

# Готовясь к встрече с проблемой сознания<sup>1</sup>

*Дэвид Дж. Чалмерс*

## 1. Введение

Сознание ставит наиболее трудные проблемы в науке об уме. Нет ничего более близкого нам, чем наш сознательный опыт и нет ничего, что было бы труднее объяснить. В последние годы все виды ментальных явлений поддались научному исследованию, но сознание упрямо сопротивляется. Многие пробовали объяснять его, но, по-видимому, не достигли своей цели. Некоторые пришли к выводу о том, что эта проблема неразрешима, и что приемлемого объяснения здесь нет.

Для того, чтобы добиться прогресса в проблеме сознания, мы должны посмотреть ей прямо в лицо. В этой статье я сначала сформулирую действительно трудную часть проблемы, отделив ее от более простых частей и обосновав, почему ее так трудно объяснять. Я подвергну критике некоторые новые работы, в которых используются редукционистские подходы к сознанию, и покажу, что эти подходы неизбежно окажутся неспособными постигнуть наиболее трудную часть проблемы. Как только эта неудача будет осознана, то откроется дверь для дальнейшего прогресса. Во второй половине этой статьи, я утверждаю, что, если мы двигаемся в направлении нового типа нередукционистского подхода, то сознанию можно дать натуралистическое объяснение. Я выдвигаю моего собственного кандидата на такое объяснение: нередукционистскую теорию, основанную на принципах структурной согласованности и организационной инвариантности, а также на двухаспектном представлении информации.

## 2. Легкие проблемы и трудная проблема

Не существует какой-то одной проблемы сознания. "Сознание" - это неоднозначный термин, относимый ко многим различным явлениям. Каждое из этих явлений должно быть объяснено, но некоторые поддаются объяснению легче, чем другие. В начале полезно разделить связанные с сознанием проблемы на "трудные" и "легкие". Легкие проблемы сознания - это те, которые, по-видимому, поддаются стандартным методам когнитивной науки, объясняющей явление в терминах вычислительных или нейронных механизмов. Трудные проблемы - это те, которые, по-видимому, не поддаются этим методам.

Легкие проблемы сознания связаны с объяснением следующих явлений:

- способность различать, категоризировать, и реагировать на внешние стимулы;
- интеграция информации когнитивной системой;
- способность описывать ментальные состояния;
- доступность для системы ее собственных внутренних состояний;
- фокус внимания;
- преднамеренный контроль поведения;
- различие между бодрствованием и сном.

Все эти явления ассоциируются с понятием сознания. Например, иногда говорят, что ментальное состояние является сознательным, когда оно выражено в словах, или когда оно внутренне доступно. Иногда систему считают осознающей некоторую информацию, когда

1 D.Chalmers. Facing up to the problem of consciousness//Journal of consciousness studies, 2 (3), 1995, pp. 200-219

она способна реагировать на основе этой информации или, более точно, когда она уделяет внимание этой информации, или когда она может ее интегрировать и использовать в сложном контроле поведения. Иногда мы говорим, что действие является осознанным именно тогда, когда оно преднамеренно. Часто, мы говорим, что некоторый организм сознателен, имея в виду, что это эквивалентно бодрствованию.

Нет никакой реальной проблемы в возможности объяснения *этих* явлений с научной точки зрения. Все они непосредственно поддаются объяснению в терминах вычислительных или нейтральных механизмов. Для того, чтобы объяснить, например, доступ к внутренним состояниям и способность их описывать, нам нужно только определить механизм, посредством которого информация о внутренних состояниях обнаруживается и становится доступной для вербального описания. Для того, чтобы объяснить интеграцию информации, нам нужно только показать механизмы, посредством которых информация присоединяется к последующим процессам и используется ими. Для объяснения сна и бодрствования будет достаточно объяснить соответствующие нейрофизиологические процессы, которые отвечают за столь контрастное поведение организма в этих состояниях. Несомненно, что для каждого случая найдется подходящая когнитивная или нейрофизиологическая модель, осуществляющая объяснительную работу.

Если бы только эти явления составляли сознание, то оно не было бы большой проблемой. Хотя мы до сих пор еще не приблизились к их полному объяснению, но у нас есть ясное представление о том, как мы могли бы это сделать. Именно поэтому я и называю эти проблемы легкими. Конечно, "легкий" - это относительное понятие. Для прояснения деталей, вероятно, потребуется столетие или два трудной эмпирической работы. Тем не менее, есть все основания полагать, что методы когнитивной науки и нейрологии приведут к успеху.

Действительно трудная проблема сознания - это проблема *переживания*. Когда мы воспринимаем и думаем, то имеет место шум от обработки информации, но существует также и субъективный аспект. Как выразился Нагель (1974), *существует нечто, похожее* на сознательный организм. Этот субъективный аспект и есть переживание. Когда, например, мы смотрим, то мы *переживаем* визуальные ощущения: ощущение качества красного, переживание тьмы и света, качества глубины в зрительном поле. Другие переживания происходят параллельно с восприятием в различных модальностях: звук кларнета, запах нафталиновых шариков. Кроме этого есть телесные ощущения, от боли до оргазмов, возникающие внутри ментальные образы, чувство качества эмоции и переживание потока сознательного мышления. Объединяет все эти состояния то, что в них, по-видимому, что-то присутствует. Все это - состояния переживания.

Бесспорно, некоторые организмы являются субъектами переживания. Но вопрос о том, каким образом эти системы являются субъектами переживания, ставит в тупик. Почему, когда наши когнитивные системы обрабатывают визуальную и слуховую информацию, у нас есть зрительные и слуховые переживания: качества глубокой синевы, ощущение среднего "С"? Как можно объяснить, почему как бы есть что-то, что развлекается ментальным образом или переживает эмоцию? Принято считать, что переживание возникает из физического основания, но у нас нет хорошего объяснения тому, почему и как это происходит. Почему физические процессы вообще должны порождать богатую внутреннюю жизнь? Объективно кажется неразумным, что они должны это делать, но все же это происходит.

Если обозначать какую-то проблему как проблему сознания, то это именно *эта* проблема. В этом центральном смысле термина "сознание" организм сознателен, если существует нечто, по-видимому, являющееся этим организмом, и ментальное состояние является сознательным, если существует нечто, по-видимому, находящееся в этом состоянии. Иногда в этом контексте также используются термины "феноменальное сознание" и "qualia", но мне кажется более естественным говорить о "сознательном переживании" или просто о "переживании". Другой полезный способ избежать недоразумений (использованный, например, Newell 1990, Chalmers 1995) состоит в том, чтобы зарезервировать термин

«сознание» для явлений переживания, используя менее нагруженный термин «осознание» для более простых явлений, описанных ранее. Если бы такое соглашение было общепринятым, то общаться было бы гораздо легче, т.к. говорящие о «сознании» в настоящее время часто не понимают друг друга.

Неопределенность термина «сознание» часто используется как философами, так и учеными, пишущими об этом предмете. Нетрудно найти работу по сознанию, которая начинается с обращения к тайне сознания, странной неосвязаемости и невыразимости субъективности, и озабоченности тем, что до сих пор у нас нет никакой теории этого явления. Понятно, что темой здесь является трудная проблема, проблема переживания. Во второй половине работы тон становится более оптимистическим и кратко излагается собственная теория сознания. При ближайшем же рассмотрении оказывается, что это теория одного из более простых явлений: способности описывать, интроспективного доступа и т.д. В заключении же автор заявляет, что сознание в конце концов все же поддалось, но у читателя остается ощущение иллюзорности этой победы. Трудная проблема осталась нетронутой.

### 3. Функциональное объяснение

Почему легкие проблемы легки, и почему трудная проблема трудна? Легкие проблемы легки именно потому, что они связаны с объяснением когнитивных способностей и функций. Для того, чтобы объяснять когнитивную функцию, нам нужно только определить механизм, который может ее осуществлять. Методы когнитивной науки хорошо подходят для этого вида объяснения, и поэтому хорошо подходят для решения легких проблем сознания. Напротив, трудная проблема трудна именно потому, что это - не проблема выполнения функций. Эта проблема остается даже тогда, когда объяснены все имеющие отношение к делу функции. (Здесь понятие "функция" используется не в узком телеологическом смысле, как нечто такое, для чего создана система, но в более широком смысле любой причинной роли в осуществлении возможного поведения системы.)

Для объяснения способности к описанию, например, нужно просто объяснить, как система может осуществлять функцию создания сообщений о внутренних состояниях. Для того, чтобы объяснить внутренний доступ, нам нужно только объяснить, как внутренние состояния могли бы подходящим образом воздействовать на систему и как она могла бы использовать информацию об этих состояниях в управлении последующими процессами. Для того, чтобы объяснить интеграцию и контроль, нам нужно объяснить, как центральные процессы системы могут собирать информацию и использовать ее в различных типах поведения. Все это - проблемы объяснения функций.

