Seth AK, Bayne T. Theories of consciousness. Nat Rev Neurosci. 2022 Jul;23(7):439-452. doi: 10.1038/s41583-022-00587-4. Epub 2022 May 3.

Аннотация

Теории сознания В последние годы наблюдается расцвет теорий о биологических и физических основах сознания. Хорошие теории направляют эмпирические исследования, позволяя нам интерпретировать данные, разрабатывать новые экспериментальные методы и расширять наши возможности по манипулированию интересующим нас феноменом. Действительно, только сформулированные в терминах теории эмпирические открытия могут в конечном счете обеспечить удовлетворительное понимание феномена. Однако в случае с сознанием неясно, как современные теории соотносятся друг с другом и можно ли их эмпирически различить. Чтобы прояснить этот сложный ландшафт, мы рассмотрим четыре известных теоретических подхода к сознанию: теории высшего порядка, теории глобального рабочего пространства, теории повторного входа и прогнозирующей обработки данных, а также интегрированную теорию информации. Мы описываем ключевые характеристики каждого подхода, определяя, какие аспекты сознания они предлагают объяснить, каковы их нейробиологические обязательства и какие эмпирические данные приводятся в их поддержку. Мы рассмотрим, как некоторые известные эмпирические дебаты могут различать эти теории, и наметим три способа, которыми необходимо разрабатывать теории, чтобы обеспечить зрелый режим проверки теорий в нейробиологии сознания. Есть веские основания полагать, что последовательная разработка, тестирование и сравнение теорий сознания приведут к более глубокому пониманию этой самой главной из тайн.

Введение

В первые десятилетия своего возрождения научное изучение сознания было сосредоточено на поиске "нейронных коррелятов сознания’ (NCC). Формально NCC сознательного состояния представляют собой минимальный набор нейронных событий, в совокупности достаточных для этого состояния; на практике поиск NCC включает в себя поиск состояний и процессов мозга, которые наиболее тесно связаны с сознанием1-3. Сосредоточение внимания на поиске NCC было полезным, поскольку понятие NCC относительно "теоретически нейтрально", и, таким образом, структура NCC обеспечивала общий язык и методологию для исследователей с различными теоретическими и даже метафизическими взглядами. Однако ограничения системы NCC становятся все более очевидными, что проявляется, например, в трудностях, связанных с отличением "истинных" NCC от нейронных предпосылок и последствий осознанности 4-7. В ответ на эти ограничения все большее внимание уделяется разработке теорий сознания. Имея в своем распоряжении теорию сознания (ToC), мы смогли бы выйти за рамки методологии, основанной на NCC, и перейти к моделям сознания, которые обеспечивают понимание причин. Действительно, наличие эмпирически подтвержденной теории должно быть главной целью науки о сознании8,9.

В то время как подход NCC ставит во главу угла поиск корреляций между мозговой активностью и сознанием, теоретический подход вместо этого фокусируется на выявлении объяснительных связей между нейронными механизмами и аспектами сознания10. Тем не менее, теоретики часто используют различные концепции того, что необходимо для обеспечения объяснительной связи между нейронной активностью и сознанием. Некоторые полагают, что удовлетворительный ТоС должен и может заполнить "пробел в объяснениях" (вставка 1) и что можно будет сделать взаимосвязь между нервной активностью и сознанием такой же прозрачной, как взаимосвязь между химической структурой воды и ее общим поведенческим профилем11. Другие сомневаются или остаются агностиками относительно того, будет ли когда-либо полностью устранен этот пробел в объяснениях, но, тем не менее, надеются на концепцию, которая могла бы объяснить определенные аспекты сознания и, тем самым, уменьшить или устранить ощущение таинственности, окружающее его биофизическую основу12,13. Третьи утверждают, что интуитивные объяснения пробелов вводят в заблуждение и наука о сознании не должна воспринимать их всерьез.14,15.

В настоящее время существует широкий спектр потенциальных ТОС (таблица 1). Примечательно, что вместо постепенного "исключения" ТОС по мере накопления эмпирических данных, их число, по-видимому, растет. Это распространение привело как к попыткам интегрировать существующие теории друг с другом16, так и к развитию "состязательного сотрудничества", при котором сторонники конкурирующих теорий заранее договариваются о том, поддержат ли результаты предлагаемого эксперимента предпочитаемую ими теорию17. Однако, как мы уже говорили, существуют серьезные проблемы как на пути интеграции теории, так и на пути состязательного сотрудничества.

В этом обзоре мы рассмотрим, как различные TOCs соотносятся друг с другом и с эмпирическими данными, и определим несколько многообещающих направлений, с помощью которых разработка теории и эмпирические исследования могут совместно поддерживать друг друга в поисках удовлетворительного научного описания сознательного опыта. Наше внимание ограничено теориями, которые либо сами по себе выражены в нейробиологических терминах, либо правдоподобно подразумевают утверждения, которые могут быть выражены в нейробиологических терминах. [Как мы увидим, некоторые "нейробиологические" теории сознания выражены на абстрактном языке функциональных отношений или теории информации и квалифицируются как "нейробиологические" только потому, что абстрактные характеристики, к которым они апеллируют, связаны с конкретными нейронными механизмами.] Мы также рассматриваем только нейробиологические теории, которые согласуются с известной физической теорией, и мы также оставляем в стороне теории, которые непосредственно связывают сознание с квантово-механическими процессами (примеры см. в ссылках 18,19).

Предварительные замечания Одна из главных причин, по которой TOCS "разговаривают" друг с другом, заключается в том, что они часто преследуют разные цели объяснения, поскольку фокусируются на разных аспектах сознания. Поэтому мы начнем с рассмотрения того, на что должен быть направлен всеобъемлющий ToC, отмечая, что даже этот вопрос является спорным, и теоретики часто расходятся во мнениях о том, какие явления должна объяснять теория сознания.

Суть проблемы сознания заключается в "опыте" или "субъективном осознании". Хотя однозначного определения этим терминам дать невозможно, исследуемый феномен можно прояснить с помощью некоторых интуитивных различий. Существует "нечто, на что это похоже", когда организм находится в сознании20, и то, каково это - находиться в одном состоянии сознания, отличается от того, каково это - находиться в другом состоянии сознания. Всеобъемлющаяй теория объяснит, почему одни организмы или системы обладают сознанием, в то время как другие - нет, и объяснит, почему состояния сознания отличаются друг от друга тем, как они проявляются.

