным обобщить принцип, согласно которому кора является органом «разума», *определенным* таким образом и в других модальностях

восприятия⁸. Последующие экспериментальные результаты подтвердили справедливость этого обобщения (например, абляция слуховой коры [у собак, модельных видов Мунка] вызвала «глухоту разума», точно так же как затылочные поражения вызвали «слепоту разума» – что впоследствии было также подтверждено в клинических случаях у людей; см. *Solms et al.*, 1996).

Если мы отождествим разум с образами памяти (и ассоциациями между ними), то не будет ничего удивительного, если мы узнаем, что, когда современники Мунка аблировали всю кору головного мозга, животные не впадали в кому; вместо этого они стали *амнестическими* (см. *Meynert*, 1884, глава 3, для обзора). Последующие исследования подтвердили это наблюдение у многих видов животных (например, *Huston and Borbely*, 1974). Сознание сохраняется в отсутствие коры головного мозга, как и волевое поведение. Как писали Дамасио и Карвальо (*Damasio, Carvalho*, 2013, с. 147), «декортицированные (не имеющие коры головного мозга) млекопитающие демонстрируют замечательное постоянство последовательного, целенаправленного поведения, которое согласуется с чувствами и сознанием».

Те же факты наблюдаются у врожденно декортицированных (гидранэнцефальных) людей. Учитывая важность этого для нашей темы, приведу пространное описание.

В обстановке домашнего окружения, от которого в решающей степени зависят эти хрупкие с медицинской точки зрения дети, они демонстрируют не только бодрствование, но и тот вид реакции на окружение, который квалифицируется как сознательный по критериям обычного неврологического обследования (*Shewmon et al.*, 1998). Доклад Шьюмона и его коллег является единственным опубликованным отчетом, основанным на оценке способностей детей с гидранэнцефалией в почти оптимальных условиях, и авторы обнаружили, что каждый из четырех детей, которых они оценивали, был в сознании. [...] Чтобы дополнить ограниченную информацию, имеющуюся в медицинской литературе, о поведении детей с гидранэнцефалией, я присоединился ко всемирной интернет-группе самопомощи, созданной родителями и основными опекунами таких детей. С февраля 2003 года я прочитал более 26 000 сообщений электронной почты, передаваемых между членами группы. Из них я сохранил

⁸ В то время «разум» и «сознание» были синонимами. Несмотря на свои многочисленные разногласия с Мейнертом, Фрейд поддерживал точку зрения, что сознание есть не что иное, как «орган чувств для восприятия психических качеств» (1900, с. 615) и, более того, что этот «орган чувств» был расположен в коре головного мозга: «Мы просто приняли взгляды на локализацию, которых придерживается анатомия мозга, которая находит "место" сознания в коре головного мозга – самом внешнем, обволакивающем слое центрального органа. Анатомия мозга не нуждается в объяснении, почему, говоря анатомически, сознание должно располагаться на поверхности мозга, вместо того чтобы быть надежно размещенным – где-то в его глубочайших недрах» (Фрейд, 1920, с. 24). Эта корковая локализация применима даже к аффективному аспекту сознания (см. Фрейд, 1940, с. 161–2).

около 1200 сообщений, содержащих информативные наблюдения или объясняющих инциденты с участием детей. В октябре 2004 года я присоединился к пяти из этих семей на одну неделю в рамках социальной встречи, включающей продолжительные визиты в «Мир Диснея» с детьми в возрасте от 10 месяцев до 5 лет. Я следил и наблюдал за их поведением в ходе многих частных и публичных событий той недели и задокументировал это в четырех часах видеозаписей. Мое впечатление от этого непосредственного контакта с детьми с гидранэнцефалией подтверждает рассказ, приведенный Шьюмоном и его коллегами. Эти дети не только бодрствуют и часто бдительны, но и проявляют отзывчивость к своему окружению в виде эмоциональных или ориентировочных реакций на события окружающей среды [...] Они выражают удовольствие улыбкой и смехом, а отвращение «суетливостью», выгибанием спины и плачем (в разной степени), их лица оживляются этими эмоциональными состояниями. Знакомый взрослый может использовать эту отзывчивость для создания игровых последовательностей, предсказуемо прогрессирующих от улыбки через хихиканье до смеха и сильного волнения со стороны ребенка. Дети по-разному реагируют на голос и инициативы знакомых, проявляют предпочтения к определенным ситуациям и стимулам по сравнению с другими, такими как определенная знакомая игрушка, мелодия или видеопрограмма, и, по-видимому, даже могут ожидать их регулярного присутствия в ходе повторяющихся ежедневных рутин. Хотя поведение варьируется от ребенка к ребенку и с течением времени во всех этих отношениях, некоторые из этих детей могут даже проявлять поведенческие инициативы в рамках строгих ограничений их двигательных нарушений, в форме инструментального поведения: издавая шум, пиная безделушки, висящие в специальной рамке, построенной для этой цели («маленькая комната»), или активируя любимые игрушки переключателями, предположительно основанными на ассоциативном обучении связи между действиями и их последствиями. Такое поведение сопровождается ситуативно соответствующими признаками удовольствия или возбуждения со стороны ребенка, указывающими на то, что оно предполагает согласованное взаимодействие между стимулами окружающей среды, мотивационно-эмоциональными механизмами и телесными действиями. [...] Кроме того, дети подвержены припадкам абсансной эпилепсии. Родители распознают эти провалы доступности в своих детях, комментируя их в таких терминах, как «она ушла разговаривать с ангелами», и они без труда распознают, когда их ребенок «вернулся». [...] Тот факт, что эти дети демонстрируют такие эпизоды, может показаться весомым доказательством их сознательного статуса (Merker, 2007, p. 79).

«Ассоциативное обучение взаимосвязи между действиями и их последствиями» не подразумевает переживания «образов памяти», но следует, конечно, заключить, что само переживание не является функцией коры головного мозга. Основы поведенческой нейробиологии предполагают, что если функция (или критический компонент функции) локализована в определенной структуре, то абляция этой структуры должна привести

к потере этой функции. В случае сознания по отношению к коре головного мозга этот критический тест провален.

Я знаю, что некоторые читатели зададутся вопросом о приведенном выше употреблении термина «сознание» (то есть в каком смысле эти животные и дети «сознательны»); и они могут сослаться на эпистемологическую проблему другого сознания (каким образом мы *знаем*, что они сознательны). Прежде чем ответить на эти вопросы, давайте рассмотрим третью проблему с кортико-центрическим подходом.

Третья проблема заключается в том, что существует структура мозга, которая *действительно* проходит критический тест, только что упомянутый. Эта структура расположена не в коре головного мозга, а в его стволе.

Оригинальные наблюдения были выполнены Моруцци и Магуном (1949) на кошках, а Пенфилд и Джаспер (1954) повторили их на людях. Сознание стирается очаговыми поражениями ствола головного мозга⁹ в области, обычно описываемой как расширенная ретикулоталамическая активирующая система (ERTAS). Недавние исследования показывают, что самые мелкие повреждения в стволе мозга, вызывающие полную потерю сознания (т. е. кому), расположены в парабрахмиальных ядрах моста (pons) или вблизи них (*Parvizi and Damasio, 2003; Golaszewski, 2016*).

Почему же тогда Крик и его последователи не искали НКС в стволе мозга? Ответ таков: по соображениям конвенциональности. После того как Моруцци и Магун не смогли подтвердить главное прогностическое умозаключение, вытекающее из классической теории, а именно что депривация сенсорных входов в кору головного мозга должна привести к потере сознания (например, сна)¹⁰, они не отказались от этой теории; вместо этого они ввели различие между «содержанием» и «уровнем» сознания. Это спасло старую теорию. Таким образом, содержание (*квалиа* сознания) по-прежнему приписывалось коре головного мозга, а новая функция регулирования уровня сознания (*количество* возбуждения или бодрствования, измеряемое по 15-балльной шкале) была приписана системе ERTAS.

Это приписывание существует и по сей день. Поэтому ближайший соавтор Крика, Кристоф Кох, говорит о глубоких ядрах ствола мозга, что «они являются *активаторами* [сознания], но не поставщиками контента» (*Koch, 2004, p. 93, выделено автором*). Это возвращает нас к вопросу, заданному выше: в каком смысле животные и дети, не имеющие коры головного мозга, сознательны? Проявляют ли они *бессмысленное* бодрствование, лишенное содержания и качества, или есть «что-то, на что это похоже», быть ими?

⁹ По иронии судьбы, в свете приведенного выше комментария Фрейда, сознание не находится «на поверхности мозга, [а вместо этого] надежно размещено где-то в его глубочайших недрах» (Фрейд, 1920, с. 24).

¹⁰ См. Мейнерт (1884): «Двигательные эффекты нашего сознания, реагирующего на внешний мир, не являются результатом сил, врожденных в мозге. Мозг, подобно неподвижной звезде, не излучает своего собственного тепла: он получает энергию, лежащую в основе всех мозговых явлений, из мира за его пределами» (перевод с англ., с. 160). Взгляды Фрейда на этот важный вопрос колебались (см. *Solms and Saling, 1990*).

На что это похоже – быть в сознании?

Вывод из изложенного здесь аргумента можно сформулировать заранее: так называемый уровень сознания является функцией вариационной свободной энергии. Свободная энергия в термодинамических терминах влечет за собой энтропию, которая в терминах информационной теории является неожиданностью (и неопределенностью), которая в нейрофизиологических терминах является возбуждением (см. *Solms and Friston, 2018*). Возбуждение лежит в основе бодрствования. Позже эти эквивалентности позволят нам приблизиться к гельмгольцевскому идеалу описания «специфического способа или формы действия [сознания] с помощью физико-математического метода»¹¹. Как говорит Пфафф (*Pfaff, 2006*), «поскольку возбуждение ЦНС зависит от неожиданности и непредсказуемости, его адекватная количественная оценка зависит от математики информации» (с. 13).

Речь идет о природе «сознания», проявляемого животными и детьми без коры головного мозга. В соответствии с тем, что Дамасио и Карвальо (2013) говорили о животных, а также с выводами Шьюмона, Холмса и Бирна (1999), Меркер (2007) заметил, что гидранэнцефальные дети проявляют «эмоциональные или ориентирующие реакции» на события окружающей среды. Более того, «они выражают удовольствие улыбкой и смехом, а отвращение "волнением", выгибанием спины и плачем (во многих градациях), их лица оживляются этими эмоциональными состояниями». Состояния включают в себя «улыбку, через хихиканье, смех и большое возбуждение со стороны ребенка». Эти дети также «проявляют предпочтения к определенным ситуациям и стимулам по сравнению с другими». И их «поведение сопровождается ситуативно соответствующими признаками удовольствия или возбуждения».

С уверенностью можно сделать вывод, что есть ощущение, каково это «быть этими детьми». В соответствии с любым критерием достоверности¹² можно признать, что они, подобно животным без коры головного мозга, проявляют основные эмоции. На самом деле животные-декортикаты проявляют чрезмерную эмоциональность (*Huston and Borbely, 1974*), как и люди, страдающие повреждением корковых структур, которые оказывают тормозящий контроль над ERTAS и лимбической системой (*Harlow, 1868*).