Как мы объясняем выполнение функции? Определяя *механизм*, который ее выполняет. Для выполнения этой задачи идеально подходит нейрофизиологическое и когнитивное моделирование. Если нам нужно детальное объяснение более низкого уровня, то мы можем определить нейронный механизм, который отвечает за эту функцию. Если нам нужно более абстрактное объяснение, то мы можем определить механизм в вычислительных терминах. Любым путем мы придем к полному и удовлетворительному объяснению. Как только мы определили нейронный или вычислительный механизм, который, например, осуществляет функцию вербального описания, то большая часть нашей работы по объяснению способности к описанию выполнена.

До известной степени этот момент тривиален. В этом состоит *концептуальное* утверждение относительно этих явлений, состоящее в том, что их объяснение связано только с объяснением различных функций, поскольку эти явления *функционально определимы*. Для установления в системе самоописания *все* это означает, что система обладает способностью к вербальному описанию внутренней информации. Для поддержания бодрствования системы все это означает, что она должна быть подходящим образом восприимчива к информации, приходящей из окружающей среды, и способна использовать подходящим образом эту информацию для управления своим поведением. Для того, чтобы понять, что это некоторый

концептуальный момент, заметим, что тот, кто говорит: "Вы объяснили выполнение функции вербального описания, но Вы не объяснили саму способность к вербальному описанию", - делает тривиальную концептуальную ошибку относительно способности к вербальному описанию. Все, что могло бы потребоваться для ее объяснения, это объяснение того, как работает соответствующая функция; то же самое верно для других рассматриваемых явлений.

В науках высокого уровня редуцирующее объяснение работает именно таким способом. Чтобы объяснить, например, ген, нам нужно определить механизм, который хранит и передает наследственную информацию от одного поколения к другому. Оказывается, что эту функцию выполняет ДНК; как только мы объяснили, как работает эта функция, мы объяснили ген. Для объяснения жизни нам нужно в конечном счете объяснить, как система может воспроизводить, приспосабливаться к окружающей среде, осуществлять обмен веществ и так далее. Все это вопросы, касающиеся выполнения функций, и поэтому они хорошо подходят для редуцирующего объяснения. То же самое верно для большинства проблем когнитивной науки. Чтобы объяснить обучаемость, мы должны объяснить способ, которым поведенческие способности системы изменяются в связи с информацией, поступающей из окружающей среды, и способ, которым новая информация может поддерживать адаптацию системы к окружающей среде. Если мы показали, как нейронный или вычислительный механизм осуществляет эту работу, то мы объяснили обучаемость. То же самое мы можем сказать о других когнитивных явлениях типа восприятия, памяти и языка. Иногда соответствующие функции должны быть охарактеризованы весьма тонко, но ясно, что до тех пор когнитивная наука вообще объясняет эти явления, она делает это, объясняя выполнение функций.

Когда дело доходит до сознательного переживания, то этот вид объяснения перестает работать. Трудной и почти уникальной трудную проблему делает именно то, что она выходит *за пределы* выполнения функций. Для того, чтобы это понять, обратите внимание на то, что даже когда мы объяснили выполнение всех когнитивных и поведенческих функций в близких к переживанию: перцептивное различение, классификацию, внутренний доступ, вербальное описание, - без ответа все же может остаться следующий дополнительный вопрос: «*Почему выполнение этих функций сопровождается переживанием?*» Простое объяснение этих функций оставляет этот дополнительный вопрос открытым.

При объяснении генов, или жизни, или обучаемости аналогичный дополнительный вопрос отсутствует. Если говорят: "Я могу понять, что Вы объяснили, как ДНК хранит и передает наследственную информацию от одного поколения к другому, но Вы не объяснили *ген*", - то при этом делается концептуальная ошибка. Все, что означает быть геном, это быть единицей, которая выполняет функцию хранения и передачи информации. Но если говорится: "Я могу понять, что Вы объяснили, как информация различается, интегрируется и сообщается, но Вы не объяснили, как это *пережито*", - то при этом не делается концептуальная ошибка. Это - нетривиальный дополнительный вопрос.

Этот дополнительный вопрос является ключевым в проблеме сознания. Почему вся обработка информации не осуществляется "в темноте", свободной от какого-либо внутреннего чувства? Почему, когда электромагнитные волны воздействуют на сетчатку, различаются и классифицируются зрительной системой, это различение и классификация переживается как ярко красное? Мы знаем, что сознательное переживание *действительно* возникает, когда выполнены эти функции, но сам этот факт, что оно возникает, и является главной тайной. Существует *объяснительный разрыв* (этот термин принадлежит Levine 1983) между функциями и переживанием, и для его преодоления нам нужен объяснительный мост. Простое объяснение функций находится по одну сторону этого разрыва, поэтому материалы для моста необходимо искать в другом месте.

Это не означает, что переживание *не имеет* никакой функции. Возможно, окажется, что оно играет важную когнитивную роль. Но какую бы роль оно ни играло, объяснение переживания - это нечто большее, чем простое объяснение функции. Возможно, даже

окажется, что в процессе объяснения функции, мы придем к ключевому инсайту, который позволит нам объяснить переживание. Если это открытие даже и случится, то оно будет *сверх-объяснительной* наградой. Не существует такой когнитивной функции, о которой мы можем сказать заранее, что ее объяснение *автоматически* объяснит переживание.

Для объяснения переживания нам нужен новый подход. Обычные объяснительные методы когнитивной науки и нейрологии недостаточны. Эти методы были развиты именно для того, чтобы объяснить выполнение когнитивных функций, и они это делают хорошо. Но как таковые, они оснащены *именно* так, чтобы объяснять выполнение функций. Когда же речь заходит о трудной проблеме, то стандартному подходу нечего сказать.

#### 4. Некоторые конкретные исследования

В последние годы в целом ряде работ проблема сознания исследовалась на основе когнитивной науки и нейрологии. Это могло бы означать, что проделанный выше анализ неточен, но в действительности тщательное изучение соответствующих работ только подтверждает этот анализ. Когда мы просто изучаем то, с каких аспектов сознания начинаются эти работы, и какими аспектами они заканчивают объяснение, то мы обнаруживаем, что целью объяснения всегда является одна из легких проблем. Я проиллюстрирую это утверждение двумя репрезентативными примерами.

Первый пример - это "нейробиологическая теория сознания" намеченная Криком и Кохом (1990; см. также Crick 1994). Эта теория сосредоточена на определенных нейронных колебаниях в коре головного мозга в диапазоне частот 35-75 герц; Крик и Кох предполагают, что эти колебания составляют основу сознания. Отчасти это связано с тем, что эти колебания, по-видимому, коррелируют с осознанием в целом ряде различных модальностей, например, в зрительной и обонятельной системах, а также, потому что они содержат механизм, посредством которого может быть осуществлено *объединение* различных единиц информации. Объединение - это процесс, который соединяет воедино отдельно представленные части информации об одном объекте для использования в дальнейшей обработке, как, например, когда интегрируется информация от различных зрительных путей о цвете и форме воспринимаемого объекта. Идя в русле других работ (например, Eckhorn and *et al* 1988), Крик и Кох предположили, что объединение может быть достигнуто синхронизацией колебаний нейронных групп, представляющих релевантные содержания. Когда две части информации объединяются вместе, то соответствующие группы нейронов будут осциллировать с одинаковой частотой и фазой.

Детали того, как это объединение могло бы быть достигнуто, все еще плохо изучены, но предполагается, что эта проблема разрешима. Что могла бы объяснить такая теория? Ясно, что она могла бы объяснить объединение информационных содержаний, и, возможно, могла бы привести к более общему пониманию процесса интеграции информации в мозге. Крик и Кох также предлагают, что эти колебания активизируют механизмы кратковременной памяти, поэтому они могут служить для объяснения этой и других форм памяти. Эта теория в конечном счете могла бы привести к общему пониманию того, как воспринятая информация объединяется и сохраняется в памяти для последующего использования.