Состояния сознания можно разделить на два класса: глобальные состояния и локальные состояния. Глобальные состояния касаются общего субъективного профиля организма и связаны с изменениями в возбуждении и поведенческой реактивности. Знакомые глобальные состояния включают бодрствование, сновидение, успокоение, состояние минимального сознания, (возможно) психоделическое состояние и так далее. Эти глобальные состояния иногда называют "уровнями" сознания, но мы предпочитаем термин "глобальные состояния", поскольку он оставляет открытой возможность того, что эти состояния не могут быть полностью упорядочены в терминах одного измерения, а вместо этого лучше всего концептуализируются как области в многомерном пространстве21.

Локальные состояния, которые часто называют "сознательными содержаниями" или состояниями, обладающими "квалиа", характеризуются тем, "на что это похоже" — находиться в них. Локальное состояние, связанное с головной болью, отличается от локального состояния, связанного с запахом кофе, поскольку ощущение головной боли отличается от ощущения запаха кофе. Локальные состояния могут быть описаны на разных уровнях детализации, от низкоуровневых перцептивных характеристик (например, цвета) до объектов и полных мультимодальных перцептивных сцен. Важное подмножество локальных состояний лежит в основе переживания самости, которое включает в себя переживания настроения, эмоций, воли, владения телом, явную автобиографическую память и тому подобное 13,22-24. Хотя нейробиологические теории, как правило, фокусируются на локальных состояниях с сенсорным и перцептивным содержанием, сознание также включает локальные состояния с когнитивным и пропозициональным содержанием, такие как мысли, возникающие при разгадывании кроссворда. Важно отметить, что локальные состояния, которые испытывает агент в определенный момент времени, не просто возникают как независимые элементы, но вместо этого связаны друг с другом как компоненты единой сознательной сцены, которая включает в себя каждое из локальных состояний агента25,26.

Второе различие заключается в различии между феноменальными свойствами сознания и его функциональными свойствами. Первый термин относится к эмпирическому характеру сознания, что подтверждается фразой "на что это похоже’. Функциональные аспекты сознания касаются роли (ов), которую(-ые) психические состояния играют в когнитивной экономике организма в силу того, что он сознателен. (Термин "функция" здесь охватывает как телеологические функции — функциональные роли, сформированные эволюцией, — так и диспозициональные функции — роль, которую процесс играет в функционировании более крупной системы, частью которой он является; см. ССЫЛКУ 27.) Например, осознание того, что вы видите кофейную чашку, может активировать целый ряд функций, таких как способность гибко относиться к чашке (возможно, пить из нее или швырять ее через всю комнату), сохранять эпизодические воспоминания о событии, обеспечивать устные отчеты о пережитом и так далее. Проводя это различие, мы не утверждаем, что феноменальные и функциональные свойства независимы (скорее всего, они не независимы), а просто указываем, что они обеспечивают четкие объяснительные цели для ТОС. Как мы увидим, некоторые TOCs фокусируются на феноменальных особенностях сознания, другие - на функциональных особенностях сознания, а третьи пытаются объяснить как функциональные, так и феноменальные особенности сознания.

Третье различие заключается в двух видах вопросов, касающихся локальных состояний ("содержимого"), на которые ToC может попытаться ответить. С одной стороны, можно спросить, почему агент находится в определенном локальном состоянии (а не в другом). С другой стороны, можно было бы спросить, почему конкретное локальное состояние имеет тот эмпирический характер, который оно имеет (а не какой-то другой эмпирический характер). Это различие может быть объяснено в связи с бинокулярным соперничеством, при котором каждому глазу предъявляются разные стимулы (например, дом для правого глаза и лицо для левого глаза), и зрительный опыт испытуемого меняется от стимула для левого глаза к стимулу для правоглаза 28. Выделите определенное время, в течение которого содержание сознания включает в себя дом, в то время как лицо сознательно не воспринимается. Здесь мы можем спросить, почему психическое состояние, соответствующее "дому", является осознанным (а психическое состояние, связанное с "лицом", - бессознательным), и мы также можем спросить, почему визуальные переживания дома имеют особый эмпирический характер, а не, скажем, эмпирический характер видения лица, слышания звонок или ощущение боли.

Примечательно, что некоторые материалы могут не поддаваться осознанию (например, низкоуровневая обработка в рамках ранних сенсорных или регуляторных систем), а другие могут быть только осознанными (например, глобально интегрированные перцептивные сцены). Таким образом, в дополнение к объяснению того, почему некоторые ментальные содержания осознаются в одних контекстах, но не в других, еще одна задача состоит в том, чтобы объяснить, почему некоторые содержания никогда не могут быть осознанными, а другие могут существовать только как сознательные.

Вместо того чтобы рассматривать весь спектр вопросов, которые мы только что обозначили, большинство TOCs стремятся объяснить только определенные аспекты сознания, возможно, в качестве шага на пути к тому, чтобы стать всеобъемлющими. Хотя некоторые ограничения сами по себе не являются возражением против оглавления, это означает, что задача межтеоретического сравнения менее проста, чем могла бы быть в противном случае. Если теории нацелены на разные аспекты сознания (скажем, одна теория сосредоточена на феноменальном характере сознания, а другая - на его функциональном профиле), то они могут оказаться не такими "противниками", какими кажутся на первый взгляд.

TOCs, которые мы рассматриваем здесь, сгруппированы в четыре категории: теории высшего порядка (HotS), теории глобального рабочего пространства (GWTs), интегрированная теория информации (IIT) и теории повторного входа и прогнозирующей обработки (PP). Хотя некоторые представления о сознании охватывают несколько категорий, а другие с большой вероятностью не подпадают ни под одну из этих категорий (таблица 1), это четырехстороннее различие между ТОС дает полезную возможность взглянуть на текущее состояние науки о сознании сквозь призму (вставка 2; другие способы понимания). группировка теорий, см., например, ссылку 29). Далее мы представим ключевые элементы каждой категории, опишем некоторые заметные различия внутри категорий и определим те аспекты сознания, которые наиболее тесно связаны с каждой категорией. Затем мы проиллюстрируем, как эти TOCs соотносятся друг с другом с точки зрения некоторых известных эмпирических дискуссий, и представим несколько предложений, которые, как мы полагаем, помогут наладить эффективный цикл между разработкой теории и экспериментальным исследованием.