Эти наблюдения могут быть связаны с тем фактом, что глубокая стимуляция (DBS) центрэнцефальных структур, таких как ERTAS и

¹¹ Это также позволит нам приблизиться к безответному стремлению молодого Фрейда (1950 [1895]) «представлять психические процессы как количественно определенные состояния». См. его более раннее замечание о том, что кванты аффекта «обладают всеми атрибутами величины (хотя мы не имеем средств ее измерения), которая способна к увеличению, уменьшению, вытеснению и разрядке» (Фрейд, 1894, с. 60, выделено автором).

¹² Разумный критерий здесь должен быть таким же, как и для любого другого научного вопроса, а именно: являются ли предсказания этой гипотезы (что эти животные и дети сознательны) неподтвержденными или нет? (см. *Panksepp и др., 2016*).

околоводопроводное серое вещество (PAG), а также лимбических контуров, возникающих из них, генерирует мощные аффективные реакции (см. *Panksepp, 1998*, для подробного изучения). Важно отметить, что в отношении вопроса о том, как мы знаем, что эти пациенты сознательны: при глубокой стимуляции мозга человека пациенты декларируют свои субъективные состояния словами (например, *Blomstedt et al., 2007*). В рамках эпистемологической проблемы другого разума (когда никогда нельзя знать достоверно, является ли еще кто-то, кроме самого себя, сознательным) не может быть более высокого стандарта доказательства для вывода о том, что верхний ствол мозга и лимбические контуры порождают аффекты¹³.

Этот вывод подтверждается также тем фактом, что препараты, воздействующие на нейромодуляторы, находящиеся в ядрах ERTAS (серотонин, дофамин, норадреналин, ацетилхолин), оказывают мощное воздействие на настроение, тревожность и т. д. Именно поэтому сегодня они представляют собой основу психофармакологии (*Meyer and Quenzer, 2005*). Другими словами, большинство психотропных препаратов действуют через ERTAS.

Правомерно утверждать, что аффекты *генерируются* в этих подкорковых структурах по той причине, что те же самые эффекты могут наблюдаться и при отсутствии коры. Это противоречит господствующему мнению, что эти ядра просто «позволяют» коре чувствовать (*Koch, 2004*). Примечательно в этой связи, что у пациентов с тотальным разрушением именно тех структур, которые конкретно идентифицируются кортико-центрическими теоретиками с функцией аффекта, а именно префронтальных долей и островковой доли (например, *Craig, 2009; LeDoux and Brown, 2017*), не только сохраняются чувственные состояния, но и, как уже упоминалось, демонстрируется чрезмерная эмоциональность (см. *Damasio et al., 2012*)¹⁴.

Хотя многих когнитивистов все еще нужно отучить от мнения, что кора головного мозга является местом сознания (см. *Panksepp и др., 2016*, для длительного обсуждения этого противоречия), вес доказательств альтернативной точки зрения о том, что процессы возбуждения, генерируемые в верхнем стволе мозга и лимбической системе, «чувствуются как нечто» сами по себе, в настоящее время ошеломляет. В сочетании с огромным

¹³ Конечно, это не означает, что другие структуры также не участвуют, даже (и это важно) включая некоторые за пределами нервной системы. Функция аффекта «локализуется» в общепринятом смысле, которого требует тойберовская парадигма «двойной диссоциации», утверждающая, что если функция А теряется с повреждением структуры X, но не структуры Y, а функция В теряется с повреждением структуры Y, но не структуры X, то функции А и В являются двумя независимыми функциями. (Здесь А = сознание; В = когниции; X = ствол мозга; Y = кора).

¹⁴ Фрейд разделял кортико-центрическую точку зрения, что даже аффекты ощущаются только тогда, когда лежащие в основе «психические энергии» пробуждают то, что он назвал «внутренней поверхностью» системы Pcpt-Cs (восприятие-сознание) в коре головного мозга (см. сноску 8 выше).

количеством доказательств, свидетельствующих о том, что корковые (когнитивные) функции *не* являются действительно сознательными (см. *Bargh and Chartrand, 1999*, для обзора), можно прийти к выводу, что классические немецкие анатомы были правы: кора просто является хранилищем «образов памяти». Кора головного мозга, очевидно, обеспечивает пространство «оперативной памяти» (*Solms and Panksepp, 2012; Ellis and Solms, 2018*). Этот вывод согласуется с радикальной пластичностью коры; настолько, что правое полушарие может полностью взять на себя функции левого, если его удалить достаточно рано (*Pulsifer* и др., 2004); и когда зрительный нерв перенаправляется в слуховую кору, она научается видеть (*Sharma* и др., 2000).

Эта линия мышления будет расширена в части «Сознание возникает вместо следа памяти», где утверждается, что кора *стабилизирует* сознание, а не порождает его; то есть что функционирование коры связывает аффективное возбуждение и тем самым преобразует его в сознательное познание.

Нельзя отрицать, что между корковым типом сознания и сознанием, генерируемым верхним стволом мозга, существует иерархическое отношение зависимости. Это не спорное утверждение; это именно то, что подразумевается под общепринятым утверждением, что ERTAS «делают возможным» сознание. При отсутствии возбуждения ствола мозга не может быть коркового сознания, но обратное неверно. Поскольку эти простые факты соответствуют золотому стандарту анализа нейропсихологических функций, а именно принципу «двойной диссоциации» (*Teuber, 1955*; см. сноску 13), мы должны заключить, что сознание генерируется в верхнем стволе мозга.

Если сознание, генерируемое стволом мозга, является первичным, то сознание фундаментально *аффективно* (см. *Panksepp, 1998; Solms, 2013; Damasio, 2018*). Процессы возбуждения, которые производят то, что условно называется «бодрствованием», составляют основу субъективных переживаний субъекта. Другими словами, *субъективный опыт переживаний вызывается аффектом*.

Данное переформулирование природы первичного сознания имеет серьезные последствия для его функционального механизма, подчеркивая выводы, сделанные в конце раздела «Проблема со сложной проблемой». Вполне разумно спросить, почему обработка визуальной информации не происходит в темноте, без какого-либо внутреннего ощущения, но неразумно спрашивать, почему аффективное возбуждение этого не делает. Как может аффективное возбуждение (то есть активация чувства) иметь место без внутреннего переживания какого-либо чувства?

Почему мы чувствуем?

Современные теоретические попытки ответить на этот вопрос были инициированы Дамасио (*Damasio, 1994*), который отождествил чувство с процессом регистрации состояний тела, в рамках биологической шкалы ценностей, посредством которой приятные и неприятные чувства

регистрируют улучшение и, соответственно, ухудшение шансов на выживание и репродуктивный успех¹⁵. По теории Дамасио, именно поэтому мы чувствуем. Его теория была существенно усилена, когда он включил (*Panksepp*, 1998) выводы о том, что чувства генерируются не в коре головного мозга, а в стволе мозга (и лимбической системе; см. *Damasio*, 2010) и что рассматриваемые нейронные цепочки регистрируют не только состояния здесь-и-сейчас (или состояния «как если бы», *Damasio*, 1994) автономного и сенсорного тела, но также и внутренние состояния мозга: мозговые системы для инстинктов¹⁶, таких как привязанность, ярость и игра (см. *Damasio*, 2018). Сдвиг вниз к стволу мозга позволил Дамасио (как и Панксеппу до него) осознать, что *элементарная форма сознания является чрезвычайно примитивной функцией*. Мой собственный вклад в эти теоретические усилия был сделан относительно поздно (*Solms and Panksepp*, 2012), и они вращались в основном вокруг точной взаимосвязи между гомеостазом¹⁷ и чувством (*Solms*, 2013, 2018a; *Solms and Friston*, 2018).

Уточню: я не утверждаю (как и Панксепп или Дамасио)¹⁸, что чувство возникает из гомеостаза как такового. Я не верю, что термостаты сознательны. Я даже не утверждаю, что все живые существа обладают сознанием (хотя все живые существа обладают гомеостазом). Даже у людей действуют гомеостатические механизмы, полностью лишённые сознания. Регуляция кровяного давления – клинически известный пример. На самом деле можно пойти гораздо дальше: как и Фрейд (и почти все

¹⁵ Данный взгляд не был оригинальным; он почти точно совпадал со взглядом Фрейда на то, что «колебания напряжения инстинктивных потребностей [...] становятся сознательными как чувства удовольствия-неудовольствия» (1940, с. 198). Дамасио (*Damasio*, 1999) признал приоритет Фрейда.

¹⁶ Нет общепринятого определения «инстинкта», но следует отметить, что этот термин используется здесь в основном в биологическом смысле, а не во фрейдистском (который, кстати, возник из-за неправильного перевода немецкого термина *Trieb*; см. *Solms*, 2018 с).

¹⁷ Многие комментаторы забывают, что термин «гомеостаз» был введен в биологию только в 1926 году. Фрейд концептуализировал ту же функцию как «влечение». В связи с этим следующий отрывок (*Solms*, 2013, с. 79–80) служит резюме настоящей статьи: «Я определяю влечение как меру требования психики к работе вследствие ее связи с телом (Фрейд, 1915а, с. 122), где "мерой" является степень отклонения от гомеостатической установки (с последствиями для выживания и репродуктивного успеха). Я не верю, что это отклонение само по себе является чем-то психическим, но "требование", которое оно порождает, ощущается в терминах удовольствия-неудовольствия. Это (ощущаемое требование) и есть аффект, который, на мой взгляд, является источником психического. "Работа", вытекающая из аффекта, – это когнитивная деятельность, функциональное назначение которой состоит в уменьшении аффекта, то есть в уменьшении ошибки предикции (свободной энергии). Цель когнитивной деятельности – привести мир в соответствие с нашими предсказаниями, а наши предсказания – в соответствие с миром. Это первым делом включает в себя обучение».

¹⁸ Дамасио (2018) приписывает чувственные состояния только существам с нервной системой (см. *Solms*, 2018b).

остальные в наши дни), я не утверждаю, что все *психические функции* человека сознательны. Это имеет важное значение для таких философов, как Нагель и Чалмерс, которые иногда забывают, что «субъектность» и «сознание» не являются синонимами (этот факт особенно проблематичен для идеи панпсихизма Чалмерса; см. *Chalmers*, 1995, 1996).

Я утверждаю другое: чувство позволяет сложным организмам регистрировать и тем самым регулировать и определять приоритеты посредством мышления и произвольных действий, отклонения от гомеостатических опорных точек *в непредсказуемых контекстах*. Эта адаптация, в свою очередь, обеспечивает обучение на основе опыта. В предсказуемых ситуациях организмы могут полагаться на автоматизированные рефлексивные ответы (в этом случае биологически жизнеспособные предикции делаются путем естественного отбора и воплощаются в фенотипе; см. *Clark*, 2016). Но если организм собирается сделать адекватный *выбор* в новых контекстах (ср. «свободная воля»), он должен сделать это при помощи *оценки* «здесь и сейчас» относительной *ценности*, придаваемой альтернативным вариантам (см. *Solms*, 2014).

Принципиально важно, что в этом процессе организм должен оставаться «впереди волны» биологических последствий своего выбора (если использовать аналогию, которая дала книге Энди Кларка (2016) замечательное название: *Серфинг неопределенности*):

«Чтобы быстро и свободно справляться с неопределенным и шумным миром, такой мозг, как наш, стал мастером предикции – серфинга по волнам шумной и неоднозначной сенсорной стимуляции, по сути, пытаюсь оставаться прямо впереди того места, где разбивается волна» (с. xiv).