Такая теория была бы ценной, но она ничего бы не сообщила нам о том, почему соответствующие содержания переживаются. Крик и Кох предполагают, что эти колебания являются нейронными *коррелятами* переживания. Это утверждение спорно: разве объединение не происходит также при бессознательной обработке информации? - но даже если это и принято, то остается *объяснительный* вопрос: «Почему эти колебания порождают переживание?» Единственное основание для объяснительной связи - это та роль, которую она играет в объединении и хранении информации, но вопрос о том, почему объединение и хранение должны сами по себе сопровождаться переживанием, никогда не ставился. Если мы не знаем, почему объединение и хранение должны породить переживание, то история о колебаниях не может нам помочь. Напротив, если бы мы знали, почему объединение и

хранение порождают переживание, то нейрофизиологические детали были бы просто сахарной глазурью на торте. Теория Крика и Коха находит свою точку опоры в предположении о связи между объединением и переживанием, и поэтому она ничего не может дать для объяснения этой связи.

Я не думаю, что Крик и Кох претендуют в конечном счете на трудную проблему, хотя некоторые авторы интерпретировали их иначе. Опубликованное интервью с Кохом дает ясное представление об ограничениях в притязаниях этой теории.

Давайте сначала забудем о действительно трудных вопросах, подобных чувствам, поскольку у них может и не быть научного объяснения. Субъективные состояние игры, боли, удовольствия, видения синего цвета, запаха розы - здесь, по-видимому, есть огромный скачок между материалистическим уровнем объяснения в терминах молекул и нейронов, и субъективным уровнем. Давайте сконцентрируемся на том, что легче исследовать, например, на зрительном осознании. Теперь Вы говорите со мной, но Вы не смотрите на меня, Вы смотрите на капучино, и поэтому Вы осознаете его. Вы можете сказать: " Это - чашка и в ней есть какая-то жидкость". Если я дам ее Вам, то Вы протянете руку и возьмете ее. Ваш ответ будет осмысленным. Это - то, что я называю осознанием. (" Что такое Сознание", *Discover*, Ноябрь 1992, с. 96)

Второй пример - подход на уровне когнитивной психологии. Это теория глобального рабочего пространства сознания Баарса, изложенная в его книге "Когнитивная Теория Сознания". Согласно этой теории, содержания сознания находятся в *глобальном рабочем пространстве*, центральном процессоре, соединяющем множество специализированных несознательных процессоров. Когда этим специализированным процессорам нужно распространить информацию на остальную часть системы, то они посылают ее в рабочее пространство, действующее для остальной части системы как своего рода коммуникационная доска, доступная для всех других процессоров.

Баарс использует эту модель для исследования многих аспектов человеческого познания и объяснения различий между сознательным и бессознательным познавательным функционированием. В конечном счете, однако, это теория *когнитивной доступности*, объясняющая то, как некоторые информационные содержания оказываются доступными всей системе, а также это теория информационной интеграции и способности к описанию. Эта теория перспективна, как теория осознания, функционального коррелята сознательного опыта, но она не дает объяснения самого переживания.

Можно было бы предположить, что согласно этой теории, содержания переживания - это в точности содержания рабочего пространства. Но даже если это и так, то ничто внутри этой теории не *объясняет*, почему переживается информация, содержащаяся в глобальном рабочем пространстве. Лучшее, что может дать эта теория, это утверждать, что информация переживается потому, что она *глобально доступна*. Но теперь тот же вопрос возникает в другой форме: почему глобальная доступность должна вызвать сознательное переживание? Как всегда, этот соединяющий вопрос остается без ответа.

Почти все выполненные в последние годы работы, которые основаны на когнитивном и нейрологическом подходах к сознанию, могут быть подвергнуты подобной критике. Например, в модели "Нейронного Дарвинизма" Эдельман (1989) изучает вопросы, относящиеся к перцептивному осознанию и концепции "Я", но ничего не говорит о том, почему при этом также должно быть переживание. Модель "Множественных проектов" Деннета(1991) в значительной степени направлена на объяснение способности к описанию определенных ментальных содержаний. Теория "Промежуточного уровня" Джакендорфа(1988) дает объяснение некоторых вычислительных процессов, которые лежат в основе сознания, но Джакендорф подчеркивает, что вопрос о том, как это "проецируется" в сознательное переживание остается непостижимым.

Исследователи, использующие эти методы, часто не эксплицируют свое отношение к проблеме сознательного переживания, хотя иногда они занимают и определенную позицию. Но даже и в этом случае встречаются очень разные точки зрения. Возможны различные

стратегии в исследованиях проблемы переживания. Было бы полезно, если бы соответствующие стратегические выборы делались более эксплицитно.

Первая стратегия состоит в том, чтобы просто *объяснять что-то другое*. Некоторые исследователи эксплицитно утверждают, что проблема переживания пока слишком трудна, и, возможно, даже в целом. Вместо этого они обращаются к более легко поддающимся проблемам типа способности к описанию или концепции "Я". Хотя я назвал эти проблемы "легкими", но они являются одними из наиболее интересных нерешенных проблем когнитивной науки, поэтому это несомненно стоящая работа. Самое худшее, что можно сказать об этом выборе, это неопределенность претензий в связи с исследованием сознания, что иногда влечет неверную интерпретацию работы.

Второй выбор состоит в том, чтобы рассмотреть более жесткую линию и *отрицать это явление*. (Разновидности этого подхода см. в Allport 1988, Dennett 1991, и Wilkes 1988.) Согласно этой линии, как только мы объяснили функции типа доступности, самоописания и т.п., то больше нет никакого явления, называемого "переживанием", которое нужно было бы объяснять. Некоторые исследователи явно отрицают это явление, придерживаясь, например, той точки зрения, что то, что не поддается проверке извне, не существует. Другие добиваются того же самого эффекта, позволяя переживанию существовать, но только, если "переживание" отождествляется с чем-то вроде способности различать и сообщать. Эти подходы ведут к более простой теории, но в конечном счете они являются неудовлетворительными. Переживание - это наиболее важный и проявленный аспект нашей ментальной жизни, и в действительности это, возможно, ключевое подлежащее объяснению понятие в науке об уме. Из-за своего статуса подлежащего объяснению переживание, подобно жизненному духу, не может быть отброшено, когда приходит новая теория. Скорее это центральный факт, который любая теория сознания должна объяснить. Теория, которая отвергает это явление, "решает" проблему, уклоняясь от вопроса.

В третьем случае некоторые исследователи заявляют, что в полной мере *объясняют переживание*. Они (в отличие от предыдущих) хотят рассматривать его очень серьезно; они представляют свою функциональную модель или теорию и утверждают, что она полностью объясняет субъективность переживания (например, Flohr 1992, Humphrey 1992). Соответствующий шаг в объяснении, однако, обычно осуществляется быстро и, как правило, выглядит как нечто магическое. После того, как сообщаются некоторые детали относительно обработки информации, на сцене внезапно появляется переживание, но остается неясным, как эти процессы приводят к его появлению. Возможно, это появление считается само собой разумеющимся, но тогда мы имеем неполное объяснение и разновидность пятой стратегии, обсуждаемой ниже.

Четвертый, более многообещающий подход, состоит в том, чтобы обратиться к этим методам в связи с *объяснением структуры переживания*. Например, доказывається, что различение, осуществленное зрительной системой, может объяснить структурные отношения между различными цветовыми переживаниями, а также геометрическую структуру зрительного поля (см., например, Clark 1992 и Hardin 1992). В общем некоторые факты, касающиеся структуры процессов обработки информации, будут соответствовать и объяснять факты о структуре переживания. Эта стратегия приемлема, но ограничена. В лучшем случае она принимает существование переживания как само собой разумеющееся и объясняет некоторые факты относительно его структуры, представляя собой один из типов нередуктивного объяснения структурных аспектов опыта (позже я буду более подробно говорить об этом). Эта стратегия полезна для многих целей, но она ничего не говорит нам о том, почему при этом должно быть переживание.

Пятая и наиболее приемлемая стратегия состоит в том, чтобы *изолировать субстрат переживания*. В конечном счете, почти все допускают, что так или иначе переживание *возникает* из процессов, протекающих в головном мозге, и имеет смысл определить тот тип процесса, из которого оно возникает. Крик и Кох, например, утверждают, что они обнаружили нейронный коррелят сознания, Эдельман (1989) и Джанкедорфф (1988) делают

аналогичные заявления. Обоснованность этих заявлений требует тщательного теоретического анализа, тем более, что переживание непосредственно не наблюдаемо в экспериментальных контекстах, но, осторожно применяемая, эта стратегия сможет пролить дополнительный свет на проблему переживания. Однако, ясно, что она неполна. Для построения удовлетворительной теории, нам нужно знать не только то, *какой* процесс приводит к возникновению переживания, но также, почему и как. Полная теория сознания должна построить объяснительный мост.