Теории высшего порядка Основное утверждение, объединяющее все HotS, заключается в том, что психическое состояние является осознанным в силу того, что оно является объектом определенного вида метапредставительного состояния. Метапредставления - это не просто представления, которые располагаются выше или глубже в иерархии обработки, но, скорее, представления, целью которых являются другие представления (рис. 1). Например, представление с содержанием <У меня есть визуальное представление движущейся точки>

это метапредставление, поскольку его содержание касается скорее собственных представлений агента о мире, чем самого мира.

Важный аспект, в котором HOT отличаются друг от друга, касается того, как они описывают природу и роль метапредставлений, ответственных за сознание. Некоторые версии подхода определяют виды метапредставлений, которые имеют решающее значение для сознания, с помощью мыслей (или мыслеподобных состояний), которые имеют концептуальное содержание30,31,32. Другие разновидности HOT были выражены в вычислительных терминах. Согласно теории самоорганизующихся метарепрезентаций, сознание включает в себя мозговые сети более высокого порядка, которые учатся переописывать представления, закодированные в сетях более низкого порядка, таким образом, что это считается метарепрезентацией33,34.

Альтернативно, теория пространства состояний более высокого порядка предполагает, что субъективные сообщения (например, такие утверждения, как "Я знаю о X") являются метакогнитивными решениями (более высокого порядка) о порождающей модели перцептивного содержания35, в то время как мониторинг перцептивной реальности утверждает, что сознательное восприятие возникает, когда сеть более высокого порядка оценивает первичный контент.представление о порядке должно быть надежным отражением внешнего мира36,37.

Как должно быть ясно из вышесказанного, HotS фокусируется на объяснении того, почему некоторые содержания осознаются, а другие - нет. Однако эти теории не ограничиваются этим конкретным направлением — у них также есть ресурсы для решения проблем, связанных с эмпирическим характером локальных состояний. Один из ярких примеров касается (обсуждаемой) интуиции о том, что содержание перцептивного опыта часто превосходит информацию, доступную в сенсорных представлениях "первого порядка", что, как утверждается, происходит в контексте периферического зрения38,39. Предположение, основанное на HOT, состоит в том, что кажущаяся "раздутой" феноменология периферийного визуального опыта вызвана искажением состояний первого порядка более высокого порядка40. Подход HOT также может быть расширен, чтобы объяснить, почему некоторые содержания не могут быть осознанными (они не могут быть объектами соответствующих метапредставительных состояний) и почему некоторые содержания обязательно осознаваемы (они обязательно сопровождаются соответствующими метапредставительными состояниями). HOTS редко фокусируются на глобальных состояниях сознания, но для них было бы естественно апеллировать к целостности процессов репрезентации, чтобы объяснить различия между глобальными состояниями.

Особенно интригующий вопрос заключается в том, объясняют ли (и если да, то как) HOTS особый феноменальный характер различных видов переживаний. Почему феноменальный характер, связанный с наблюдением заката, так сильно отличается от феноменального характера, связанного с головной болью? Общая форма ответа более высокого порядка на этот вопрос заключается в том, что феноменальный характер состояния определяется свойствами, которые приписывает ему соответствующее метапредставительное состояние. Большинство примеров этого подхода сосредоточены на визуальном восприятии40, но также предпринимались попытки объяснить феноменальный характер эмоциональных состояний41 и метакогнитивных состояний, таких как "каково это" - чувствовать уверенность в принятии перцептивного решения42,43. В конечном счете, любая полностью редуктивная версия подхода высшего порядка должна объяснять, почему представление различных свойств порождает ту феноменологию, которую оно создает (или идентично ей), и как нейронная активность позволяет в первую очередь представить соответствующие свойства.

Представления о сознании более высокого порядка - это, прежде всего, представления о том, что делает психическое состояние осознанным, и, как таковой, этот подход не привязан к какому-либо конкретному взгляду на функцию (ы) сознания. Действительно, некоторые авторы принижают идею о том, что сознание имеет какую-либо особую функцию44. Другие версии подхода более высокого порядка отождествляют функциональную роль сознания с метакогнитивными процессами, связанными с оценками достоверности и мониторингом ошибок45. Однако, хотя взгляды более высокого порядка допускают, что сознательные психические состояния могут сопровождаться сознательными метакогнитивными суждениями — такими, как те, которые используются при явном мониторинге эффективности или в отчетах о субъективной уверенности, — большинство версий этого подхода не требуют, чтобы сознательное восприятие всегда сопровождалось соответствующим сознательным метакогнитивным состоянием. Вместо этого, чтобы метапредставления были осознанными, они сами должны быть объектами подходящего метапредставительного состояния.

Что касается нейронной основы сознания, то акцент на метапредставлении побудил теоретиков высшего порядка уделять особое внимание передним отделам коры головного мозга, особенно префронтальной коре [30], учитывая связь этих областей со сложными когнитивными функциями. Однако, хотя большинство исследователей предполагают, что вовлечение передних отделов связано с сознанием, существуют разногласия по поводу того, какие именно передние области (или процессы) необходимы для этого 46.

Глобальные теории рабочего пространства GWT основаны на архитектуре "классной доски" в области искусственного интеллекта, в которой классная доска является централизованным ресурсом, с помощью которого специализированные процессоры обмениваются информацией и получают ее. Первый GWT consciousness47 был разработан на когнитивном уровне. Он предположил, что осознанные психические состояния - это те, которые "глобально доступны" для широкого спектра когнитивных процессов, включая внимание, оценку, память и устный отчет.

Основное утверждение GWTs заключается в том, что именно широкая доступность информации для таких потребительских когнитивных систем является основой осознанного опыта (рис. 2).

Это основное утверждение с тех пор было развито в нейронную теорию, часто называемую “теорией глобального нейронного рабочего пространства”, согласно которой сенсорная информация получает доступ к сознанию, когда она "транслируется" в анатомически широко распространенном нейронном рабочем пространстве, которое реализуется в ассоциативных областях коры головного мозга более высокого порядка с определенной целью. (хотя и не исключительно) акцент на префронтальную кору головного мозга 48,49. Доступ к глобальному рабочему пространству обеспечивается за счет нелинейного сетевого "зажигания", в котором повторяющаяся обработка усиливает и поддерживает нейронные репрезентации50. Акцент на зажигании и трансляции — по сравнению с метапредставлением — это один из способов, которым GWT отличается от горячего подхода.