Здесь предлагается предположение, что эта императивная *прогностическая* функция, которая дает адаптивное преимущество, позволяющее организмам выживать в новых условиях, выполняется чувством (см. часть «Чтобы быть точным» ниже для разъяснения ключевой роли *контекста* в приоритизации аффектов и, таким образом, в «наполнении оттенками вкуса» нашего сознания). В настоящем предложении это каузальный вклад квалиа (см. *Solms and Friston*, 2018).

Аффективные квалиа, соответственно, работают следующим образом: отклонение от гомеостатической точки фиксации (увеличение неопределенности) ощущается как неудовольствие, а возвращение к ней (уменьшение неопределенности) – как удовольствие¹⁹. В мозге существует

¹⁹ С точки зрения предложенного здесь, таким образом, в отличие от точки зрения Фрейда, влечение является чувством (*Solms*, 2013). Влечение буквально приводит психику в бытие. До того как влечение ощущается, это не влечение – это простой гомеостаз, который может регулироваться автономными рефлексомы и поведенческими стереотипами. Настоящая точка зрения также отличается от фрейдовской концептуализации удовольствия-неудовольствия как отклонения к и от точки фиксации, в противоположность фрейдовскому континууму, и концептуализации нирваны как этой точки фиксации, а не как чего-то «за пределами» принципа удовольствия (см. *Solms*, 2018b).

множество типов (или «вкусов») удовольствия и неудовольствия (*Panksepp, 1998*)²⁰. Этот тип удовольствия/неудовольствия позволяет идентифицировать затрагиваемую потребность, что помогает организму минимизировать вычислительную сложность (т. е. сосредоточиться на рассматриваемой потребности, а не на состоянии организма в целом – и тем самым минимизировать метаболические затраты; см. *Solms and Friston, 2018*, сноска 7). Все потребности нельзя ощутить сразу. Приоритетность потребностей – то есть определение того, какая потребность будет ощущаться, очевидно, должна в решающей степени зависеть от *контекста* (то есть соотношения потребностей к другим потребностям и соотношения потребностей к возможностям)²¹. Таким образом, чувство расширяется в экстероцепцию (то есть контекстуализируется: «Я чувствую *это об этом*») и трансформируется в когнитивное сознание (то есть оно «связывается»; см. часть «Пробуждение сознания вместо памятного следа»). Это, в свою очередь, порождает добровольное действие – и то, что мы вольно называем *мышлением* – и, в более длительных временных масштабах, обучением на основе опыта (мышление, как учил нас Фрейд, влечет за собой виртуальное действие, а не реальное действие, и тем самым спасает жизни)²².

Сознание (определяемое таким образом) – это биологический императив; это средство, с помощью которого сложные организмы контролируют и поддерживают свою функциональную и структурную целостность в неизвестных ситуациях. По своей сути субъективная и качественная природа этого процесса самооценки объясняет, «как и почему» оно [сознание] ощущается как нечто имеющее отношение к организму, для организма (ср. Нагель, 1974). В частности, увеличение неопределенности по отношению к любому биологическому императиву *просто* «плохо» с точки зрения (от первого лица) такого организма – на самом деле это экзистенциальный кризис, – в то время как уменьшение неопределенности *просто* «хорошо». Это дает очень важную подсказку к тому, как можно решить «сложную проблему». Сознание адаптивно определяет, какие неопределенности должны ощущаться (т. е. выстраиваться в приоритетном порядке) в любом данном контексте. Короче говоря, сознание – это *ощущаемая неопределенность*. Вскоре мы увидим, как и почему возникает перспектива первого лица.

²⁰ Противоречивые требования различных потребностей, которые представляют эти многочисленные «вкусы», лежат в основе психического конфликта и (что не менее важно) объясняют многие формы поведения, которые мы наблюдаем в природе, и в психопатологии, которые ни в коем случае не являются «самосохраняющимися».

²¹ Панксепп (1998) и Меркер (2007) предоставляют убедительные доказательства того, что этот процесс приоритизации вращается вокруг «треугольника принятия решений» среднего мозга (Панксепп называет его «я»), в котором потребности регистрируются в околосинаптическом сером веществе (PAG), а возможности – в верхнем колликуле.

²² См. понятие Фрейда о том, что мышление находится между влечением и действием. Содержание этого параграфа неизбежно чрезмерно плотное. Эти чрезвычайно сложные вопросы требуют больше места, чем позволяет журнальная статья. См. *Solms*, в печати, для более подробного объяснения.

Однако здесь мы должны столкнуться с тем, что философы называют «проблемой представимости».

Можно представить, что функция, которую я только что описал, могла бы выполняться несознательными «чувствами» (см. «философские зомби»), если бы эволюция нашла другой способ для живых существ упреждающе регистрировать и приоритизировать (в отношении себя и для себя) подобные изначально качественные экзистенциальные динамики в условиях неопределенности. Тот факт, что нечто могло бы быть сделано по-другому, не означает, что это не делается так, как это происходит в нервной системе позвоночных. В этом отношении сознание ничем не отличается от любой другой биологической функции. Например, для передвижения не *обязательно* нужны ноги (как сказал Жан-Мартен Шарко, «теория хороша, но она не мешает вещам существовать») (Фрейд, 1893, с. 13). Похоже, аргумент о проблеме представимости возник только потому, что мы искали НКС не в том месте. Можно предположить, что проблема никогда бы не возникла, если бы мы начали с вопроса о том, как и почему чувства (как голод) возникают в связи с потребностями жизни, а не о том, почему опыт связан с познанием.

В следующей части я сведу функцию сознания к его формальной сути. Но я хочу завершить настоящую часть кратким описанием его анатомической реализации.

Ядра мониторинга тела в спинном мозге (спинальные ганглии), верхнего ствола мозга и промежуточного мозга (например, ядро одиночного пути, самое заднее поле, парабрахияльное ядро, циркумвентрикулярные органы и гипоталамус) могут пойти только так далеко в связи с удовлетворением эндогенных потребностей на основе внутренних (автономных) корректировок. Сверх этого предела требуется *внешнее* действие. В этот момент вегетативные рефлекс становятся *влечениями*. То есть interoцептивные (в основном медиальные гипоталамические) «детекторы потребностей» запускают не только вегетативные рефлекс, но также и, в соответствии с критическим процессом приоритизации, выполняемым «треугольником принятия решений» среднего мозга (см. сноску 21 выше), – чувства голода, жажды и т. д. Через последний общий путь возбуждения ERTAS эти влечения обычно²³ запускают дофаминергически опосредованное поведение «добывания» (то есть поведение, которое Панксепп (1998) называет «*поиск*», а Берридж (1996) – «желание»). Добывание отражает филогенетически определенную предикцию, а именно предикцию, что все, что мне нужно, будет найдено там в мире. Различие между «*поиском*» Панксеппа (то есть безобъектным влечением) и «желанием» Берриджа (то есть целенаправленной мотивацией)

²³ Я говорю «обычно», потому что «добывание» обычно является наиболее адаптивным ответом на неопределенность контекста. Однако могут быть выбраны всевозможные другие инстинкты, которые обусловлены обучением через опыт, часто уже не распознаются как инстинкты совсем (см. то, что сказано ниже об обучении по отношению к инстинкту поиска, который служит наглядным примером).

отражает влияние обучения на первичный инстинктивный механизм желания, посредством которого аффективный *поиск* становится когнитивным «желанием» (через соответствие потребности и ее удовлетворения)²⁴. Это облегчает формирование причинно-следственных связей ДП между конкретными потребностями и их адекватными целями и объектами, что, в свою очередь, приводит к итеративному циклу «ошибки предикции вознаграждения», который кодирует непрерывное обучение на основе опыта (см. *Schultz, 2016*).

К счастью, живые организмы не обязаны узнавать все о мире с нуля. Каждый фенотип наделен врожденными предикциями относительно биологически значимых ситуаций, с которыми он обязательно столкнется²⁵. Панксепп (*Panksepp, 1998*) называет это «эмоциональные» и «сенсорные» аффекты (но важно признать, что слово «аффект» оправдано только в той мере, в какой ощущаются соответствующие инстинктивные и рефлексивные предикции, то есть в той мере, в какой они дают остаточную неопределенность, которая требует выбора и обучения на основе опыта). Примерами «эмоциональных» аффектов (каждый из которых отмечен своими собственными ведущими нейромодуляторами и типами рецепторов) являются страх, ярость, привязанность и игра; а примерами таких «сенсорных» аффектов являются боль, удивление и отвращение (см. *Panksepp, 1998*). Например, поведение страха (застывание и бегство) – это врожденные предикции; но каждый человек должен научиться тому, *чего бояться и что еще можно сделать* в ответ. То, что делают позвоночные для удовлетворения своих потребностей, всегда есть сочетание врожденного и выученного поведения.

Остаточная неопределенность (неудовлетворенные потребности, то есть нерешенные проблемы различных типов), возникающая в каждом таком цикле поведения, автоматически оценивается, как описано выше, механизмами, расположенными главным образом в РАГ (околоводопроводном сером веществе) – конечной точке всех аффективных контуров²⁶. Меркер (*Merker, 2007*) соответственно описывает РАГ как часть «синэнцефалического узкого места», где восприятие, действие и аффект объединяются и делается выбор, «что делать дальше»²⁷ (Важно признать,

²⁴ Это, по-видимому, совпадает с концепцией Фрейда (1950 [1895]) о когнитивных эффектах «опыта удовлетворения», т. е. желаемого катексиса и т. д. О роли опиоидов в таких переживаниях см. *Berridge и др. (2009)*. Панксепп (1998) и Шульц (2016) предлагают совершенно разные объяснения роли дофамина.

²⁵ Фрейд поддержал концепцию базовых эмоций, хотя он классифицировал их иначе, чем мы делаем это сегодня, и он объединил их со своей концепцией первичной фантазии, которая повлекла за собой несостоятельное понятие унаследованных эпизодических воспоминаний. (См. Фрейд, 1916–1917, с. 395).

²⁶ Очаговые поражения РАГ вызывают стойкие вегетативные состояния, и глубокая стимуляция мозга в этой области вызывает мощные аффекты различного рода – не только отрицательные – в зависимости от того, какая часть РАГ стимулируется.

²⁷ См. сноску 21 выше. Фрейд (1900) также поместил систему Cs. (Сознание) на моторную часть аппарата, но он, очевидно, имел в виду корковые моторные механизмы.

что конечное расположение PAG в только что описанном цикле делает его функционально «надкорковым», несмотря на то что он *анатомически* подкорковый; см. *Merker, 2007*). Активность PAG, таким образом, приводит к пересмотру выбора восприятия/действия через нейромодуляторные корректировки ERTAS (и более специфических высших лимбических цепочек). Вот как простое чувство становится «чувством об этом»²⁸.

Обратите внимание, что только что описанный цикл оценки включает в себя непрерывную оценку событий окружающей среды и внутренней среды (через ядра мониторинга тела) – оба из которых являются «внешними» для нервной системы, хотя по очевидным биологическим причинам внутренние неопределенности почти всегда будут превосходить внешние (представьте себе последствия недооценки приоритета изменений в оксигенации, гидратации или терморегуляции). Вот почему сознание является квинтэссенцией аффективного.