## 5. Дополнительный ингредиент

Мы видели, что имеются систематические причины, по которым обычные методы когнитивной науки и нейрологии не могут объяснить сознательное переживание. Они просто представляют собой неверные методы: ничто из того, что они могут нам дать, не может привести к объяснению. Для объяснения сознательного переживания нам нужен *дополнительный ингредиент*. Это утверждение бросает вызов тем, кто серьезно относится к трудной проблеме сознания: каков ваш дополнительный ингредиент, и почему *именно он* должен объяснить сознательное переживание?

Недостатка в предлагаемых дополнительных ингредиентах нет. Некоторые предлагают введение хаоса и нелинейной динамики. Некоторые думают, что ключом является неалгоритмическая обработка информации. Некоторые надеются на будущие открытия в нейрофизиологии. Некоторые предполагают, что ключ к этой тайне находится на уровне квантовой механики. Нетрудно понять, почему выдвигаются все эти предположения. Ни один из старых методов не работает, поэтому решение нужно искать в *чем-то* новом. К сожалению, все эти предположения страдают от тех же самых старых проблем.

Неалгоритмическая обработка информации, например, выдвинута Пенроузом (1989; 1994) в связи с той ролью, которую она могла бы играть в процессе сознательного математического инсайта. Аргументы относительно математики являются спорными, но даже если они оправданы и получено объяснение неалгоритмической обработки информации в человеческом мозге, то все равно это будет только объяснением *функций*, связанных с математическим рассуждением и т.п.. Для неалгоритмического процесса, так же как и для алгоритмического процесса, главный вопрос остается без ответа: почему этот процесс должен вызывать переживание? Неалгоритмическая обработка информации не играет никакой особой роли в ответе на *этот* вопрос.

То же самое справедливо относительно нелинейной и стохастической динамики. Они могли бы дать новое объяснение динамики функционирования когнитивных способностей, весьма отличное от того, что дается стандартными методами в когнитивных науках. Но из динамики можно получить только еще больше динамики. Вопрос же относительно переживания здесь столь же таинственен, как и везде. В связи с новыми открытиями в нейрофизиологии этот момент еще более понятен. Эти открытия могут помочь нам существенно продвинуться в понимании функционирования мозга, но в отношении любого нейронного процесса, который мы изолируем, всегда возникнет тот же самый вопрос. Трудно себе представить, что еще ожидает сторонник нового нейрофизиологии сверх объяснения когнитивных функций. Ведь мы же не обнаружим внезапно необыкновенное сияние внутри нейрона!

Возможно, наиболее популярным "дополнительным ингредиентом" является квантовая механика (см., например, Nameroff 1994). Привлекательность квантовых теорий сознания может происходить из Закона Минимизации Тайны: сознание - это тайна и квантовая механика - это тайна, поэтому, возможно, что эти две тайны имеют общий источник. Однако, квантовые теории сознания страдают от тех же самых трудностей, что и нейронные или вычислительные теории. Квантовые явления имеют некоторые замечательные функциональные свойства, типа индетерминизма и нелокальности. Естественно думать, что эти свойства могут играть некоторую роль в объяснении когнитивных функций таких, как



случайный выбор и интеграция информации, и поэтому эти гипотезы не могут быть исключены *априорно*. Но когда дело доходит до объяснения переживания, то квантовые процессы оказываются в той же самой лодке, что и любые другие. Вопрос о том, почему эти процессы должны вызвать переживание, остается полностью без ответа.

(Особая привлекательность квантовых теорий связана с тем фактом, что в некоторых интерпретациях квантовой механики сознание играет активную роль в "коллапсе" квантовой волновой функции. Такие интерпретации спорны, но в любом случае они не дают никакой надежды относительно *объяснения* сознания в терминах квантовых процессов. Скорее, в этих теориях существование сознания *предполагается*, а затем используется для объяснения квантовых процессов. В лучшем случае эти теории сообщают нам кое-что относительно физической роли, которую может играть сознание. Они ничего не сообщают нам о том, как оно возникает.)

В конечном счете та же самая критика применима к *любому* чисто физическому объяснению сознания. Для любого заданного нами физического процесса следующий вопрос останется без ответа: почему этот процесс должен вызывать переживание? В отношении любого процесса такого типа концептуально оправданно утверждать, что он мог бы происходить и в отсутствие переживания. Из этого следует, что простой учет физического процесса не скажет нам, почему возникает переживание. Проявление переживания выходит за пределы того, что может быть получено из физической теории.

Чисто физическое объяснение хорошо подходит для физических *структур*, объясняя макроскопические структуры в терминах микроструктурных элементов; оно дает удовлетворительное объяснение выполнению *функций*, объясняя эти функции в терминах физических механизмов, которые их осуществляют. Это связано с тем, что из физического основания можно *извлечь* только факты относительно структур и функций: как только даны внутренние детали физического процесса, его структурные и функциональные свойства получаются как автоматическое следствие. Но структура и динамика физических процессов производят только еще больше структуры и динамики, поэтому структуры и функции - это все, что мы можем ожидать от объяснения на основе этих процессов. Факты относительно переживания не могут быть автоматическим следствием какого-либо физического процесса, поскольку концептуально последовательно утверждать, что любой данный процесс мог бы существовать и без переживания. Переживание может *возникать* из физического, но не может быть *следствием* физического.

Мораль, извлекаемая из всего этого, состоит в том, что *вы не можете "дешево" объяснить сознательное переживание*. Замечательный факт состоит в том, что редуктивные методы, т.е. методы, объясняющие явление высокого уровня полностью в терминах более фундаментальных физических процессов, хорошо работают во многих областях. В некотором смысле большинство биологических и когнитивных явлений *можно* объяснить "дешево". А именно в том смысле, что они являются автоматическим следствием более фундаментальных процессов. Было бы замечательно, если бы редуктивные методы могли также объяснить и переживание (в течение длительного времени я надеялся, что это так). К сожалению, имеются систематические причины, по которым эти методы должны потерпеть неудачу. Редуктивные методы приводят к успеху в большинстве областей потому, что структуры и функции - это то, что нам нужно объяснить в этих областях и это то, что может быть вызвано физическим. Когда возникает проблема, выходящая за пределы объяснения структур и функций, то эти методы оказываются бессильными.

Это может напомнить нам заявление виталиста, что жизнь нельзя объяснить с позиций физики, но это разные случаи. В своем скептицизме виталист руководствовался сомнением в том, что физические механизмы могут обеспечить выполнение многих замечательных функций, связанных с жизнью: сложное адаптивное поведение и репродуктивность. Концептуальное требование, состоящее в том, что объяснение функций - это все, что здесь необходимо, принималось имплицитно, а недостаток детального знания биохимических механизмов заставлял виталиста сомневаться в том, что какой-либо физический процесс

может выполнить необходимую работу, и выдвигать в качестве альтернативного объяснения гипотезу жизненного духа. Как только оказывается, что физические процессы могут обеспечивать выполнение соответствующих функций, сомнения виталиста исчезают.

В случае переживания, с другой стороны, физическое объяснение функций не подвергается сомнению. Ключевым является действительно *концептуальное* утверждение о том, что объяснения функций не достаточно для объяснения переживания. Это - основополагающая концептуальная позиция, на которую не повлияют дальнейшие нейрологические исследования. То же самое верно относительно различия между переживанием и *жизненным духом*. Жизненный дух был выдвинут в качестве объяснительного основания для соответствующих функций, и поэтому может быть отвергнут, если эти функции объясняются без него. Переживание - это не объяснительное основание, но объясняемое на своей собственной основе, поэтому оно не является кандидатом на такое устранение.

Заманчиво было бы утверждать, что, в конечном счете, все виды загадочных явлений должны быть объяснены в физических терминах. Но при условии, что все они являются проблемами, относящимися к наблюдаемому поведению физических объектов, сводимыми к проблемам объяснения структур и функций. По этой причине этим явлениям *может быть* дано физическое объяснение, даже если в некоторых пунктах и имеются серьезные основания подозревать, что такого рода объяснения ожидать не приходится. В случае сознания заманчивая индукция на основе этих случаев потерпит неудачу, поскольку сознание не является проблемой физических структур и функций. Проблема сознания загадочна совершенно в другом смысле. Ее анализ показывает, что сознательное переживание не полностью поддается редуктивному объяснению.

## 6. Нередуктивное объяснение

В этой точке некоторые отказываются от поиска решения проблемы, придерживаясь того мнения, что мы никогда не построим теорию сознательного переживания. МакГинн(1989), например, утверждает, что проблема слишком трудна для наших ограниченных умов; мы "познавательнo замкнуты" относительно этого явления. Другие утверждают, что сознание вообще находится вне науки.