Как и HotS, GWT фокусируются на вопросе о том, что делает репрезентацию осознанной, и теоретики GWT редко пытались объяснить феноменальные различия между различными видами переживаний (хотя, хотя, см. ссылку 51).

Возвращаясь к нашему примеру с бинокулярным соперничеством, GWT view стремится объяснить, почему в определенный момент времени психическое состояние, соответствующее "дому", является осознанным (в то время как психическое состояние, соответствующее "лицу", является бессознательным), но оно не дает прямого объяснения эмпирического контраста между наблюдением дома на экране. с одной стороны, и видеть чье-то лицо с другой.

Относительное молчание GWTs по вопросу об эмпирическом характере согласуется с общей тенденцией таких теорий фокусироваться на функциональных, а не на феноменальных аспектах сознания. Фактически, GWT часто прямо предлагаются как объяснения "сознательного доступа"49; то есть как объяснения того, почему определенные представления доступны для гибкого использования широким спектром систем-потребителей (в то время как другие - нет). Основным функциональным свойством, на которое обращают внимание GWTs, является способность сознательных состояний управлять поведением и познанием гибким, зависящим от контекста образом. GWT также дают четкое представление о том, как сознание связано с другими когнитивными процессами, такими как внимание и рабочая память. Согласно GWTs, внимание выбирает и усиливает определенные сигналы, позволяя им проникать в рабочее пространство (и, таким образом, быть осознанными); в то время как сознание и рабочая память тесно связаны, поскольку посещаемые элементы рабочей памяти осознанны и используют глобальное рабочее пространство для трансляции49.

GWT учитывают изменения в глобальных состояниях сознания с точки зрения изменений функциональной целостности рабочего пространства. На нейронном уровне глобальная потеря сознания отражается в нарушении функциональных или динамических связей в лобно-теменных областях, которые рассматриваются как "узловые" узлы в глобальном рабочем пространстве52, и в том, что функциональные связи становятся все более ограниченными до шаблонов, непосредственно отражающих лежащую в их основе структурную связь53-55.

Один из важных вопросов, поднятых GWTs, касается того, что именно требуется для того, чтобы рабочее пространство можно было квалифицировать как "глобальное"25,56. Имеет ли значение количество (и тип) систем-потребителей, на которые может транслироваться рабочая область, или вид трансляции, осуществляемой внутри рабочей области? Или оба этих фактора имеют отношение к тому, что считается "глобальной рабочей областью’? На эти вопросы необходимо ответить, если мы хотим знать, какие прогнозы дают GWT в отношении сознания, например, младенцев, людей с повреждениями головного мозга, пациентов с раздвоением мозга, животных, не являющихся людьми, и систем искусственного интеллекта.

Интегрированная теория информации IIT начинается совсем с другого направления, чем HotS или GWTs, продвигая математический подход к описанию феноменологии. Теория начинается с предложения аксиом о феноменологическом характере сознательного опыта (то есть о свойствах, которые считаются самоочевидно истинными и применимыми ко всем возможным формам сознания), и из этих аксиом вытекают утверждения о свойствах, которым должен удовлетворять любой физический субстрат сознания. Затем IIT предполагает, что физические системы, которые реализуют эти свойства, обязательно реализуют и сознание (57-60, см. рис. 3).

В частности, ИИТ предполагает, что сознание следует понимать в терминах "причинно–следственной связи", связанной с непреодолимыми максимумами интегрированной информации, генерируемой физической системой. Интегрированная информация, в свою очередь, связана с теоретико—информационной величиной F (Phi), которая, в широком смысле, измеряет, сколько информации генерируется системой в целом по сравнению с ее частями, рассматриваемыми независимо. В ИИТ сознание является неотъемлемым, фундаментальным свойством системы и определяется как природой составляющих ее причинно-следственных механизмов, так и их состоянием60.

В отличие от HotS и GWTs, IIT связывает сознание в первую очередь с задними областями коры головного мозга (так называемой задней "горячей зоной", охватывающей теменную, височную и затылочную области) на том основании, что эти области обладают нейроанатомическими свойствами, которые, предположительно, хорошо подходят для генерации высокого уровня интегрированной информации59. Также, в отличие от GWTs и HotS, которые связывают сознание с аспектами обработки информации в коре головного мозга (то есть функциональными описаниями того, что делает система), IIT не относится к "обработке информации’ как таковой. Вместо этого он связывает сознание со свойствами внутренней причинно–следственной структуры системы, а именно с причинной способностью системы влиять на саму себя. Согласно ИИТ, любая система, генерирующая ненулевой максимум (неснижаемой) интегрированной информации, обладает сознанием, по крайней мере, в некоторой степени.

Из-за этого ИИТ, по-видимому, подразумевает, что уже существуют небиологические системы, обладающие сознанием 61.

IIT является достаточно всеобъемлющим, предлагая описание как глобальных состояний, так и локальных состояний сознания (59, см. рис. 3). Глобальные состояния связаны с количеством несводимой интегрированной информации, генерируемой системой, измеряемой F. Таким образом, ИИТ поощряет одномерную концепцию глобальных состояний, поскольку она приравнивает уровень сознания организма к его значению F. Эмпирический характер локальных состояний может быть понят в терминах "концептуальных структур", которые IIT рассматривает как "формы" в многомерном пространстве, определяемом механистической причинно–следственной структурой системы. Эти формы лежат в основе определенных видов феноменального характера (или идентичны им). Например, пространственная природа визуального восприятия связана с причинно–следственной структурой, определяемой сетчатыми механизмами, присутствующими в ранней зрительной коре [62]. Глобальное единство сознания объясняется интегральным аспектом интегрированной информации — ее ассоциацией с информацией, генерируемой "целым", сверх информации, генерируемой ‘частями’. Наконец, согласно IIT, содержимое является сознательным (а не бессознательным) тогда и только тогда, когда оно включено в причинно–следственный "комплекс" (где комплекс представляет собой подмножество физической системы, которое содержит максимум несводимой интегрированной информации).

Возвращаясь к примеру с бинокулярным соперничеством, ИИТ объясняет, почему испытуемые сообщают, что видели дом (а не лицо), апеллируя к гипотезе о том, что комплекс, лежащий в основе их сообщения, связан с концептуальной структурой, соответствующей содержанию "дом" (а не содержанию "лицо"), и это объясняет, почему испытуемые сообщают о доме (а не о лице). эмпирический контраст между видением дома и лица в терминах ‘формы’ соответствующей концептуальной структуры.