Если быть точным

Как возникает гомеостаз?

Если сознание возникает через гомеостатический механизм, как предполагают вышеприведенные физиологические²⁹ соображения, то многое зависит от вопроса: как возникает гомеостаз? Ответ на этот вопрос должен привести к абстракции, которую мы ищем (т. е. к абстракции, которая выходит за рамки психологических и физиологических «проявлений»).

По данным Фристонa (*Friston, 2013*), ответом является *минимизация свободной энергии*. Для того чтобы самоорганизующиеся системы, включая все живые существа, такие как мы, существовали, они должны *противостоять энтропии* (количественно выражаемой как свободная энергия, но о важной роли точного взвешивания см. ниже)³⁰. То есть

²⁸ Фрейд (1900) также изобразил функциональное совпадение между Cs. interoцепцией (сознания) и Pcpt. экстероцепцией (восприятие), и в конце концов он объединил эти две системы под одной рубрикой Pcpt.-Cs. (восприятие-сознание) (Freud, 1917). Однако, опять же, он явно имел в виду корковые системы.

²⁹ В этом случае я поставил физиологические соображения выше психологических. Параллельный комментарий в этих сносках обращает внимание на то, что физиологические выводы, к которым мы пришли, сильно напоминают психологические выводы, к которым пришел Фрейд. Для него чувства (принцип удовольствия) были основой психической жизни, включая когнитивную деятельность.

³⁰ Фрейд (1920) заключил эту фундаментальную биофизическую динамику в своей второй теории влечения. До этого он сформулировал его как компромисс – «принцип постоянства», – который, по его мнению, осуществляется сетью «постоянно катектированных нейронов» (Фрейд, 1950 [1895]), «великим резервуаром» его более позднего «эго», «связанная энергия» которого придавала негэнтропийную силу «вторичному процессу».

самоорганизующиеся системы могут сохраняться в течение долгого времени, если будут удерживаться только в «предпочтительных» состояниях, а не рассредоточиваться по всем возможным состояниям и тем самым рассеиваться. Это фундаментальная предпосылка жизни, да и вообще любой самоорганизации. Здесь нам не нужно задаваться вопросом о том, как возникает жизнь. Однако обоснование механизма сознания в его связи с важнейшими предпосылками жизни – неплохая отправная точка, поскольку обычно предполагается, что все сознательные вещи живы, хотя не все живые существа сознательны.

Чтобы система сопротивлялась энтропии, должны быть выполнены три условия. (1) Должна существовать граница, которая отделяет внутреннее и внешнее состояния системы и тем самым изолирует систему от мира. Давайте назовем первые состояния «система», а вторые – «не-система», нежели «мир», по причинам, которые вскоре будут объяснены. (2) Должен существовать механизм, регистрирующий влияние диссипативных (рассеивающих) внешних сил, то есть свободной энергии. Назовем этот механизм «сенсорными состояниями» системы. (3) Должен существовать механизм, противодействующий этим диссипативным силам, то есть связывающий свободную энергию. Назовем этот механизм «активными состояниями» системы, такими как двигательные и вегетативные рефлекс³¹.

Согласно Фристон³² (2013), эти функциональные условия, которые позволяют самоорганизующимся системам существовать и сохраняться во времени, возникают естественным образом (и являются обязательными) в любой эргодической³² рандомной динамической системе, обладающей *покрытием Маркова*³³. Это покрытие устанавливает указанные выше

³⁰ Под этим я подразумеваю способность подавлять нейронный разряд (называемый в терминологии Фрейда «свободно плавающей энергией»), которую он приравнивал к действию Второго закона (см. его принцип «нейронной инерции», прямого предка «влечения к смерти»). В рамках процессов переработки предикций Фристона эта же неэнтропийная сила приписывается скоплениям предиктивных нейронов (которые непосредственно эквивалентны Ψ -нейронам долговременной памяти Фрейда), которые подавляют передачу сенсорных сигналов – Ψ -нейронов кратковременной памяти по Фрейду, – тем самым минимизируя «ошибку предикции» и все энтропийные всплески, которые она порождает, измеряемые как «свободная энергия». (См. *Carhart-Harris and Friston, 2010*).

³¹ См. концепции Фрейда: (1) «Q-экраны» или «барьеры стимулов», (2) «ф-нейроны» или «система Rcpt восприятие» и (3) «M-нейроны» или «система Cs сознание», соответственно. Кстати, большинство исследователей Фрейда, похоже, не понимают, что Q в термодинамике количественно определяет тепло.

³² «Эргодичность» – это статистическое свойство, при котором среднее значение любой измеримой функции случайной динамической системы сходится в одну точку в течение достаточного периода времени. Короче говоря, динамические системы, которые обладают измеримыми характеристиками в течение периодов времени, должны быть (почти) эргодическими.

³³ «Покрытие Маркова» вызывает статистическое разделение внутренних и внешних состояний и скрывает последнее от первого. Само покрытие Маркова состоит из двух наборов («сенсорных» и «активных» состояний), которые влияют друг на друга по кругу: внешние состояния вызывают сенсорные состояния, которые влияют, но не подвержены влиянию внутренних состояний, в то время как внутренние состояния вызывают активные состояния, которые влияют, но не подвержены влиянию со стороны внешних состояний.

граничные условия и является вероятностной конструкцией, которая зависит от того, что влияет на что (и что не влияет на что). Правила причинного влияния Маркова обеспечивают предпосылку (1) разделения между системой и не-системой (т. е. самим покрытием) и снабжают первую (2) рецепторными способностями (сенсорными состояниями покрытия) и (3) эффекторными способностями (активными состояниями покрытия). Важно признать, что эти сенсорные и активные способности являются свойствами покрытия, а не состояний, с которыми они взаимодействуют, что подразумевает, что система, изолированная покрытием Маркова, может «знать» состояния не-системы только *опосредованно*. Другими словами, система может лишь делать предположения о внешних состояниях на основе «чувственных впечатлений» на покрытии Маркова.

На самом деле, для того чтобы противостоять диссипативным силам, *необходимо*, чтобы система делала предположения о внешних состояниях. Это означает, что система должна включать в себя *модель* мира, которая затем становится основой, исходя из которой система действует. Такие модели, как и все модели, несовершенны. Они могут (и должны) улучшаться по мере поступления новых свидетельств. Другими словами, выводы, которые модель генерирует для системы относительно внешних условий (выводы, сформированные на основе сенсорных последствий ее действий), принимают форму предикций, и эти предикции должны постоянно проверяться и пересматриваться³⁴. Таким образом, восприятие и действие влекут за собой непрерывные процессы *проверки гипотез*, в результате чего система со временем обновляет свою модель – свои «убеждения»³⁵. Этот императив негэнтропических самоорганизующихся систем и есть, в двух словах, то, что Фристон называет «активным нахождением выводов». Математически качество этой модели коррелирует с обоснованностью модели, а именно вероятностью сенсорных колебаний согласно модели. В этом случае свободная энергия обеспечивает функцию сенсорных состояний, которые должны уменьшаться при росте обоснованности модели. Другими словами, самоорганизацию – и неявно любую форму гомеостаза – можно рассматривать как минимизацию свободной энергии (или, проще говоря, самосвидетельствование).

Следует добавить, что если рассматриваемая самоорганизующаяся система является нервной системой, то, как бы странно это ни звучало, важно признать, что все остальные системы организма (например, внутренние органы) являются «внешними» по отношению к нервной системе³⁶. Нервные системы воспринимают, представляют и воздействуют на

³⁴ Фрейд назвал бы такие предикции «бессознательными фантазиями».

³⁵ Этот процесс сенсорной выборки напоминает фрейдовский образ системы Ucs (бессознательное), периодически пальпирующей систему Pcpt-Cs (восприятие-сознание) катектическими щупальцами (Фрейд, 1925b, с. 231). «Убеждения» в том смысле, в каком они здесь используются, принимаются за вероятностные распределения, параметры (достаточная статистика которых относится к вероятностным распределениям) или соответствуют состояниям системы.

³⁶ Фрейд (1950 [1895]) говорит о «самом соматическом элементе», порождающем Q (который он называет Qn) в силу «возрастающей сложности внутренних органов организма» (с. 297).

все другие системы организма (как вегетативные, так и сенсомоторные) именно так, как я описал. Нервные системы эволюционировали совместно с другими системами из-за возрастающей сложности организмов, которая (сложность) требует согласования многочисленных гомеостатических требований, возникающих из различных систем. Таким образом, нервные системы являются метасистемами, выполняющими метагомеостатические функции от имени всего организма. Гомеостатическая регуляция организма в целом, можно сказать, делегируется нервной системе.

Таким образом, гомеостаз объясняется каузальной динамикой, обусловленной самим существованием покрытий Маркова; с точки зрения нее самоорганизующиеся системы генерируют тип работы, который связывает свободную энергию и поддерживает систему в типично занимаемых ею («предпочтительных» или «ценных») состояниях. Понятие предпочтительных состояний самоорганизующихся систем тождественно понятию гомеостатических точек равновесия. Нет необходимости воспроизводить здесь математические формулировки, количественно определяющие соответствующую динамику самоорганизующихся систем (см. *Friston, 2013*), поскольку они касаются предпосылок жизни вообще, а не сознания в частности. Уравнения, имеющие критическое значение для наших целей, я представлю в следующей части.

Надеюсь, из вышесказанного ясно, что, хотя я использовал квазифизиологические термины, такие как «сенсорные» и «моторные», и квазипсихологические, такие как «знание», «вывод», «убеждение», «ценность» и «предикция», фактические механизмы, которые я описал, являются одновременно физиологическими и психологическими. Это (их абстрактная онтология) является их главным свойством в свете того, что я сказал в части «Проблема со сложной проблемой». Как мы теперь увидим, одни и те же абстракции могут быть расширены для объяснения функции сознания в обоих его (психологических и физиологических) проявлениях. В самом деле, именно поэтому оправданно использовать для этих механизмов квазифизиологические и квазипсихологические термины.

Теперь мы подходим к сути дела.

Как возникает сознание?

Впервые я высказал мнение в 1997 году, что проблема сознания будет решена только в том случае, если мы сведем ее психологические и физиологические проявления к одной лежащей в основе абстракции (*Solms, 1997*)³⁷. Мне потребовалось много лет, чтобы понять, что эта абстракция вращается вокруг динамики свободной энергии и неопределенности (*Solms, 2013, 2014*).

³⁷ Объединяющей абстракцией Фрейда был «психический аппарат». Философские следствия его часто повторяемого утверждения о том, что инструмент психики бессознателен «сам по себе», недостаточно оценены (см. *Wakefield, 2018*). Отсюда его лаконичное замечание: «Бессознательное является надлежащим посредником между соматическим и психическим, возможно, долгожданным "недостающим звеном"» (Письмо Георгу Гроддеку от 5 июня 1917 года; см. *Groddeck, 1977*).

Минимизация свободной энергии – это основная функция гомеостаза, функция, которая, как показывают мои выводы, а также выводы других исследователей на независимых (клинико-анатомических) основаниях, выполняется теми же ядрами ствола мозга, которые вовлечены решающим образом в генерацию сознания. Другими словами, функции гомеостаза и сознания физиологически реализуются в одной и той же части мозга. Это понимание привело к совместной работе, которая позволила нам с Фристоном расширить вариационную формулировку свободной энергии механизма гомеостаза, чтобы объяснить источник самого сознания (*Solms and Friston, 2018*)³⁸.