Я думаю, что этот пессимизм преждевремен. Отказываться рано: именно здесь и становится интересно. Когда исключены простые методы объяснения, мы должны исследовать альтернативы. Учитывая, что редуктивное объяснение терпит неудачу, естественно обратиться к *нередуктивному* объяснению.

Хотя поразительное количество явлений оказалось полностью объяснимым в терминах объектов более простых, чем они сами, но это положение не является универсальным. В физике иногда случается, что некоторая сущность рассматривается как *фундаментальная*. Фундаментальные сущности не объясняются в терминах чего -нибудь более простого. Вместо этого, они принимаются в качестве базовых и на этом основании строится теория о том, как они связаны со всем остальным в этом мире. Например, в девятнадцатом веке оказалось, что электромагнитные процессы нельзя полностью объяснить в терминах механических процессов, к которым обращались предыдущие физические теории, так Максвелл и другие ввели электромагнитный заряд и электромагнитные силы, как новые фундаментальные компоненты физической теории. Для объяснения электромагнетизма онтология физики должна была быть расширена. Понадобились новые фундаментальные свойства и законы для удовлетворительного объяснения этого явления.

Масса, пространство-время - таковы другие качества, которые в физических теориях рассматриваются как фундаментальные. Попыток объяснить эти качества в терминах чего -то более простого не предпринималось. Но это не исключает возможность построения теории массы или пространства-времени. Существует сложная теория того, как эти качества взаимосвязаны и в какие фундаментальные законы они входят. Эти основные принципы

используются для объяснения многих известных явлений, связанных с массой, пространством и времени на более высоком уровне.

Я предполагаю, что теория сознания должна рассматривать опыт как фундаментальный. Мы знаем, что теория сознания требует дополнения к нашей онтологии *чего-то* фундаментального, поскольку все в физической теории совместимо с отсутствием сознания. Мы могли бы добавить некоторое совершенно новое нефизическое качество, из которого можно получить переживание, но трудно понять, что чем оно могло бы быть. Предпочтительнее взять само переживание, как фундаментальное качество этого мира по аналогии с массой, зарядом и пространством-временем. Если мы рассматриваем переживание как фундаментальное, то мы можем заняться конструированием теории переживания.

Там, где имеется это фундаментальное качество, имеются и фундаментальные законы. Нередуктивная теория переживания добавит новые принципы к перечню основных законов природы. Эти основные принципы будут в конечном счете нести объяснительное бремя в теории сознания. Также, как мы объясняем известные явления высокого уровня, связанные с массой, в терминах более фундаментальных принципов, включающих массу и другие сущности, мы могли бы объяснять известные явления, связанные с переживанием, в терминах более фундаментальных принципов, включающих переживание и другие сущности.

В частности нередуктивная теория опыта определит базовые принципы, говорящие нам о том, как опыт зависит от физических качеств мира. Эти *психофизические* принципы не будут противоречить физическим законам, поскольку, по-видимому, физические законы уже формируют замкнутую систему. Скорее они будут дополнять физическую теорию. Физическая теория дает теорию физических процессов, а психофизическая теория сообщает нам, как эти процессы вызывают переживание. Мы знаем, что опыт зависит от физических процессов, но мы также знаем, что эта зависимость не может быть получена только из физических законов. Новые базовые принципы, постулируемые нередуктивной теорией, дают нам дополнительный ингредиент, который нам необходим для построения объяснительного моста.

Конечно, рассматривая переживание как фундаментальное качество, мы не имеем возможности сказать, почему оно есть. Но это относится к любой фундаментальной теории. Ничто в физике не говорит нам, почему есть материя, но мы не рассматриваем это, как аргумент против теорий материи. Определенные качества этого мира принимаются некоторой научной теорией как фундаментальные. Теория материи может объяснить все виды фактов о материи, показывая, как они следуют из фундаментальных законов. То же самое относится и к теории переживания.

Эта позиция квалифицируется как разновидность дуализма, т.к. она постулирует фундаментальные свойства сверх физических свойств. Но это - невинная разновидность дуализма, полностью совместимая с научным представлением о мире. Ничто в этом подходе не противоречит физической теории; нам просто нужно добавить *соединяющие* принципы для объяснения того, как переживание возникает из физических процессов. В этой теории нет ничего особенно духовного или мистического - по форме она подобна физической теории с несколькими фундаментальными сущностями, связанными в соответствии с фундаментальными законами. Что и говорить, это слегка расширяет онтологию, но Максвелл сделал то же самое. Действительно, общая структура этой позиции полностью натуралистична и позволяет, в конечном счете, свести вселенную к сети базовых сущностей, подчиняющихся простым законам, она в конечном счете допускает существование теории сознания, заданную в терминах таких законов. Если эту позицию нужно как-то обозначить, то "*натуралистический дуализм*" мог бы быть подходящим названием.

Если эта точка зрения верна, то в некоторых отношениях теория сознания будет больше похожа на теорию в физике, чем на теорию в биологии. Биологические теории не включают какие-либо принципы, которые являются фундаментальными в этом смысле, поэтому

биологическая теория сложна и запутанна, но физические теории в той мере, в какой они имеют дело с фундаментальными принципами, стремятся к простоте и элегантности. Фундаментальные законы природы - это часть фундаментального инвентаря этого мира, а физические теории говорят нам о том, что этот фундаментальный инвентарь замечательно прост. Если теория сознания также включает фундаментальные принципы, то мы должны ожидать от нее того же самого. Принципы простоты, элегантности, и даже красоты, которыми руководствуются физики в поиске фундаментальной теории, также применимы и к теории сознания.

(Техническое примечание: Некоторые философы утверждают, что хотя имеется *концептуальный* разрыв между физическими процессами и переживанием, но при этом не обязательно имеется метафизический разрыв, так что переживание в некотором смысле может быть физическим (например, Levine 1983; Loar 1990; Byrne 1993; Papineau 1994; Sturgeon 1994). Обычно эта линия аргументации усиливается обращением к понятию *апостериорной* необходимости (Kripke 1980). Я думаю, что эта позиция основана на неверном понимании *апостериорной* необходимости, или требует совершенно нового вида необходимости, в который у нас нет оснований верить; детали см. Chalmers 1995 (также Jackson 1994 и Lewis 1994). В любом случае, эта позиция все же признает наличие *объяснительного* разрыва между физическими процессами и переживанием. Например, принципы, соединяющие физическое и переживаемое, не выводимы из законов физики, поэтому такие принципы должны быть приняты как *объяснительно* фундаментальные. Поэтому даже с этой точки зрения объяснительная структура теории сознания будет подобна той, которую я описал.)

## 7. К теории сознания

Сейчас не слишком рано начинать работу над теорией. Мы уже в состоянии понять некоторые ключевые факты о взаимосвязях между физическими процессами и переживанием, а также относительно регулярностей, которые их соединяют. Как только редуктивное объяснение отложено в сторону, мы можем выложить эти факты на стол, чтобы они могли играть свою роль, как исходные данные в нередуктивной теории сознания и как ограничения на основные законы, которые составят окончательную теорию.

Имеется очевидная проблема, которая затрудняет развитие теории сознания: недостаточность объективных данных. Сознательное переживание не является непосредственно наблюдаемым в экспериментальном контексте, поэтому мы не можем по нашему желанию получать данные о взаимосвязи между физическими процессами и переживанием. Однако, все мы имеем доступ к богатому источнику данных в нашем собственном случае. Можно получить много важных регулярностей, связывающих переживание и обработку информации, исследуя наш собственный опыт. Имеются также хорошие косвенные источники данных, когда можно положиться на вербальное описание субъекта, как на показатель переживания. Эти методы имеют свои ограничения, но таким образом мы имеем более, чем достаточно данных для построения теории.

Философский анализ также полезен в плане извлечения денег из имеющихся в нашем распоряжении данных. Этот вид анализа может дать нам некоторое количество принципов, связывающих сознание и познание, сильно ограничивая, таким образом, вид окончательной теории. Метод мысленного экспериментирования, как мы увидим, может также принести существенные плоды. Наконец, тот факт, что мы ищем *фундаментальную* теорию означает, что при ее развитии мы можем обращаться к таким неэмпирическим ограничениям как простота, однородность, и т.п.. Мы должны стремиться к систематизации имеющейся в нашем распоряжении информации, к ее пополнению насколько это возможно посредством осторожного анализа, а затем к получению максимально простой возможной теории, которая объяснит эти данные, оставаясь при этом вероятным кандидатом на часть фундаментального инвентаря мира.