Несмотря на то, что IIT обеспечивает более полное рассмотрение различных аспектов сознания, чем большинство TOCs, в нем сравнительно мало говорится о том, как сознание связано с другими аспектами разума, такими как внимание, обучение и память, и в нем еще не уделено особого внимания важности воплощения и привязанности к окружающей среде для сознания (последнее также относится и к другим аспектам сознания). быть проблемой для HotS и GWT). Тем не менее, теоретики ИИТ сделали первые шаги к решению некоторых из этих проблем, например, разработав показатели "соответствия сложности", которые отслеживают обмен информацией между агентом и его средой, и сформулировав модели на основе агентов, в которых агенты могут эффективно взаимодействовать со своим окружением. для демонстрации увеличенных объемов интегрированной информации62-64.

Повторный вход и прогнозирующая обработка Наконец, мы рассмотрим два общих подхода к пониманию сознания, которые подчеркивают важность нисходящей передачи сигналов в формировании и обеспечении осознанного восприятия. Первые — теории повторного входа — являются теориями сознания как такового и связывают сознательное восприятие с нисходящими (повторяющимися, реентерабельными) сигналами65,66. Вторая группа — теории ПП — не являются первоочередными ТОС, а представляют собой более общие описания функций мозга (и тела), которые могут быть использованы для формулирования объяснений и предсказаний относительно свойств сознания67.

Теории повторного входа основаны на нейрофизиологических данных, раскрывающих важность нисходящей сигнализации для сознательного (обычно визуального) восприятия (примеры см. в ссылках 68-70). В одной из известных теорий повторного входа — теории локальной повторяемости — Ламме утверждает, что локализованной повторяющейся или реентерабельной обработки в перцептивной коре головного мозга достаточно для возникновения сознания (при условии наличия других благоприятных факторов, таких как возбуждение ствола мозга), но теменные и лобные области могут потребоваться для передачи содержания перцептивного опыта или опираясь на них, можно рассуждать и принимать решения65,71 (рис. 4).

Вообще говоря, теории ПП имеют две мотивации. Одна из них связана с рассмотрением проблемы восприятия как одной из причин возникновения сенсорных сигналов 72,73. Другой — на примере принципа свободной энергии74 — апеллирует к фундаментальным ограничениям, касающимся контроля и регулирования, которые применимы ко всем системам, сохраняющим свою организованность с течением времени (75-77, но см. 78). И то, и другое приводит к представлению о том, что мозг реализует процесс "минимизации ошибок предсказания"79, который приближен к байесовскому выводу посредством взаимного обмена (обычно нисходящими) предсказаниями восприятия и (обычно восходящими) ошибками предсказания (80, хотя см. ссылку 81). Некоторые выражения PP, такие как активный логический вывод, добавляют представление о том, что ошибки сенсорного прогнозирования могут быть сведены к минимуму не только путем обновления прогнозов, но и путем выполнения действий для получения ожидаемых сенсорных данных, что позволяет осуществлять прогностический контроль82,83.

Хотя теории PP возникли не как TOCS, было высказано предположение, что они могут обеспечить систематические корреляции между нейронными механизмами и феноменологическими свойствами67, где "систематичность" означает наличие объяснительной силы, основанной на теоретических соображениях, в отличие от простых эмпирических корреляций, как в стандартном подходе NCC. С этой точки зрения, теории ПП отвечают многим требованиям, предъявляемым к ТОС, которые мы изложили ранее, но их лучше всего рассматривать как теории для науки о сознании, а не как теории сознания, поскольку существует множество точек зрения на то, как именно ПП соотносится с сознанием84,85.

Теории ПП обычно рассматривают локальные состояния сознания с точки зрения содержания нисходящих перцептивных предсказаний73,79,86,87: неофициально перцептивное содержание определяется "наилучшей догадкой" мозга о причинах его сенсориума. Эмпирический характер локального состояния определяется природой используемых перцептивных предсказаний. Например, феноменология "объективности" в зрении может быть объяснена условными предсказаниями о сенсорных последствиях действий87,88, тогда как феноменология эмоциональных состояний может быть объяснена ролью интероцептивных предсказаний в регуляции физиологического состояния организма89,90. Пример эмоций показывает, что теории ПП в большей степени, чем другие, рассмотренные здесь, охватывают вопросы, связанные с осознанным саморазвитием13,77,91.

ПП может объяснить различие между сознательным и бессознательным состояниями с точки зрения того, является ли психическое состояние частью текущей "наилучшей догадки" (или оптимального последующего вывода) во время перцептивного вывода. В примере бинокулярного соперничества ПП предусматривает две конкурирующие перцептивные гипотезы (наилучшие предположения), одна из которых "побеждает", что приводит к доминированию восприятия. Сенсорные сигналы из альтернативной гипотезы накапливаются в виде ошибок предсказания, которые в конечном итоге приводят к перцептивному переходу, и в этот момент сенсорные сигналы, объясняемые ранее доминировавшим наилучшим предположением, теперь становятся источником необъяснимой ошибки предсказания, и таким образом цикл повторяется 92,93. (Как уже упоминалось, эмпирический контраст между домом и лицом можно было бы объяснить свойствами соответствующих перцептивных предсказаний.) В тех разновидностях ПП, которые делают упор на активный вывод, изменение содержания сознания может произойти только в том случае, если обновление перцептивных убеждений происходит посредством действия (где действие может быть явным, например, саккадическое движение глаз, или скрытым, например, смещение фокуса внимания)76,94.

Теории ПП, как правило, не имеют дела с глобальными состояниями сознания, но для них было бы естественно апеллировать к целостности соответствующих прогностических процессов при объяснении различий между глобальными состояниями 95, во многом таким же образом, как ГОРЯЧИЕ источники могут апеллировать к целостности соответствующего механизма метапредставления.

Что касается функциональных аспектов сознания, то как повторный вход, так и ПП-подходы дают четкое представление о взаимосвязи между сознанием и вниманием. В теории локальной повторяемости, как и в GWTs, внимание обеспечивает избирательное усиление сенсорных сигналов, так что они достигают префронтальных и теменных областей, обеспечивая сознательный доступ71. В PP внимание ассоциируется с процессом "точного взвешивания", при котором предполагаемая точность сенсорных сигналов модулируется способами, интуитивно эквивалентными изменению отношения сигнал/шум или "усиления" этих сигналов74,96; а при активном умозаключении, как уже упоминалось, выборка внимания может быть необходима для изменения содержания сознания76,94.