Читатели, возможно, уже заметили, что динамика покрытия Маркова формирует два фундаментальных свойства сознания, а именно (элементарные формы) *самости* и *интенциональности*. Верно, что эти динамики также порождают элементарные свойства тел, а именно *изолирующую мембрану* (эктодерму сложных организмов, из которой происходит нервная пластинка) и *адаптивное поведение*. Это замечательный факт. Она лежит в основе двухаспектного монизма.

Часть «Вначале был аффект» была посвящена в основном анатомии и физиологии гомеостаза; теперь мы также проясняем его психологию, объясняя более глубокий механизм. Основой того, что мы называем психологией, является *субъективная* наблюдательная перспектива. Тот факт, что самоорганизующиеся системы должны контролировать свои собственные внутренние состояния, чтобы сохраниться (то есть существовать, выживать), является именно тем, что приводит к *активным* формам субъективности. Само понятие самости оправдывается этим экзистенциальным императивом. Это источник и цель разума.

Самость невозможна, если самоорганизующаяся система не контролирует свое внутреннее состояние по отношению ко внешним диссипативным силам «не-Я». «Я» может существовать только в противоположность «не-Я». Это в конечном счете порождает философскую проблему других разумов. На самом деле свойства покрытия Маркова *объясняют* проблему других разумов: внутренние состояния самоорганизующейся системы могут регистрировать скрытые внешние (не-системные) состояния только опосредованно, через сенсорные состояния их собственного покрытия.

Мы видели, что разум возникает в результате экзистенциального императива самоорганизующихся систем, чтобы контролировать свои собственные внутренние состояния по отношению к потенциально аннигиляционным, энтропийным силам³⁹. Такой мониторинг, по своей сути, является ценностным процессом. Он основан на биологической этике (которая лежит в основе всей эволюции) в том смысле, что выживание –

³⁸ Когда мы это сделали, я испытал нечто похожее на то, что описал Фрейд более века назад, когда он писал: «Все, казалось, сходилось вместе, шестеренки были в сетке, вещь создавала впечатление, что это действительно машина и скоро будет работать сама по себе». [...] Конечно, я не могу сдержать восторга» (Письмо Флиссеу от 20 октября 1895 года; Фрейд, 1950 [1892–99]).

³⁹ См. формулировку Фрейда о нарциссизме («ненависть как отношение к объектам старше любви»; Фрейд (1915b), с. 139), который стал основой «параноидно-шизоидной позиции» Мелани Кляйн.

это «хорошо». Этот императив формализуется с точки зрения минимизации свободной энергии.

Такая негэнтропийная динамика самоорганизующихся систем является абсолютной предпосылкой для эволюции разума. Однако в этой динамике нет ничего, что отличало бы сознательные психические процессы от бессознательных. Иными словами, в такой протоментальной динамике нет ничего, что объясняло бы возникновение чувства, в отличие от жизненных потребностей. Верно, что динамика, описанная выше, вращается вокруг ценности, но ценности, о которых идет речь, в принципе, все еще можно выразить в чисто количественных терминах (например, $10 > 9$). Нет необходимости вводить качественные термины в динамику минимизации свободной энергии.

Что же тогда лежит в основе перехода от бессознательных (количественных, «протоментальных») состояний к сознательным (качественным, истинно «психическим»)? Похоже, что переход в основном вращается вокруг возрастающей сложности. Однако это относится к сложности конкретного типа, не только к сложности комплексной обработки информации в целом (ср. Тонони, 2012). С точки зрения самосвидетельствования сложность приобретает весьма конкретное значение⁴⁰ (это следует из того, что обоснованность модели – это разница между точностью и сложностью. Поскольку обоснованность модели активно увеличивается за счет минимизации свободной энергии, точность предикций повышается с сопутствующим увеличением сложности. Другими словами, увеличение сложности модели всегда сопровождается способностью делать более точные предикции).

Организмы эволюционируют, увеличивая собственную сложность, – по очевидным адаптивным причинам – по мере того как они диверсифицируются в многочисленные подсистемы (и разделяют вегетативный труд между ними). Например, они развивают пищеварительную, дыхательную, терморегуляторную и иммунную системы. Каждая такая специализированная система управляется собственным гомеостатическим императивом. Метаболический энергетический баланс, оксигенация, гидратация и терморегуляция (например) – это не одно и то же, хотя каждый из них вносит свой вклад в общий императив минимизации свободной энергии в организме. Если различные требования специализированных гомеостатических систем будут вычисляться дифференцированно (как они и должны), то из этого следует, что возрастающая сложность требует некоторой формы разделения количественных величин. Такое разделение может быть достигнуто только посредством некоторой формы *качественной дифференциации* между наборами переменных (например, $10 \times X$ стоит больше, чем $10 \times Y$; где X и Y – *категориальные переменные*). Можно думать об этом разделении как о чем-то сродни «цветовому кодированию» или «вкусам» различных наборов данных. Это проявляется во множестве различных облиций – от функциональной специализации в нейронных

⁴⁰ Технически это относительная энтропия между последующими и предшествующими убеждениями или распределениями вероятностей по внешним или не-собственным состояниям.

системах до факторизации фундаментальных конструкций, которые мы используем для моделирования мира (например, системы «что» и «где» в мозге). Как отмечалось выше, обоснованность модели – это разница между точностью и сложностью, которая требует нюансирования при увеличении сложности (ср. принцип Оккама). Разделение позволяет проще представить то, что происходит «снаружи», в терминах внешних или «не-Я» состояний. Важно отметить, что такого рода разделение существенно важно для моделей, обобщающих новые ситуации.

Другими словами, потребность в разделении становится необходимостью, когда относительная величина различных величин *изменяется* с течением времени. Например: голод превосходит усталость до определенного значения, после чего усталость превосходит голод; или голод превосходит усталость в определенных обстоятельствах, но не в других (например, $10 \times X$ в настоящее время [но не всегда] стоит больше $10 \times Y$). Такие изменения требуют, чтобы система не только разделила свои рабочие усилия по отношению к своим различным потребностям, но и *определила их приоритетность* с течением времени.

Этот императив достигает своего апогея в активных состояниях системы, которые неизбежно порождают узкое место. Например, организмы не могут есть и спать одновременно. Точно так же они не могут одновременно поворачиваться налево и направо. Когда дело доходит до действий, должен быть сделан решающий выбор.

Все эти *контекстуальные* факторы становятся более предвидимыми, когда мы рассматриваем также, как организмы выживают в новых (непредсказуемых) средах. Вполне возможно, что чрезвычайно сложный набор алгоритмов может эволюционировать (независимо от того, насколько громоздкими они могут стать), чтобы вычислять относительные потребности выживания во всех предсказуемых ситуациях и определять приоритеты действий на этой основе. Но как организм выбирает между X и Y, если последствия этого выбора непредсказуемы? Физиологические соображения, рассмотренные в предыдущей части, предполагают, что он делает это, *чувствуя* свой путь через проблему, где направление чувства (удовольствие либо неудовольствие), в соответствующей модальности, предсказывает направление ожидаемой неопределенности (уменьшение либо увеличение) в пределах этой модальности⁴¹.

При выборе наилучшего образа действий мы должны обратиться к нашей модели мира, чтобы предсказать последствия некоторого поведения

⁴¹ Общим источником путаницы здесь является тот факт, что дофаминергическая модальность поиска (обсуждаемая в части «В начале был аффект») положительно взаимодействует с неопределенностью. Его сущностная недеklarативная предикция переводится как: «взаимодействие с источником неопределенности предоставляет максимальные возможности для разрешения этой неопределенности». Поэтому в случае этого инстинкта отсутствие взаимодействия с неопределенностью – это «плохо» (см. анергия, абулия, ангедония, безнадежность). Концептуальное различие в аффективной нейронауке нашего времени между «аппетитными» и «завершающими» удовольствиями устраняет источник озадаченности Фрейда в его пожизненных попытках установить психофизику удовольствия-неудовольствия в связи с колебаниями напряжения потребностей влечения.

с точки зрения ожидаемой свободной энергии. *Ожидаемая свободная энергия как раз и есть неопределенность относительно последствий любого предполагаемого действия.* Поэтому императив минимизации ожидаемой свободной энергии становится необходимым для выбора действий, которые минимизируют неопределенность и реализуют привычные, предпочитаемые сенсорные состояния.

Прежде чем мы рассмотрим, что это может повлечь за собой в формальных, математических терминах, я хочу пояснить, что эволюционные соображения, которые мы только что рассмотрели, предполагают *постепенный* переход от протоментальных к ментальным состояниям (то есть от бессознательной к сознательной субъективности). Субъективные ценности (то есть системно-центрические ценности) вычисляются уже на уровне вегетативного гомеостаза. Это подразумевает возможность гедонической валентности. Но качественно ощущаемый аспект гедонической ценности *не должен* регистрироваться самоорганизующейся системой до тех пор, пока множество таких ценностей не будет дифференцированно вычислено и расставлено по приоритетам в изменчивых и новых контекстах, где неопределенность сама по себе становится основным фактором выбора действия.

Вычислительно такие контекстуальные факторы формализуются с точки зрения точного взвешивания. «Точность» – чрезвычайно важный аспект активного и перцептивного предположения; это *репрезентация неопределенности*. Точность, связанная с величиной, оценивает ее надежность или обратную дисперсию (например, визуальные – относительно слуховых – сигналы получают большую точность в дневное время по сравнению с ночным). Эвристически точность можно рассматривать как уверенность, которую дают вероятностные представления о состояниях не-системы – или, что более важно, о том, какие действия «я должен выбрать».

Это фундаментальный момент, сформулированный в *Solms and Friston (2018)*. Мы пришли к выводу, что в то время, как гомеостаз требует не более чем постоянной корректировки активных состояний системы (M) и/или предположений о ее сенсорных состояниях (φ), в соответствии с ее прогностической моделью (ψ) внешнего мира (Q) или вегетативного тела (Qn), которая может быть скорректирована автоматически на основе текущих регистраций ошибки предикции (e), количественно выраженной как свободная энергия (F), – рассмотренные только что контекстуальные соображения требуют дополнительной способности корректировать точное взвешивание (ω) всех соответствующих величин. Эта способность обеспечивает формальное (механистическое) объяснение произвольного поведения – выбора.

При наличии указанных выше величин⁴² можно описать любую самоорганизующуюся (т. е. самосвидетельствующую) систему следующей динамикой (схемы 1, 2).

Где свободная энергия и ошибка предикции обозначаются так (схема 2).

Более подробное описание мышления, лежащего в основе этих уравнений широкого мазка кисти, можно найти в *Solms and Friston (2018)* и в справочных материалах, содержащихся в этой статье.