Такие теории в силу невозможности решающих intersubъектных экспериментальных тестов всегда будут сохранять элемент спекуляции, которого нет в других научных теориях. Тем не менее, мы, несомненно, можем строить теории, совместимые с имеющимися в нашем распоряжении данными, и сравнивать их друг с другом. Даже в отсутствии междисциплинарного наблюдения для оценки таких теорий имеются многочисленные критерии: простота, внутренняя согласованность, согласованность с теориями в других областях, способность воспроизводить известные нам свойства переживания, и даже в целом соответствовать здравому смыслу. Возможно, что даже после выполнения всех этих ограничений останутся значительные неопределенности, но по крайней мере мы можем развивать приемлемые возможности. Только, когда возможные теории будут развиты, мы сможем их оценить.

Нередуктивная теория сознания будет состоять из некоторого числа *психофизических принципов*, соединяющих свойства физических процессов со свойствами переживания. Эти принципы можно рассматривать, как способ, которым переживание возникает из физического. В конечном счете, эти принципы должны сообщить нам, какой тип физических систем будет ассоциироваться с переживанием, и для этого типа систем они должны сообщить нам, какой вид физических свойств обуславливает возникновение переживания, а также какой вид переживания мы должны ожидать для данной физической системы. Это трудная задача, но не видно причин, по которым мы не могли бы начать.

В нижеследующем изложении я представляю моих кандидатов на психофизические принципы, которые могли бы войти в теорию сознания. Первые два из них - это *небазовые принципы*, они суть систематические связи между обработкой информации и переживанием на относительно высоком уровне. Эти принципы могут играть существенную роль в развитии и ограничении теории сознания, но они не соотносятся с достаточно фундаментальным уровнем, чтобы считаться базовыми законами. Последнее предложение - это мой кандидат на *основной принцип*, который мог бы служить краеугольным камнем фундаментальной теории сознания. Он особенно умозрителен, но это такой вид умозрения, который необходим, если мы вообще хотим получить удовлетворительную теорию сознания. Я могу только кратко представить здесь эти принципы; более подробно я обосновываю их в Chalmers (1995).

**1. Принцип структурной согласованности.** Это принцип согласованности между *структурой сознания* и *структурой осознания*. Вспомним, что термин "осознание" использовался ранее для обозначения различных функциональных явлений, ассоциированных с сознанием. Теперь я использую его для обозначения несколько более специфического процесса в когнитивном основании переживания. А именно, под содержаниями осознания понимаются те информационные содержания, которые доступны центральным системам и которые осуществляют контроль за поведением типичным способом. Короче говоря, об осознании можно думать как о *непосредственной доступности для глобального контроля*. В первом приближении содержания осознания - это те содержания, которые являются непосредственно доступными и потенциально поддающимися описанию, по крайней мере, в системе, которая использует язык.

Осознание - это исключительно функциональное понятие, но, тем не менее, оно глубоко связано с сознательным переживанием. В известных нам случаях везде, где мы находим сознание, мы находим осознание. Везде, где есть сознательное переживание, в когнитивной системе имеется соответствующая информация, которая доступна для контроля поведения и вербального описания. И наоборот, по-видимому, всякий раз, когда информация доступна для описания и глобального контроля, то имеется сознательное переживание. Таким образом, существует прямое соответствие между сознанием и осознанием.

Это соответствие может быть развито далее. Центральным фактом относительно переживания состоит в том, что оно имеет сложную структуру. Визуальное поле, например, имеет сложную геометрию. Имеются также отношения подобия и различия между различными переживаниями, их относительными интенсивностями. По крайней мере частично переживание каждого субъекта можно охарактеризовать и проанализировать в

терминах этих структурных свойств: сходство и различие отношений, воспринятая область, относительная интенсивность, геометрическая структура и так далее. Существенный факт также состоит в том, что у каждого из этих структурных свойств имеется соответствующее свойство в структуре обработки информации при осознании.

Возьмем в качестве примера цветовые ощущения. Для каждого различия в переживании цвета существует соответствующее различие в обработке информации. Переживаемые нами различные эмпирические цвета образуют сложное трехмерное пространство, изменяющееся по оттенку, насыщенности и интенсивности. Свойства этого пространства могут быть получены из рассмотрения процесса обработки информации: исследование зрительных систем показывает, что волновые формы света различаются и анализируются по трем различным осям, и именно эта трехмерная информация существенна для последующей обработки. Трехмерная структура феноменального цветового пространства поэтому непосредственно соответствует трехмерной структуре зрительного осознания. Это именно то, что мы ожидаем. Кроме этого каждое цветовое различие соответствует некоторой поддающейся описанию информации, и поэтому некоторому различию, представленному в структуре обработки информации.

Или более прямо: геометрическая структура зрительного поля непосредственно отражена в структуре, которая может быть получена из обработки зрительной информации. Каждое геометрическое отношение соответствует чему-то поддающемуся описанию и поэтому является когнитивно представимым. Если бы в нашем распоряжении был только результат обработки информации в зрительной и когнитивной системе некоторого субъекта, то мы не могли бы *непосредственно* наблюдать его зрительный опыт, но тем не менее мы могли бы делать выводы о структурных свойствах этого опыта.

В целом, любая сознательно пережитая информация будет также когнитивно представлена. Тонкая структура зрительного поля будет соответствовать некоторой тонкой структуре в обработке зрительной информации. То же самое относится к переживанию в других модальностях, и даже к несенсорному опыту. Внутренние ментальные образы имеют геометрические свойства, которые представлены в обработке информации. Даже эмоции имеют структурные свойства, типа относительной интенсивности, которые непосредственно соответствуют некоторой структурной особенности в обработке информации; там, где имеется большая интенсивность, обнаруживается большее влияние на последующие процессы. В общем именно потому, что структурные свойства переживания доступны и могут быть вербально описаны, они будут непосредственно представлены в структуре осознания.

Именно этот изоморфизм между структурами сознания и осознания и составляет принцип структурной согласованности. Этот принцип отражает центральный факт, состоящий в том, что даже при том, что когнитивные процессы концептуально не влекут за собой факты, касающиеся сознательного переживания, но тем не менее сознание и осознание не плавают независимо друг от друга, а очень тесно согласованы.

Этот принцип имеет свои ограничения. Он позволяет нам восстановить структурные свойства переживания из свойств обработки информации, но не все свойства переживания являются структурными. Имеются свойства, типа внутренней природы ощущения красного, которые полностью не поддаются структурному описанию. Сам факт познаваемости обратных цветов, когда переживания красного и зеленого инвертированы, но структурно идентичны, показывает, что структурные свойства ограничивают переживание, но не исчерпывают его. Тем не менее, наше напряжение при попытке оставить неизменными структурные свойства инвертированных, но функционально идентичных систем, показывает, насколько важен принцип структурной согласованности для концепции нашей ментальной жизни. Он не является *логически* необходимым принципом, поскольку, в конце концов, мы можем представить себе, что вся обработка информации происходит вообще без переживания, но, тем не менее, это сильное и известное ограничение на психофизическую взаимосвязь.

Принцип структурного соответствия позволяет осуществить очень полезный тип косвенного объяснения переживания в терминах физических процессов. Например, мы можем использовать факты, касающиеся нейронной обработки визуальной информации, для косвенного объяснения структуры цветового пространства. Факты, касающиеся нейронной обработки, могут влечь за собой и объяснять структуру осознания; если мы принимаем принцип согласованности, то структура переживания также будет объяснена. Эмпирическое исследование могло бы даже дать нам лучшее понимание структуры осознания летучей мыши, проливая косвенный свет на раздражающий вопрос Нагеля, чему подобен опыт летучей мыши. Этот принцип обеспечивает естественную интерпретацию большей части проделанной работы по объяснению сознания (например, Clark 1992 и Hardin 1992 по цветам, и Akins 1993 по летучим мышам), хотя обращаются к нему неявно. Он настолько привычен, что практически всеми считается само собой разумеющимся, и является центральным пунктом в когнитивном объяснении сознания.

Соответствие между сознанием и осознанием также позволяет осуществлять естественную интерпретацию исследований в нейрологии, направленных на идентификацию *субстрата* (или *нейронного коррелята*) сознания. Были выдвинуты различные конкретные гипотезы. Например, Крик и Кох (1990) предполагают, что колебания с частотой 40 гц могут быть нейронными коррелятами сознания, в том время как Либет (1993) предполагает, что основной в этом вопросе является продолжительная по времени нейронная активность. Если мы примем принцип соответствия, то наиболее непосредственным физическим коррелятом сознания является осознание: процесс, посредством которого информация становится непосредственно доступной для глобального контроля. Различные конкретные гипотезы могут быть проинтерпретированы, как эмпирические предположения о том, как может быть достигнуто осознание. Например, Крик и Кох предполагают, что колебания с частотой 40 гц - это ворота, через которые информация интегрируется в рабочей памяти и таким образом становится доступной для дальнейшей обработки. Аналогично этому, естественно предположить, что продолжительная по времени активность Либета существенна именно потому, что только такой вид активности становится глобально доступным. Тот же самое применимо к другим предлагаемым коррелятам типа "глобального рабочего пространства" Баарса (1988), "высококачественных репрезентаций" Farah (1994), и "селективного доступа к системам действия" Shallice (1972). Все они могут рассматриваться, как гипотезы о *механизмах осознания*, механизмах, которые делают информацию, непосредственно доступной для глобального контроля.