Оценка теорий сознания Диапазон данных, к которым обращались в связи с дебатами между конкурирующими теориями сознания, огромен, и мы не можем надеяться представить здесь их полный перечень. Вместо этого мы предлагаем выборочный обзор некоторых текущих дискуссий, подчеркивая разнообразие данных, которые могут быть использованы для оценки КТВ. (Некоторые другие эмпирические данные, обычно используемые в поддержку каждого оглавления, описаны в подписях к рис. 1-4.) В качестве исходного пункта важно признать целостный характер оценки теории.

Теории не подтверждаются ни одним открытием, и, как правило, они не опровергаются ни одним экспериментом. Вместо этого подтверждение теории, как правило, представляет собой постепенный процесс, в ходе которого одна теория выигрывает у своих конкурентов, предоставляя описание целевого явления, которое объясняет широкий спектр данных и может быть интегрировано с успешными теориями в соседних областях97-99.

Одним из очевидных источников ограничений, накладываемых на теорию, является структура сознания.

Хотя в связи с TOCs обсуждался ряд структурных особенностей, одной из структурных особенностей, особенно полезной для противопоставления TOCs, является единство сознания — тот факт, что переживания, которые испытывает отдельный агент в данный момент, по-видимому, всегда возникают как компоненты единого сложного опыта, который полностью отражает то, что происходит в данный момент. это все равно, что быть таким агентом25. Различные TOCs по-разному относятся к единству сознания. IIT уделяет значительное внимание единству сознания. Это не только предполагает, что сознание всегда едино, но и апеллирует к утверждению, что сознание обязательно едино, чтобы мотивировать ассоциацию сознания с (максимумами) несводимой интегрированной информации. Хотя GWT не подчеркивают единство сознания так, как это делает IIT, связь сознания с трансляцией в функционально интегрированном рабочем пространстве предполагает, что у них тоже могут быть ресурсы для обеспечения правдоподобного описания единства сознания. Другие теории TOCs, такие как HotS и теории повторного входа/PP, имеют более двойственное отношение к единству сознания, стремясь либо только подчеркнуть это свойство, либо полностью его игнорировать. Контраст в отношении к единству сознания среди TOCs обусловлен, по крайней мере частично, более фундаментальными разногласиями по поводу того, является ли сознание (обязательно) единым. Хотя единство сознания обещает стать важным ограничением для TOCS, для того, чтобы это обещание было реализовано, нам необходимо лучше понять, в каких отношениях сознание (обязательно) является единым.

Вторым источником ограничений являются нейронные данные. Например, общепризнано, что мозжечок не является ни необходимым, ни достаточным для сознания. Оглавление должно учитывать этот факт и объяснять, почему мозжечок не связан с сознанием. Некоторые TOCs с готовностью дают такое объяснение — например, IIT утверждает, что мозжечок не вовлечен в сознание, поскольку его архитектура плохо подходит для генерирования высокоуровневой интегрированной информации59. Но объяснения такого рода дают конкретную поддержку теории только в том случае, если предоставленная учетная запись является более правдоподобной, чем учетные записи, которые могут быть предоставлены ее конкурентами, и в настоящее время остается открытым вопрос о том, выполняется ли это условие. Например, сторонники HotS могут утверждать, что мозжечку не хватает способности поддерживать метапредставления соответствующего типа; сторонники GWTs могут утверждать, что мозжечок не реализует глобальное рабочее пространство; а теоретики реентерабельности и PP могут указывать на отсутствие в мозжечке интенсивной повторяющейся сигнализации65.

Хотя общепризнано, что ToC должен объяснять, почему мозжечок не вовлечен в сознание, существуют другие виды нейронных данных, которые гораздо более противоречивы с точки зрения TOCs. Важным примером может служить дискуссия о роли префронтальных ("передних отделов мозга’) процессов в сознании.

Используя различные экспериментальные парадигмы, многие исследования с помощью нейровизуализации выявили участие префронтальной коры в сознательном (в отличие от бессознательного) восприятии48, основанное как на региональной активности100,101, так и на функциональной связи между лобной и другими областями 102. Небольшое количество исследований на приматах также показало, что содержание сознания может быть расшифровано по паттернам префронтальной активности во время бинокулярного соперничества, непрерывного подавления вспышек и быстрого последовательного предъявления зрительных стимулов103-105; смотрите также ССЫЛКУ 106 для более сложного изображения, на котором информация, относящаяся к контенту, была декодирована из широкого спектра как активированных, так и деактивированных областей коры головного мозга во время задачи распознавания объектов. Данные о повреждениях и данные, полученные в результате стимуляции мозга, также были использованы для доказательства того, что префронтальная активность имеет решающее значение для сознания; обзор приведен в ссылке 30.

Сторонники HotS и GWTs апеллируют к этим выводам, чтобы подкрепить свои выводы конкурирующими теориями. В ответ сторонники теорий ИИТ и повторного входа утверждают, что наблюдаемая префронтальная активность является (необязательным) следствием сознания и, вероятно, связана с когнитивным доступом к содержимому сознания и способностью предоставлять поведенческие отчеты, а не с сознательным восприятием как таковым 107,108 (но см. ССЫЛКУ 109). Те, кто отстаивает эту точку зрения на "заднюю часть мозга", утверждают, что задних кортикальных процессов - охватывающих части перцептивной и теменной коры и прекунеус - достаточно для восприятия, и что процессы, происходящие в передней части мозга, не являются необходимыми. Это утверждение подтверждается так называемыми исследованиями "без отчета", которые, как правило, выявляют снижение активности префронтальной коры, когда испытуемые не предоставляют четких отчетов о своих ощущениях6,110 (но см. ссылку 111). Сторонники теории "задней части мозга" опираются на положительные данные в пользу тесной связи между задней активностью и сознанием. Например, одно инновационное исследование, в ходе которого изучалось содержание сознания во время сна с использованием парадигмы последовательного пробуждения, показало, что активность в задних отделах коры головного мозга предсказывает, сообщит ли человек о пережитом сновидении как на стадии быстрого, так и на не-быстрой фазах сна112 (но см. ссылку 113). Наконец, интерпретация исследований декодирования с точки зрения "передней части мозга" может быть оспорена, поскольку демонстрация того, что сознательное содержимое может быть "считано" из определенной области, не доказывает, что сам мозг "считывает" это содержимое из этой области таким образом, чтобы оно представляло собой релевантный вид информации. метапредставление или глобальная трансляция.