Физиологически точность обычно связана с *постсинаптическим усилением* корковых нейронов, сообщающих об ошибках предикции. Это как раз и есть функция ERTAS – модуляторных нейронов (см. часть «В начале был аффект»). В этом смысле точность может быть связана – через минимизацию свободной энергии – с *избирательным возбуждением* (и, таким образом, как формализовано тремя зависимостями в уравнении 1, с действием [1a], восприятием [1b] и аффектом [1c], соответственно).

$$\frac{\partial}{\partial t} M = -\frac{\partial F}{\partial M} = -\frac{\partial F}{\partial e} \frac{\partial e}{\partial M} = \frac{\partial \Phi}{\partial M} \cdot \omega \cdot e \quad (1a)$$

$$\frac{\partial}{\partial t} Q = -\frac{\partial F}{\partial Q} = -\frac{\partial F}{\partial e} \frac{\partial e}{\partial Q} = -\frac{\partial \psi}{\partial Q} \cdot \omega \cdot e \quad (1b)$$

$$\frac{\partial}{\partial t} \omega = -\frac{\partial F}{\partial \omega} = \frac{1}{2} \cdot (\omega^{-1} - e \cdot e) \quad (1c)$$

Схема 1

$$F = \frac{1}{2} \cdot (e \cdot \omega \cdot e - \log(\omega)) \quad (2)$$

$$e = \Phi(M) - \psi(Q) \quad (3)$$

Схема 2

⁴² ω – точность; ψ – предикция; ϕ – восприятие; M – действие; Q – [предполагаемый] мир; F – свободная энергия; e – ошибка предикции. Читатели-психоаналитики распознают некоторые из этих величин от Фрейда (1950 [1895]). Мы используем те же символы в знак признания глубоких инсайтов, содержащихся в его «Проекте», хотя стало необходимым – наряду с некоторыми другими инсайтами, записанными в сносках выше, – использовать их несколько иначе, чем то, что имел в виду Фрейд.

Полезно понимать, что каждый нейрон ошибки предикции (или нейронная популяция) снабжен специфическим и изменяющимся постсинаптическим усилением и, следовательно, имплицитным представлением точности. Точность – это не одно значение; каждое ощущение и действие и каждая иерархическая абстракция, включая каждую предикцию и последующий сигнал ошибки, должны быть снабжены точностью, которая должна быть оптимизирована.

Из приведенных выше уравнений также ясно, что точность (сознание) *контролирует* влияние ошибок предикции на действие (мотивацию) и восприятие (внимание). Концептуально точность является ключевым фактором минимизации свободной энергии и включения или активации ошибок предикции. Другими словами, точность определяет, какие ошибки предикции выбираются и в конечном счете как мы представляем мир и наши действия в нем.

В этом смысле точность играет роль демона Максвелла⁴³ – выбор прохождения молекул (то есть сенсорных сигналов), чтобы опровергнуть Второй закон термодинамики. В этой аналогии сознание есть не что иное, как деятельность демона Максвелла (то есть оптимизация точности по отношению к свободной энергии). То есть в этой аналогии сознание соответствует не прохождению молекул, которые активируются демоном (то есть перцептивными последовательностями передачи сообщений в корковых иерархиях), а активности самого демона.

Это различие лежит в основе предубеждения (Коха и других) о том, что нейромодуляция просто «включает» сознательное содержание. Концептуальный прорыв, о котором здесь говорится, связан с пониманием того, что остаточная ошибка в каждом цикле действия/восприятия (зарегистрированном в PAG, см. часть «В начале был аффект») *ощущается* как неопределенность – то есть каждая из различных категорий (или вкусов) ошибки обладает собственным аффективным «содержанием». Здесь неудовольствие (в рамках рассматриваемой модальности) означает *возрастающую неопределенность* в модальности, а удовольствие означает, что *все идет так, как ожидалось*. Эта ощущаемая неопределенность причинно обуславливает (ERTAS) корректировку последующих сенсомоторных приоритетов и ожиданий (т. е. ω). То есть она определяет избирательное возбуждение. Вот в чем суть дела.

Обратите внимание, что это предложение обращается к понятию *активации* ожиданий или представлений в том смысле, что отсутствие точности ошибки предикции может не вызвать никакой нейронной реакции. Другими словами, без точности ошибки предикции могут быть изолированы в точке их формирования в сенсорном эпителии (или на каком

⁴³ Демон Максвелла – это мысленный эксперимент, созданный Джеймсом Клерком Максвеллом, чтобы показать, как может быть нарушен Второй закон термодинамики: короче говоря, демон управляет маленькой дверью между двумя газовыми камерами. Когда молекулы газа приближаются, демон открывает и закрывает дверь, так что быстрые молекулы проходят в другую камеру, в то время как медленные молекулы остаются в первой, тем самым уменьшая энтропию.

бы уровне иерархии прогностической обработки они ни возникали). Физиологически такие состояния встречаются каждый день, например при стереотипных поведенческих автоматизмах и во время сна (*Hobson, 2009; Hobson and Friston, 2014*)⁴⁴.

Различие между *интероцептивной* и *экстероцептивной* точностью является центральным в данном аргументе. Если мозг является симпатическим органом предположения, усваивающим экстероцептивные (сенсорные/моторные) и интероцептивные (вегетативные) данные посредством предикции, то их (предикций) соответствующая точность о чем-то говорит (см. *Brentano, 1874*).

Предполагается, что интероцептивная точность является приоритетной, поскольку вероятностные убеждения, связанные с тем, что Панксепп называет гомеостатическими аффектами (например, голод, жажда, сонливость), нельзя не принимать во внимание. Убеждения организма на этом уровне иерархии диктуются фенотипом, а не опытом. Это означает, что все, что следует в иерархии, ведущей от центра головного мозга к сенсомоторной периферии, подчинено аффекту. Вот почему я описываю настройку ω как таковую как «аффект». Само сознание аффективно. Все остальное (от мотивации и внимания, ведущих к действию и восприятию, а тем самым и к обучению) – все это есть функция от аффекта. Аффект *обязывает* организм вступать во взаимодействие с внешним миром и тем самым определяет все его активное, субъективно воплощенное взаимодействие с ним.

Ничто из этого не может продолжаться в темноте.

Интроспективная точность изначально связана с самостью и интенциональностью (и, следовательно, с выживанием). Ее компульсивное качество постепенно размывается по мере прохождения иерархии центробежной обработки через инстинктивные и сенсорные аффективные механизмы и связанные с ними недеklarативные поведенческие стереотипы, через декларативные системы долгосрочной памяти к постоянно меняющейся периферии кратковременной памяти (см. часть «Сознание возникает вместо следа памяти» ниже).

Аффективная ценность, скрытая в ω , должна быть неотъемлемым свойством любой самоорганизующейся системы, которая активно и контекстуально сопротивляется Второму закону термодинамики. Оптимизация точности определяет степень, в которой это значение будет ощущаться (т. е. выражаться через избирательное включение обновления убеждений) для целей выбора. Проясню: легко представить себе организм (или машину), в котором значения точности установлены таким образом, что реакции системы на ошибку предикции автоматизированы. Действительно, большие участки нервной системы человека (не говоря уже об остальном теле) организованы именно таким образом.

⁴⁴ Можно легко утверждать, что тот же самый механизм – то есть установка точных значений, чтобы ошибки предикции не вызывали никакого ответа, – лежит в основе вытеснения (см. *Solms, 2018a*). Это то, к чему сводится понятие Фрейда о вытеснении как «неудаче перевода» в нынешних рамках (Письмо Флиссу, 6 декабря 1896 года; Фрейд, 1950 [1892–99]).

Следует отметить, что качественные колебания аффекта (то есть ω) возникают непрерывно в результате периодических сравнений между предсказанными сенсорными состояниями, основанными на генеративной модели внутреннего тела (Q_n) и мира ($\psi(Q)$), и образцами реальных сенсорных состояний (ϕ). Эта повторяющаяся оценка сенсорных состояний порождает изменения субъективного качества только при *изменении* амплитуды ошибок предикции, сигнализируя об изменении неопределенности относительно состояния дел и, в частности, ожидаемых последствий действия (M). Только по этой причине следует сказать – как любезно попросил меня уточнить один из моих рецензентов, – что нирвана, к которой стремится описанная здесь идеальная самоорганизующаяся система, никогда не может быть достигнута в реальной биологической системе по той простой причине, что изменения (как внешние, так и внутренние) всегда происходят⁴⁵.

Ниже мы кратко рассмотрим связь этой способности с нейронной пластичностью. Трудно представить себе сложную самоорганизующуюся систему, гибко адаптирующуюся к изменяющимся и новым условиям в отсутствие некоторой такой способности. Это, на мой взгляд, и есть то, как и почему возникает сознание.

«Сознание возникает на месте следа памяти»

Эта часть будет непропорционально короткой (см. *Solms*, 2018a, d, в печати, для более полного описания).

Выше мы видели, что сознательные селф-состояния в своей основе являются аффективными состояниями. Сознание в его самой элементарной форме – это своего рода сигнальный механизм, который направляет поведение самоорганизующихся систем, когда они переживают ситуации, выходящие за пределы их предпочтительных состояний, поскольку они не оснащены автоматизированными (или автоматизируемыми) предикциями, чтобы иметь с ними дело.

В части «В начале был аффект» я объяснил, что предикции, которые возвращают нас, сложные организмы, к нашим предпочтительным состояниям, обеспечиваются в первую очередь инстинктивным поведением, которое является врожденным инструментом выживания. Эти инструменты хорошо служат нам и волей-неволей используются, однако они не могут с должным вниманием отнестись к сложности экологических ниш, в которых мы на самом деле находимся. По этой причине врожденные предикции должны дополняться обучением через опыт.

Вот почему мы *чувствуем* инстинктивные эмоции: мы чувствуем их, потому что они не предсказывают и не могут предсказать все отклонения. Короче говоря, то, что мы чувствуем, – это остаточная ошибка предикции и связанная с ней неопределенность, когда мы находимся в непредвиденных ситуациях. Это (чувство в рамках определенной модальности)

⁴⁵ Как поэтически говорит нам песня Говорящих Голов, «небеса – это место, где никогда ничего не происходит».

направляет выбор, который с течением времени порождает новые, приобретенные предикции, как описано в части «В начале был аффект».

Но идеалом такого эмоционального обучения является автоматизация приобретенных предикций (некоторые из них, такие как обуславливание страха, автоматизируются с самого начала; другие, такие как привязанность, закрепляются в течение более длительных периодов времени). Естественно, нам нужно создавать новые предикции, которые, по крайней мере, так же надежны, как и врожденные, и в той мере, в какой мы достигаем этого (то есть в той мере, в какой ошибки предикции уменьшаются), в той степени приобретенные эмоциональные предикции автоматизируются путем консолидации, вплоть до уровня процедурных систем памяти (которые «трудно выучить и трудно забыть», см. *Squire, 2004*). Таким образом, приобретенные предикции становятся похожими на инстинктивные не только по своим функциональным свойствам⁴⁶, но и по своей подкорковой анатомической локализации.

Самым важным функциональным свойством недеklarативных воспоминаний является сам факт их недеklarативности. Это сводится к тому, что подкорковые следы памяти не могут быть восстановлены в виде *образов* по той простой причине, что они не состоят в корковых картах сенсорных поверхностных органов⁴⁷. Они влекут за собой более простые причинно-следственные связи такого рода, которые были описаны выше как «ассоциативное изучение связи между действиями и их последствиями».