Из согласованности сознания и осознания следует, что механизм осознания будет сам являться коррелятом сознательного переживания. Вопрос только в том, *какие* механизмы в мозге, управляющие глобальной доступностью, являются эмпирическими; возможно, что имеется несколько таких механизмов. Но если мы принимаем принцип согласованности, то у нас есть основания предполагать, что те процессы, которые *объясняют* осознание, будут в то же самое время составлять часть *основания* сознания.

**2. Принцип организационной инвариантности.** Этот принцип утверждает, что любые две системы с одной и той же тонкой *функциональной организацией* будут обладать качественно идентичным опытом. Если бы причинные паттерны нейронной организации были дублированы, например, в кремнии, так, что каждому нейрону соответствовал бы кремниевый элемент с тем же самым паттерном взаимодействий, то это привело бы к возникновению идентичных переживаний. Согласно этому принципу, не конкретность физического носителя системы существенна для возникновения опыта, но абстрактный паттерн причинного взаимодействия между ее компонентами. Этот принцип, конечно, неоспорен. Некоторые исследователи (например, Searle 1980) полагают, что сознание связано с определенной биологией, так что кремниевый дубликат человека не должен быть сознательным. Я, однако, полагаю, что анализ следующего мысленного эксперимента может оказать существенную поддержку этому принципу.

Очень кратко: предположим (с целью *сведения к абсурду*), что этот принцип ложен, и

что имеется две функционально изоморфные системы с различными переживаниями. Возможно, что только одна из систем сознательна или обе, но их переживания различны. Для наглядности давайте предположим, что одна система сделана из нейронов, а другая - из кремния, и что одна переживает красный цвет тогда, когда другая - синий. Эти две системы имеют идентичную организацию, поэтому мы можем представить себе постепенное преобразование одной в другую, возможно, заменой каждого нейрона кремниевыми чипами с идентичной локальной функцией. Мы, таким образом, получаем спектр промежуточных состояний, каждое из которых имеет ту же самую организацию, но со слегка различной физической основой и слегка различными переживаниями. В этом спектре должны быть две системы, *A* и *B*, которые различаются меньше, чем на десятую часть, но при этом их переживания различны. Эти две системы физически идентичны за исключением того, что небольшая нейронная цепь *A* заменена на кремниевую цепь *B*.

Ключевой шаг в этом мысленном эксперименте состоит в том, чтобы взять нейронную цепь в *A* и установить ее вместе с изоморфной кремниевой цепью, с переключателем между ними. Что произойдет, когда мы щелкнем переключателем? В соответствии с гипотезой, сознательное переживание системы изменится, скажем, от красного до синего. Это следует из того факта, что после этой замены система является по существу версией *B*, в то время как до нее она была *A*.

Но ведь предполагалось, что система не может *заметить* изменений! Ее причинная организация не изменилась, поэтому все ее функциональные состояния и поведенческие предрасположенности не изменились. С системой не произошло ничего необычного. Нет оснований для такой мысли: "Хмм! Только что произошло что-то странное!". В общем, структура любой такой мысли должна отражаться в обработке информации, но здесь структура обработки остается неизменной. Если бы появилась такая мысль, то она бы существовала полностью независимо от системы и была бы совершенно неспособна повлиять на дальнейшую обработку информации. (Если бы она повлияла на последующую обработку информации, то вопреки гипотезе эти системы были бы функционально различны.) Мы можем даже пощелкать выключателем несколько раз так, чтобы красный и синий цвета танцевали бы перед "внутренним оком" системы. Согласно гипотезе, система никогда не сможет заметить это "танцующее qualia".

Это рассуждение я рассматриваю как *сведение к абсурду* первоначального предположения. Это центральный факт относительно переживания, хорошо знакомый нам из нашего собственного опыта: всякий раз, когда переживания значительно изменяются, и мы обращаем на это внимание, то мы в состоянии заметить это изменение; если бы это не имело место, то мы пришли бы к сомнительной возможности, что наши переживания все время танцуют перед глазами. Эта гипотеза имеет тот же самый статус, как и возможность того, что мир был создан пять минут назад: возможно, что это логически непротиворечиво, но это неприемлемо. Исходя из очень вероятного предположения, что изменения в переживании соответствуют изменениям в обработке информации, мы приходим к заключению, что первоначальная гипотеза является невозможной, и что у любых двух функционально изоморфных систем должен быть идентичный опыт. В технических терминах это можно сформулировать так: хотя логически философские гипотезы "отсутствия qualia" и "инверсии qualia" возможны, но эмпирически и номологически они невозможны.

(Кого-то может беспокоить то, что кремниевый дубликат нервной системы невозможен по техническим причинам. Вопрос открыт. Принцип инвариантности утверждает только, что *если* изоморфизм возможен, то тип сознательного опыта изоморфных систем будет идентичным.)

Здесь можно сказать и больше, но изложенного достаточно для создания основного аромата. Еще раз: этот мысленный эксперимент позволяет на основе известных фактов о соответствии между сознанием и когнитивной обработкой информации прийти к сильному выводу об отношении между физической структурой и переживанием. Если этот аргумент проходит, то единственные физические свойства, непосредственно имеющие отношение к



появлению переживания, - это *организационные* свойства. Это утверждение представляет собой дальнейшее сильное ограничение на теорию сознания.

**3. Теория двойного аспекта информации.** Два предшествующих принципа *не были основополагающими*. Они формулируются с помощью понятий высокого уровня типа "осознания" и "организации", и поэтому не прилежат уровню, конституирующему фундаментальные законы в теории сознания. Тем не менее, они действуют как сильные ограничения. Далее необходимы *базовые* принципы, удовлетворяющие этим ограничениям и в конечном счете способные объяснить их.

Основной принцип, который я предлагаю, по существу использует понятие *информации*. Я понимаю информацию более или менее в смысле Шеннона(1948). Там, где имеется информация, существуют *информационные состояния*, вложенные в *информационное пространство*. Информационное пространство имеет базовую структуру отношений *различия* между его элементами, характеризующую те отношения, в которых разные элементы пространства являются сходными или различными, возможно эти отношения сложны. Информационное пространство - это абстрактный объект, но, следуя Шеннону, мы можем рассматривать информацию как *физически воплощенную*, когда имеется пространство физических состояний, различия между которыми могут быть переданы некоторым причинным способом. Переданные состояния можно рассматривать как образующие информационное пространства. По выражению Бейтсона(1972), физическая информация - *это различие, которое имеет значение*.

Принцип двойного аспекта происходит из наблюдения, состоящего в том, что существует прямой изоморфизм между некоторыми физически воплощенными информационными пространствами и некоторыми феноменальными (или *переживаемыми*) информационными пространствами. На основании того же типа наблюдений, которые вошли в принцип структурного соответствия, мы можем заметить, что различия между феноменальными состояниями имеют структуру, непосредственно соответствующую различиям, которые воплощены в физических процессах, в частности, тем различиям, которые существенны для способов реализации глобальной доступности и контроле. А именно, мы обнаруживаем *одно то же* абстрактное информационное пространство, воплощенное в физической обработке и в сознательном переживании.

Это приводит к естественной гипотезе о том, что информация (или по крайней мере некоторая информация) имеет два основных аспекта, физический аспект и феноменальный аспект. Это утверждение имеет статус основного принципа, который мог бы стать основополагающим и объяснить возникновение переживания из физического. Переживание возникает благодаря своему статусу одного аспекта информации, когда другой аспект оказывается воплощенным в физической обработке.