Хотя некоторые аспекты споров о соотношении переднего и заднего отделов мозга действительно касаются нейробиологических данных - например, мнения о том, где проходят анатомические границы префронтальной коры, расходятся [107,109 — в основе этого лежит разногласие по поводу взаимосвязи между сознанием и когнитивным доступом: разумно ли рассматривать доступность контента для устного сообщения и прямого контроля поведения как показатель сознания, или исследования мозговой основы сознания должны оставаться нейтральными в отношении того, как именно связаны сознание и когнитивный доступ114 (см. также вставку 3)? Дискуссия по этому вопросу отражается в том, как различные TOCs относятся к когнитивному доступу. Авторы GWT ставят когнитивный доступ в центр своего представления о сознании, предполагая не только то, что содержимое сознания всегда доступно для когнитивного доступа, но и то, что процессы, лежащие в основе когнитивного доступа (а именно, зажигание и глобальная трансляция), служат основой сознательного опыта (см. ссылку 111 для получения недавнего нюанса с этой точки зрения). Другие теории, такие как IIT и локальные рекуррентные теории, отрицают тесную взаимосвязь между сознанием и когнитивным доступом, утверждая, что психические состояния могут быть осознанными, не будучи доступными для прямого контроля мыслей и действий, а также что психические состояния в принципе могут быть доступны для прямого контроля мыслей и действий, не будучи доступными для прямого контроля мыслей и действий. сознательный. Хотя подходы более высокого порядка не предполагают какой-либо конкретной взаимосвязи между сознанием и когнитивным доступом, на практике их сторонники обычно предполагают, что содержимое сознания будет доступно когнитивно (например, см. ссылку 46), хотя, возможно, и не наоборот.

Возможно, самым мощным источником данных для оценки конкурирующих TOCS являются новые предсказания. Многие из наиболее значимых событий в истории науки были связаны с подтверждением новых предсказаний115. Например, общая теория относительности получила сильную поддержку благодаря тому факту, что она предсказала как приближение перигелия Меркурия, так и то, как будет отклоняться звездный свет, падающий на поверхность Солнца. Если бы ТоС сделал новые предсказания, которые подтвердились, это вызвало бы сильную поддержку, особенно в сравнении с теориями, которые не смогли сделать соответствующего предсказания или сделали разные и несовместимые прогнозы.

Многие из новых предсказаний, которые делают современные ТОС, трудно проверить.

Например, как в отчетах reentry, так и в отчетах IIT предсказывается, что активность задней коры может поддерживать сознательный опыт без участия передних областей, но в настоящее время у нас нет надежных методов проверки таких утверждений, поскольку проверка основывается на субъективном отчете (или, по крайней мере, на исполнительном контроле в той или иной форме), который, в свою очередь, требует активности передней коры. активность. Более того, ИИТ предсказывает, что сознание широко распространено в природе, в том числе во многих небиологических системах, и может даже проявляться в таких простых системах, как фотодиоды и отдельные атомы (хотя, что интересно, не в нейронных сетях с прямой связью, см. 61). Это предсказание противоречит широко распространенным предположениям о распределении сознания, но оно не может быть разумно оценено в отсутствие надежных методов обнаружения присутствия сознания в таких системах (вставка 3).

В некоторых случаях методологические достижения могут привести к новым предсказаниям, которые можно проверить. Одно из поразительных предсказаний, вытекающих из ИИТ, заключается в том, что изменения в нейронной структуре могут привести к изменениям в сознательном опыте, даже если эти изменения не приводят к изменениям в нейронной активности116. Например, неактивные нейроны в зрительной коре могут способствовать визуальному восприятию, в то время как неактивированные нейроны не будут этого делать. Это предположение возникает потому, что в ИИТ для сознания важна причинно–следственная структура, определяемая нейронными механизмами. Это означает, что если кто–то вмешается в нейронные механизмы, чтобы изменить причинно—следственную структуру, то сознание может измениться, даже если соответствующая нейронная динамика не изменится - предсказание, которое особенно противоречит здравому смыслу в случае, когда динамика отсутствует (то есть для неактивных нейронов). Подобные гипотезы, которые с трудом вытекают из других обсуждаемых здесь теорий, могут быть проверены с помощью точных интервенционных методов, таких как оптогенетика, на животных моделях принятия перцептивных решений117.

Особенно плодотворный подход к оценке конкурирующих TOCs основан на временном профиле сознательной (в отличие от бессознательной) обработки, что отражается, например, в связанных с событиями потенциалах (ERP) в электрофизиологических записях. Некоторые теоретики (например, см. ссылку 118) утверждают, что сознательное восприятие начинается рано (через 120-200 мс) после предъявления стимула, ссылаясь на данные, указывающие на устойчивую корреляцию между осознанным восприятием и ранним началом специфичных для модальности негативных реакций, называемых реакциями негативного осознания, в то же время подвергая сомнению надежность этих данных. ранее обсуждались признаки более позднего появления, такие как P3b (положительный результат ERP, наблюдаемый через ~300 мс после начала стимуляции). Ранняя негативность, отмеченная Дембски и его коллегами, была обнаружена как при зрении, так и при прослушивании, что позволило им утверждать, что существует общая ранняя реакция, которая четко указывает на осознанность восприятия. Такие данные свидетельствуют в пользу ИИТ и локального повторного входа в сознание (но более позднюю кросс-модальную характеристику сознательного восприятия смотрите в ССЫЛКЕ 119). Другие теоретики120,121 приводят доводы в пользу гораздо более позднего появления (примерно через 250-400 мс) перцептивного сознания. Помимо обсуждаемого P3b, поздние сообщения мотивируются различными перцептивными феноменами, которые, по-видимому, соответствуют этим временным рамкам, включая психологический рефрактерный период, мигание внимания и постдиктивные эффекты — последний представляет особый интерес, поскольку показывает, что отсроченный сигнал может ретроспективно активировать сознательное восприятие122. Возможные поздние нейронные признаки сознательного восприятия включают в себя обмен информацией на расстоянии и динамику бифуркаций49,111. Данные, свидетельствующие в пользу позднего появления представлений о перцептивном сознании, в целом подтверждают выводы о более высоком уровне и глобальном рабочем пространстве. Спор между "ранним" и "поздним" описаниями перцептивного опыта, вероятно, останется центральной темой обсуждения в обозримом будущем. Обратите внимание, что вопрос о временном профиле сознательной обработки отличается как от восприятия длительности123, так и от временных характеристик сознательного "момента"124,125, оба из которых отражают аспекты сознательного содержания, которые должны быть объяснены с помощью оглавления.