Корковые (декларативные) системы памяти, напротив, всегда готовы на основе ошибок предикции оживить психические образы, которые они представляют. Другими словами, декларативные системы легко возвращают следы ДП в состояние КВП *сознательной* рабочей памяти – для того чтобы обновить их⁴⁸. Это неизбежно влечет за собой активацию (то есть отбор) заметных корковых репрезентаций – их значимость определяется (и «приправляется») соответствующими ошибками предикции и дисперсией. Этот процесс (который Фристон называет «удивлением») не следует путать с сенсорным аффектом удивления. Чувственный аффект, о котором идет речь, может быть *любым из* гомеостатических, эмоциональных или сенсорных аффектов.

Важно отметить, что ощущаемые аффекты обычно включают в себя как выбранный сигнал ошибки, так и последующую корректировку корковых (а в более длительных временных рамках – подкорковых) точностей. Но поскольку последний (когнитивный) компонент прогностической работы связывает первую (аффективную) свободную энергию,

⁴⁷ См. Фрейд (1923): понятие «телесного эго» выводится из корковых проекций сенсорно-моторной периферии.

⁴⁸ Это свойство декларативной долговременной памяти в точности совпадает с тем, что Фрейд называл системой Pcs предсознательное, хотя, на мой взгляд, Pcs состоит как из словесных презентаций, так и презентаций вещей (как семантических, так и эпизодических следов). Конечно, в Ucs бессознательном (в недеklarативной памяти) нет презентаций вещей, есть только программы стереотипных действий.

то и рассматриваемые сознательные состояния будут напоминать сознательное мышление, а не чувства⁴⁹. Даже сознательное мышление требует присутствия субъекта опыта, но процесс становится бессознательным, как только это возможно. Это точно совпадает с тем фактом, что чувство сохраняется (требуется) только до тех пор, пока не решена когнитивная задача. Сознательная когнитивная способность – это крайне ограниченный ресурс (см. Закон Миллера, выше), который должен использоваться экономно.

В этих немногих словах мы объяснили сознательную часть познания – ту часть, которая остается «после объяснения выполнения всех [других] функций» (*Chalmers*).

Из вышесказанного, как мы надеемся, ясно, что основная задача когнитивного (коркового) сознания состоит в том, чтобы *задерживать* моторные реакции на аффективные «требования, предъявляемые разуму для работы»⁵⁰. Эта задержка позволяет думать. Таким образом становится ясно, что важнейшей функцией коры головного мозга является стабилизация недеklarативных исполнительных процессов, которая позволяет поднять их на более высокий «катектический уровень» (то есть в состояние связанности), что является самой сутью того, что мы называем (по уважительной причине) *рабочей* памятью.

Вышеописанный разворот процесса консолидации (реконсолидация; *Nader et al.*, 2000) делает следы ДП лабильными благодаря буквальному растворению белков, которые первоначально «связывали» их в цепочку (*Hebb*, 1949). Это повторяющееся ощущение и переосмысление своего пути через декларируемые задачи является по представленному здесь предложению функцией когнитивных квалий, которые так доминировали в современных исследованиях сознания. Короче говоря, сознательная реконсолидация – это предиктивная работа в действии. Вспоминается неясное изречение Фрейда (1920): «Сознание возникает на месте следа памяти» (то есть лабильный след – это не след, это состояние того, что Фрейд называл «разрядкой» влечения; см. *Solms*, 2015).

Перцептивное/когнитивное сознание (активизируемое через внимание), как и сам аффект, является продуктом неопределенности. Недекларативные (подкорковые) следы памяти гораздо менее неопределенны, более точны, но и менее сложны, чем декларативные (корковые). Таким образом, относительная степень точности, обычно присущая корковым и подкорковым ошибкам предикции, совпадает с относительной

⁴⁹ Это примерно соответствует фрейдовскому различию между свободными и связанными катексисами. Однако не следует упускать из виду тот факт, что целью мышления является автоматизация. Короче говоря, связанные катексисы просто терпимы Эго (см. компромиссный «принцип постоянства» Фрейда). Идеальным состоянием Эго остается нирвана (любопытное состояние, в котором нет остаточной свободной энергии и точность становится бесконечной).

⁵⁰ Это в точности совпадает с понятием Фрейда о «вторичном процессе». Фрейд описал различие между свободной и связанной нервной энергией как свой «глубочайший инсайт» и добавил: «Я не вижу, как мы можем избежать делать это» (Фрейд, 1915а, с.188).

пластичностью (сопротивляемостью изменениям) связанных с ними убеждений.

Нужно только добавить, что экстероцептивные сенсомоторные модальности «приправлены» сознанием точно так же, как и интероцептивные, и по той же причине. Это облегчает фрагментирование релевантных данных (и тем самым снижает вычислительную сложность), в то время как селф-система серфит неопределенность в контекстуально изменяющихся условиях (роль точного взвешивания в этих условиях по отношению к различным перцептивным модальностям и, что наиболее интересно, по отношению к языку и внутренней речи подробно обсуждается *Нохву*, 2013, и *Clark*, 2016).

Эти лаконичные формулировки создают основу для новой, интегрированной теории аффективного и когнитивного сознания (и бессознательного).

Заключение

В этой статье я обратил внимание на два препятствия на пути решения «сложной проблемы» сознания – философское и научное – и предложил, как эти препятствия можно устранить. Первая – это популярная идея о том, что мозг «производит» сознание, то есть что физиологические процессы буквально *превращаются* в опыт посредством какой-то любопытной метафизической трансформации. Вторым препятствием является общепринятое представление о том, что сознание является функцией коры головного мозга, то есть что визуальное осознание (или любая другая форма сознательного познания) служит образцовым примером сознания. Принятие двухаспектной монистической позиции по философской проблеме «разум – тело» позволяет нам найти причинный механизм сознания не в проявленном мозге, а в его *функциональной организации*, которая в конечном счете лежит в основе как физиологических, так и психологических проявлений опыта. Чтобы выйти за пределы образного языка дуализма, эта объединяющая (монистическая) организация должна быть описана в *абстрактных* терминах (то есть не в физиологических или психологических терминах, а скорее в математических). «На этом фоне» я (как Дамасио и др.) предполагаю, что давно искомый механизм сознания должен быть найден в *расширенной форме гомеостаза*, которая описывает способ функционирования как глубоких ядер ствола мозга, которые обеспечивают НКС аффективного возбуждения, так и самого переживания чувства (которое, по-видимому, является основополагающей формой сознания). Этот тип гомеостаза (формализованный здесь как минимизация свободной энергии) влечет за собой генерацию аффектов (формализованных как ошибки гомеостатического предсказания), которые должны быть контекстуально приоритетными по отношению друг к другу и несистемным событиям (формализованным как взвешивание точности), что приводит к модуляции восприятия и действия (формализованной как коррекция ошибок) на основе ощущаемой неопределенности. Этот модуляторный процесс возбуждения, в свою очередь,

приводит к обучению на основе опыта посредством реконсолидации, которая дает огромное адаптивное преимущество перед более простыми типами гомеостаза – такими, как те, которые обнаруживаются в вегетативных (непроизвольных) нервных системах и холодильниках, – преимуществом является способность к целенаправленному поведению для сохранения жизни в непредвиденных ситуациях.

Авторство

Автор подтверждает, что он является единственным автором данной работы и одобряет ее публикацию.

Признательность

Я хотел бы выразить благодарность Карлу Фристону за ценные исправления раздела «Если быть точным».

СПИСОК ЛИТЕРАТУРЫ

1. Bargh J., and Chartrand T. (1999). The unbearable automaticity of being. *Am. Psychol.* 54, 462–479. doi: 10.1037/0003-066X.54.7.462
2. Berridge K. (1996). Food reward: brain substrates of wanting and liking. *Neurosci. Biobehav. Rev.* 20, 1–25. doi: 10.1016/0149-7634(95)00033-B
3. Berridge K., Robinson T., and Aldridge J. W. (2009). Dissecting components of reward: ‘liking’, ‘wanting’, and learning. *Curr. Opin. Pharmacol.* 9, 65–73. doi: 10.1016/j.coph.2008.12.014
4. Blomstedt P., Hariz M., Lees A., Silberstein P., Limousin P., Yelnik J., et al. (2007). Acute severe depression induced by intraoperative stimulation of the substantia nigra: a case report. *Parkinsonism Relat. Disord.* 14, 253–256. doi: 10.1016/j.parkreldis.2007.04.005
5. Brentano F. (1874). *Psychology From an Empirical Standpoint*. London: Routledge.
6. Carhart-Harris R., and Friston K. (2010). The default mode, ego functions and free energy: a neurobiological account of Freudian ideas. *Brain* 133, 1265–1283. doi: 10.1093/brain/awq010
7. Chalmers D. (1995). Facing up to the problem of consciousness. *J. Consciousness Stud.* 2, 200–219.
8. Chalmers D.J. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press.
9. Clark A. (2016). *Surfing Uncertainty*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780190217013.001.0001
10. Craig A.D. (2009). How do you feel – now? The anterior insula and human awareness. *Nat. Rev. Neurosci.* 10, 59–70. doi: 10.1038/nrn2555
11. Crick F. (1994). *The Astonishing Hypothesis*. New York, NY: Scribner.
12. Damasio A. (1994). *Descartes’ Error*. New York, NY: Putnam.
13. Damasio A. (1999). Commentary to Panksepp, emotions as viewed by psychoanalysis and neuroscience. *Neuropsychoanalysis* 1, 38–39. doi: 10.1080/15294145.1999.10773242
14. Damasio A. (2010). *Self Comes to Mind*. New York, NY: Pantheon.
15. Damasio A. (2018). *The Strange Order of Things*. New York, NY: Pantheon.

16. *Damasio A., and Carvalho G.* (2013). The nature of feelings: evolutionary and neurobiological origins. *Nat. Rev. Neurosci.* 14, 143–152. doi: 10.1038/nrn3403
17. *Damasio A., Damasio H., and Tranel D.* (2012). Persistence of feeling and sentience after bilateral damage of the insula. *Cereb. Cortex* 23, 833–846. doi: 10.1093/cercor/bhs077
18. *Du Bois-Reymond E.* (1918). Letter to Hallmann, 1842. *Jugendbriefe von Emil Du Bois-Reymond an Eduard Hallmann.* Berlin: Dietrich Reimer. p. 108.
19. *Ellis G., and Solms M.* (2018). *Beyond Evolutionary Psychology: How and Why Neuropsychological Modules Arise.* Cambridge: Cambridge University Press.
20. *Freud S.* (1893). Charcot, Standard Edn. Vol. 3. London: Hogarth Press. p. 11–23.
21. *Freud S.* (1894). The Neuro-Psychoses of Defence, Standard Edn. Vol. 3. London: Hogarth Press. p. 45–61.
22. *Freud S.* (1900). The Interpretation of Dreams, Standard Edn. London: Hogarth Press, 4 and 5.
23. *Freud S.* (1901). The Psychopathology of Everyday Life. Standard Edn. London: Hogarth Press. p. 6.
24. *Freud S.* (1915a). The Unconscious, Standard Edn. Vol. 14. London: Hogarth Press. p. 166–204.
25. *Freud S.* (1915b). Instincts and Their Vicissitudes, Standard Edn. Vol. 14. London: Hogarth Press. p. 117–140.
26. *Freud S.* (1916–17) Introductory Lectures in Psychoanalysis, Standard Edn. London: Hogarth Press. p. 15–16.
27. *Freud S.* (1917). Metapsychological Supplement to the Theory of Dreams. Standard Edn. Vol. 14. London: Hogarth Press. p. 222–235.
28. *Freud S.* (1920). Beyond the Pleasure Principle, Standard Edn. Vol. 18. London: Hogarth Press. p. 7–64.
29. *Freud S.* (1923). The Ego and the id. Standard Edn. Vol. 19. London: Hogarth Press. p. 12–59.
30. *Freud S.* (1925a). An Autobiographical Study, Standard Edn. London: Hogarth Press. p. 20.
31. *Freud S.* (1925b). A Note Upon “the mystic writing-pad,” Standard Edn. Vol. 16. London: Hogarth Press. p. 227–232.
32. *Freud S.* (1940). An Outline of Psychoanalysis, Standard Edn. Vol. 23. London: Hogarth Press. p. 144–207.
33. *Freud S.* (1950 [1892-99]). Extracts From the Fliess Papers, Standard Edn. Vol. 1. London: Hogarth Press. p. 174–280.
34. *Freud S.* (1950 [1895]) Project for a Scientific Psychology. Standard Edn. Vol. 1. London: Hogarth Press. p. 281–397.
35. *Friston K.* (2013). Life as we know it. *J. R. Soc. Interface* 10:20130475. doi: 10.1098/rsif.2013.0475
36. *Golaszewski S.* (2016). Coma-causing brainstem lesions. *Neurology* 87:2433. doi: 10.1212/WNL.0000000000003417
37. *Groddeck G.* (1977). *The Meaning of Illness: Selected Psychoanalytic Writings Including His Correspondence With Sigmund Freud.* London: Hogarth.