В настоящее время TOCS, как правило, используются в качестве "повествовательных структур’ в науке о сознании. Несмотря на то, что они помогают интерпретировать нейронные и поведенческие данные, все еще редко при разработке исследования учитываются вопросы подтверждения теории126.

Хотя в использовании теорий таким образом нет ничего плохого, дальнейший прогресс будет зависеть от экспериментов, которые позволят проверить их и устранить неоднозначность.

Мы завершаем наш обзор, определив три проблемы, которые необходимо решить, чтобы в науке о сознании процветал зрелый подход к проверке теорий.

Во-первых, ТОС должны разрабатываться с высокой точностью, поскольку теории, которые апеллируют только к расплывчатым и неточным построениям, могут генерировать только расплывчатые и неточно сформулированные прогнозы. Например, HotS и PP, а также теории повторного вхождения должны определять виды метапредставлений, реентерабельных или прогнозирующих процессов, которые характерны для сознания или конкретных аспектов сознания; ИМ необходимо точно определить их значение для функционального профиля сознания и влияния окружающей среды и воплощения на сознание; и GWTS должны предоставить принципиальный отчет о том, какие рабочие пространства считаются "глобальными" в соответствующем смысле.

Многообещающим подходом здесь является использование вычислительных моделей для придания механистической специфики ТОС, которые, возможно, были сформулированы в относительно абстрактных или концептуальных терминах. В дополнение к разработке точных прогнозов, такие модели могут также обеспечить общий язык, на котором можно сравнивать относительные достоинства конкурирующих ТОС, что может быть особенно полезно для сравнения ТОС, полученных из разных исходных точек. Например, вычислительные модели могли бы выявить общие принципы нисходящей передачи сигналов между HotS и теориями повторного входа и PP, а также прояснить различия между метапредставлением (например, см. ссылку 35) и глобальной трансляцией (например, см. ссылки 127,128), которые отделяют HotS от GWTs129. Разработка вычислительных моделей может также позволить переосмыслить различия между TOCS в терминах (потенциально распределенных) процессов, а не, как это популярно в настоящее время, в терминах широких нейроанатомических областей (например, как в дебатах между "передней частью мозга" и "задней частью мозга’ теоретики111). Ключевой задачей вычислительного подхода является разработка моделей, которые учитывают не только функциональные особенности сознания, но и его феноменологические свойства — задача, которую можно описать общими терминами "вычислительная феноменология" и "вычислительная нейрофеноменология" (примеры см. в ссылках 37,130). Это создает дополнительную проблему, связанную с тем, как проверить правильность или устранить неоднозначность между вычислительными моделями, использующими феноменологические данные (например, см. ссылку 131).; задача, которую можно решить, по крайней мере частично, путем сбора субъективных отчетов на соответствующих уровнях феноменологической детализации (вставка 3).

Помимо того, что ТОС должны быть более конкретными, они также должны быть более всеобъемлющими. По большей части, TOCS, как правило, фокусируются на определенных видах локальных состояний (перцептивный опыт с акцентом на зрение), на определенных видах глобальных состояний (обычное бодрствующее осознание) и на определенных видах сознательных существ (взрослые люди). Хотя есть веские причины, по которым теоретики склонны сосредотачиваться на ограниченном классе состояний сознания и существ — доступность для экспериментов является важным фактором, — полностью всеобъемлющий ToC должен отдавать должное богатому разнообразию сознания. Что касается локальных состояний, то ТОС должны выходить за рамки восприятия и учитывать также, например, аффект, темпоральность, волю и мышление.

Что касается глобальных состояний, то TOCS должны выходить за рамки обычного бодрствования и учитывать также отличительные режимы сознания, связанные, например, со сновидениями, медитацией, расстройствами сознания и психоделическим состоянием. Что касается сознательных существ, то TOCs должны выходить за рамки опыта взрослых и рассматривать вопросы, касающиеся сознания у человеческих младенцев, животных, не являющихся людьми, и даже искусственных систем. Хотя нет ничего плохого в том, что TOCS имеют ограниченную направленность, теории, которые дают более полное представление о сознании, имеют очевидные преимущества перед теми, которые этого не делают, особенно если они могут выявить объяснительные связи между различными аспектами сознания.

Третий вопрос, который необходимо решить, - это проблема измерения: как можно определить достоверные показатели сознания132. Решение этой проблемы имеет решающее значение, поскольку подробные и всеобъемлющие TOCs вряд ли принесут большую пользу, если у нас не будет возможности проверить их прогнозы. Полезно различать две (тесно связанные) версии проблемы измерения. Первая касается обнаружения сознательного содержания. Здесь основная задача состоит в том, чтобы определить способы различения сознательных и бессознательных психических состояний, которые не приводят к противоречивым предположениям о функциональном профиле сознания (например, о том, что содержание сознания должно быть отчетным или иным образом доступным для когнитивного контроля высокого уровня)114,133,134. Другая версия проблемы измерения фокусируется не на содержимом, а на существах. Вопросы здесь включают в себя то, как мы могли бы определить распределение сознания в животном царстве.135; обладают ли сознанием определенные классы церебральных органоидов136 или систем искусственного интеллекта135-137; когда сознание впервые появляется в онтогенезе138; и когда оно сохраняется в контексте черепно-мозговой травмы139. Здесь также задача состоит в разработке способов измерения сознания, позволяющих избежать противоречивых предположений о его функциональном профиле (вставка 3).

Конечно, вышеупомянутые проблемы уже решаются, в той или иной степени, исследователями сознания. Эти усилия в настоящее время дополняются такими инициативами, как модель состязательного сотрудничества, которая поощряет сторонников TOCs к разработке экспериментов с конкретной целью дифференциации альтернативных TOCS17.

Сознание остается предметом научных споров, однако есть все основания полагать, что последовательная разработка, тестирование и сравнение TOCs приведут к гораздо более глубокому пониманию этой самой главной из тайн.