38. *Harlow J.M.* (1868). Recovery from the passage of an iron bar through the head. *Publicat. Massachusetts Med. Soc.* 2, 327–347.
39. *Hebb D.* (1949). *The Organization of Behavior*. New York, NY: John Wiley.
40. *Hobson J.A.* (2009). REM sleep and dreaming: towards a theory of protoconsciousness. *Nat Rev Neurosci.* 10, 803–813. doi: 10.1038/nrn2716
41. *Hobson J.A., and Friston K.* (2014). Consciousness, dreams, and inference: the Cartesian theatre revisited. *J. Consciousness Stud.* 21, 6–32.
42. *Hohwy J.* (2013). *The Predictive Mind*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199682737.001.0001
43. *Huston J., and Borbely A.* (1974). The thalamic rat: general behaviour, operant learning with rewarding hypothalamic stimulation, and effects of amphetamine. *Physiol. Behav.* 12, 433–448. doi: 10.1016/0031-9384(74)90121-8
44. *Kandel E., Schwartz J., Jessell T., Siegelbaum S., and Hudspeth A.* (2012). *Principles of Neural Science*, 5th Edn. New York, NY: Elsevier.
45. *Kihlstrom J.* (1996). “Perception without awareness of what is perceived, learning without awareness of what is learned,” in *The Science of Consciousness: Psychological, Neuropsychological and Clinical Reviews*, ed, M. Velmans (London, Routledge), p. 23–46.
46. *Koch C.* (2004). *The Quest for Consciousness*. New York, NY: WH Freeman.
47. *LeDoux J., and Brown R.* (2017). A higher-order theory of emotional consciousness. *Proc. Natl. Acad. Sci. U.S.A.* 114, E2016–E2025. doi: 10.1073/pnas.1619316114
48. *Levine J.* (1983). Materialism and qualia: the explanatory gap. *Pac. Philos. Q.* 64, 354–361. doi: 10.1111/j.1468-0114.1983.tb00207.x
49. *Merker B.* (2007). Consciousness without a cerebral cortex: a challenge for neuroscience and medicine. *Behav. Brain Sci.* 30, 63–134. doi: 10.1017/S0140525X07000891
50. *Meyer J., and Quenzer L.* (2005). *Psychopharmacology: Drugs, The Brain, and Behavior*. Sunderland, MA: Sinauer Associates.
51. *Meynert T.* (1884). *Psychiatrie. Klinik der Erkrankungen des Vorderhirns, begründet auf dessen Bau, Leistungen und Ernährung*, Vienna: Wilhelm Braumüller.
52. *Mongillo G., Barak O., and Tsodyks M.* (2008). Synaptic theory of working memory. *Science* 319, 1543–1546. doi: 10.1126/science.1150769
53. *Moruzzi G., and Magoun H.* (1949). Brain stem reticular formation and activation of the EEG. *Electroencephalog. Clin. Neurol.* 1, 455–473. doi: 10.1016/0013-4694(49)90219-9
54. *Munk H.* (1878). Weitere Mittheilungen zur Physiologie der Grosshirnrinde. *Arch für Physiol.* 2, 162–177
55. *Munk H.* (1881). *Ueber die Funktionen der Grosshirnrinde*. Berlin: August Hirschwald.
56. *Nader K., Schafe G.E. and Le Doux J.* (2000). Fear memories require protein synthesis in the amygdala for reconsolidation after retrieval. *Nature* 406, 722–726. doi: 10.1038/35021052
57. *Nagel T.* (1974). What is it like to be a bat? *Philos. Rev.* 83, 435–450. doi: 10.2307/2183914

58. *Panksepp J.* (1998). *Affective Neuroscience*. Oxford: Oxford University Press.
59. *Panksepp J., Lane R.D., Solms M. and Smith R.* (2016). Reconciling cognitive and affective neuroscience perspectives on the brain basis of emotional experience. *Neurosci. Biobehav. Rev.* 76, 187–215. doi: 10.1016/j.neubiorev.2016.09.010
60. *Parvizi J., and Damasio A.* (2003). Neuroanatomical correlates of brainstem coma. *Brain* 126, 1524–1536. doi: 10.1093/brain/awg166
61. *Penfield W. and Jasper H.* (1954). *Epilepsy and the Functional Anatomy of the Human Brain*. Oxford: Little & Brown.
62. *Pfaff D.* (2006). *Brain Arousal and Information Theory*. Cambridge, MA: Harvard University Press. doi: 10.4159/9780674042100
63. *Pulsifer M., Brandt J., Salorio C., Vining E., Carson B. and Freeman J.* (2004). The cognitive outcome of hemispherectomy in 71 children. *Epilepsia* 45, 243–254. doi: 10.1111/j.0013-9580.2004.15303.x
64. *Schultz W.* (2016). Dopamine reward prediction error coding. *Dialogues Clin. Neurosci.* 18, 23–32.
65. *Searle J.* (1997). *The Mystery of Consciousness*. New York, NY: New York Review of Books.
66. *Searle J.* (2017). “Foreword,” in *Biophysics of Consciousness: A Foundational Approach*. eds R. Poznanski, J. Tuszynski and T. Feinberg (New York, NY: World Scientific) 129–148.
67. *Sharma J., Angelucci A. and Sur M.* (2000). Induction of visual orientation modules in auditory cortex. *Nature* 404, 841–847. doi: 10.1038/35009043
68. Shewmon, D., Holmse, D., and Byrne, P. (1999). Consciousness in congenitally decorticate children: developmental vegetative state as a self-fulfilling prophecy. *Dev. Med. Child Neurol.* 41, 364–374. doi: 10.1017/S0012162299000821
69. *Solms M.* (1997). What is consciousness? [and response to commentaries]. *J. Amer. Psychoanal. Assn.* 45, 681–778. doi: 10.1177/00030651970450031201
70. *Solms M.* (2013). The conscious id. [and response to commentaries]. *Neuropsychoanalysis* 15, 5–85 doi: 10.1080/15294145.2013.10773711
71. *Solms M.* (2014). A neuropsychoanalytical approach to the hard problem of consciousness. *J. Integr. Neurosci.* 13, 173–185. doi: 10.1142/S0219635214400032
72. *Solms M.* (2015). Reconsolidation: turning consciousness into memory. *Behav. Brain Sci.* 38, 40–41. doi: 10.1017/S0140525X14000296
73. *Solms M.* (2018a). What is ‘the unconscious’ and where is it located in the brain? *Ann. NY Acad. Sci.* 1406, 90–97.
74. *Solms M.* (2018b). Review of damasio, the strange order of things. *J. Am. Psychoanal. Ass.* 66, 579–586. doi: 10.1177/0003065118780182
75. *Solms M.* (2018c). Extracts from the revised standard edition of Freud’s complete psychological works. *Int. J. Psychoanal.* 99, 11–57. doi: 10.1080/00207578.2017.1408306
76. *Solms M.* (2018d). The neurobiological underpinnings of psychoanalytic theory and therapy. *Front. Behav. Neurosci.* 12:294. doi: 10.3389/fnbeh.2018.00294

77. *Solms M.* (in press). *Consciousness Itself: Feeling and Uncertainty*. London: Profile Books.
78. *Solms M. and Friston K.* (2018). How and why consciousness arises: some considerations from physics and physiology. *J. Conscious. Stud.* 25, 202–238.
79. *Solms M., Kaplan-Solms K. and Brown J.W.* (1996). “Wilbrand’s case of ‘mind blindness’,” in *Classic Cases in Neuropsychology*, eds C. Code, Y. Joannette, A. Lecours, and C.-W. Wallesch (London: Psychology Press), 89–110.
80. *Solms M. and Panksepp J.* (2012). The ‘id’ knows more than the ‘ego’ admits. *Brain Sci.* 2, 147–175. doi: 10.3390/brainsci2020147
81. *Solms M. and Saling M.* (1990). *A Moment of Transition: Two Neuroscientific Articles by Sigmund Freud*. London: Karnac.
82. *Squire L.* (2004). Memory systems of the brain: a brief history and current perspective. *Neurobiol. Learn. Mem.* 82, 171–177. doi: 10.1016/j.nlm.2004.06.005
83. *Strachey J.* (1962). *The Emergence of Freud’s Fundamental Hypotheses*. Standard Edn (London: Early Psycho-Analytic Publications), 3, 62–68.
84. *Teuber H.-L.* (1955). Physiological psychology. *Ann. Rev. Psychol.* 6, 267–296. doi: 10.1146/annurev.ps.06.020155.001411
85. *Tononi G.* (2012). *Phi: A Voyage from the Brain to the Soul*. New York, NY: Pantheon.
86. *Wakefield J.* (2018). *Freud and Philosophy of Mind, Volume 1: Reconstructing the Argument for Unconscious Mental States*. London: Palgrave Macmillan doi: 10.1007/978-3-319-96343-3

The Hard Problem of Consciousness and the Free Energy Principle

M. Solms

*(Transl. from English: Elena Kalinina;
Scientific. ed.: Ekaterina Shapovalova)*

Mark Solms, psychoanalyst and neuropsychologist, the Chair of Neuropsychology at the University of Cape Town and Groote Schuur Hospital (Departments of Psychology and Neurology) and is the President of the South African Psychoanalytical Association. Co-chair of the Nueropschoanalytic Association (NPSA)

This article applies the free energy principle to the hard problem of consciousness. After clarifying some philosophical issues concerning functionalism, it identifies the elemental form of consciousness as affect and locates its physiological mechanism (an extended form of homeostasis) in the upper brainstem. This mechanism is then formalized in terms of free energy minimization (in unpredicted contexts) where decreases and increases in expected uncertainty are felt as pleasure and displeasure, respectively. Emphasis is placed on the reasons why such existential imperatives feel like something to and for an organism

Keywords: hard problem, consciousness, free energy, predictive processing, affect, Freud.