Этот принцип основывается на нескольких соображениях, которые здесь я могу только кратко обозначить. Во-первых, соображение о наличии такого типа физических изменений, которые соответствуют изменениям в сознательном переживании, позволяет предположить, что эти изменения значимы благодаря их роли в образовании *информационных изменений*, т.е. различий в абстрактном пространстве состояний, которые разделены в точном соответствии с их каузальными различиями вдоль некоторых каузальных путей. Во-вторых, если должен выполняться принцип организационной инвариантности, то нам нужно найти некоторое фундаментальное организационное свойство переживания, к которому нужно будет присоединять, а информация - это свойство организации *par excellence*. В-третьих, этот принцип дает некоторую надежду на объяснение принципа структурного соответствия в терминах структуры, существующей в информационных пространствах. В-четвертых, анализ когнитивного объяснения наших *суждений* и *утверждений* о сознательном переживании - суждений, являющихся функционально объяснимыми, но однако глубоко связанными с сознательным переживанием - позволяет предположить, что объяснение по существу включает в себя информационные состояния, воплощенные в когнитивной обработке информации. Из этого следует, что теория, основанная на информации, допускает глубокое

соответствие между объяснением опыта и объяснения наших суждений и требований относительно него.

Уиллер(1990) предположил, что информация фундаментальна для физики вселенной. Согласно этой доктрине «из бита», законы физики могут быть описаны в терминах информации посредством постулирования различных состояний, которые вызывают различные последствия без фактического описания того, чем эти состояния *являются*. Существенно только их положение в информационном пространстве. Если это так, то информация также является естественным кандидатом для фундаментальной теории сознания. Мы приходим к концепции мира, в котором информация является действительно фундаментальной и имеет два основных аспекта, один, соответствующий физическим, а другой феноменальным свойствам мира.

Конечно же, принцип двойного аспекта чрезвычайно спекулятивен, а также неопределен, т.к. ряд ключевых вопросов остаются без ответа. Очевидный вопрос: *вся* ли информация имеет феноменальный аспект. Одна возможность состоит в том, что нам нужно дополнительное ограничение на фундаментальную теорию, указывающее, какой *вид* информации имеет феноменальный аспект. Другая возможность состоит в том, что никакого такого ограничения нет. Если нет, то переживание распространено гораздо более широко, чем мы, возможно, предполагаем, т.к. информация - всюду. Сначала это кажется противоречащим интуиции, но дальнейшее размышление, как мне кажется, придает этому утверждению определенное правдоподобие и элегантность. Там, где присутствует простая обработка информации, там есть простое переживание, а там, где сложная - там сложный опыт. Мышь имеет более простую структуру обработки информации, чем человек, и, соответственно, более простой опыт; возможно ли, что термостат, максимально простая структура обработки информации, имеет максимально простое переживание? В самом деле, если переживание - это действительно фундаментальное свойство, то было бы удивительно если бы оно возникало только время от времени; наиболее фундаментальные свойства более равномерно распространены. В любом случае, этот вопрос совершенно открыт, но я полагаю, что эта позиция не столь неправдоподобна, как о ней часто думают.

Как только фундаментальная связь между информацией и переживанием выложена на стол, то открывается дверь для более значительных метафизических спекуляций о природе мира. Например, часто отмечается, что физика характеризует свои основные объекты только *внешне*, в терминах их отношений к другим объектам, которые в свою очередь характеризуются извне и так далее. Внутренняя природа физических объектов остается вне рассмотрения. Некоторые утверждают, что нет никаких внутренних свойств, но тогда мы остаемся в мире, который является чистым причинным потоком (чистым потоком информации) без свойств для связи. Если же допускается существование внутренних свойств, то становится возможным приведенное выше естественное умозрительное предположение о том, что внутренние свойства физических объектов, т.е. те свойства, к которым возникновение сознания в конечном счете имеет отношение, сами являются феноменальными свойствами. Можно было бы сказать, что феноменальные свойства - это внутренний аспект информации. Это может дать ответ на вопрос о причинной релевантности переживания - естественное беспокойство ввиду такой картины мира, в которой физическая область является причинно замкнутой, а переживание дополнительно к физическому. Информационная точка зрения позволяет нам понять, как переживание в силу своего статуса внутренней природы физического может являться тонким видом причинности. Эту метафизическую спекуляцию, по-видимому, лучше игнорировать с целью развития научной теории, но при обсуждении некоторых философских проблем она является стимулирующей.

## 8 Заключение

Теория, которую я представил, умозрительна, но это - только некоторая предлагаемая версия теории. Я подозреваю, что принципы структурной согласованности и

организационной инвариантности будут существенными пунктами в любой удовлетворительной теории сознания; статус теории двойного аспекта информации менее определен. В самом деле, сейчас это скорее идея, чем теория. Для того, чтобы иметь некоторую надежду на успех в объяснении, она должна быть определена более полно и обрести более сильную форму. Однако, размышление над тем, что правдоподобно и что неправдоподобно, где теория работает и где нет, может только привести к улучшению этой теории.

Большинство существующих теорий сознания или отрицают это явление, объясняя что-то другое, или оценивают эту проблему, как вечную тайну. Я надеюсь, что мне удалось показать возможность прогресса в этой проблеме, даже принимая ее всерьез. Для дальнейшего продвижения необходимо дальнейшее исследование, более разработанные теории, более осторожный анализ. Трудная проблема - это трудная проблема, но нет никаких причин полагать, что она всегда останется нерешенной.<sup>2</sup>

## Литература

- Akins, K. 1993. What is it like to be boring and myopic? In (B. Dahlbom, ed.) *Dennett and his Critics*. Blackwell.
- Allport, A. 1988. What concept of consciousness? In (A. Marcel and E. Bisiach, eds.) *Consciousness in Contemporary Science*. Oxford University Press.
- Baars, B.J. 1988. *A Cognitive Theory of Consciousness*. Cambridge University Press.
- Bateson, G. 1972. *Steps to an Ecology of Mind*. Chandler Publishing.
- Byrne, A. 1993. *The Emergent Mind*. Ph.D. Dissertation, Princeton University.
- Chalmers, D.J. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press.
- Clark, A. 1992. *Sensory Qualities*. Clarendon.
- Crick, F. and Koch, C. 1990. Toward a neurobiological theory of consciousness. *Seminars in the Neurosciences* 2:263-275.
- Crick, F. 1994. *The Astonishing Hypothesis: The Scientific Search for the Soul*. Scribners.
- Dennett, D.C. 1991. *Consciousness Explained*. Little-Brown.
- Edelman, G. 1989. *The Remembered Present: A Biological Theory of Consciousness*. Basic Books.
- Farah, M.J. 1994. Visual perception and visual awareness after brain damage: A tutorial overview. In (C. Umiltà and M. Moscovitch, eds.) *Consciousness and Unconscious Information Processing: Attention and Performance 15*. MIT Press.
- Flohr, H. 1992. Qualia and brain processes. In (A. Beckermann, H. Flohr, and J. Kim, eds.) *Emergence or Reduction?: Prospects for Nonreductive Physicalism*. De Gruyter.
- Hameroff, S.R. 1994. Quantum coherence in microtubules: A neural basis for emergent consciousness? *Journal of Consciousness Studies* 1:91-118.
- Hardin, C.L. 1992. Physiology, phenomenology, and Spinoza's true colors. In (A. Beckermann, H. Flohr, and J. Kim, eds.) *Emergence or Reduction?: Prospects for Nonreductive Physicalism*. De Gruyter.
- Humphrey, N. 1992. *A History of the Mind*. Simon and Schuster.
- Jackendoff, R. 1987. *Consciousness and the Computational Mind*. MIT Press.
- Jackson, F. 1994. Finding the mind in the natural world. In (R. Casati, B. Smith, and S. White, eds.) *Philosophy and the Cognitive Sciences*.
- Kripke, S. 1980. *Naming and Necessity*. Harvard University Press.
- Levine, J. 1983. Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly* 64:354-61.
- Lewis, D. 1994. Reduction of mind. In (S. Guttenplan, ed.) *A Companion to the Philosophy of Mind*. Blackwell.
- Libet, B. 1993. The neural time factor in conscious and unconscious events. In (G.R. Block and J. Marsh, eds.) *Experimental and Theoretical Studies of Consciousness (Ciba Foundation Symposium 174)*. John Wiley and Sons.
- Loar, B. 1990. Phenomenal states. *Philosophical Perspectives* 4:81-108.
- McGinn, C. 1989. Can we solve the mind-body problem? *Mind* 98:349-66.
- Nagel, T. 1974. What is it like to be a bat? *Philosophical Review* 4:435-50.
- Newell, A. 1990. *Unified Theories of Cognition*. Harvard University Press.
- Papineau, D. 1993. Physicalism, consciousness, and the antipathetic fallacy. *Australasian Journal of Philosophy* 71:169-83.
- Penrose, R. 1989. *The Emperor's New Mind*. Oxford University Press.
- Penrose, R. 1994. *Shadows of the Mind*. Oxford University Press.
- Searle, J.R. 1980. Minds, brains and programs. *Behavioral and Brain Sciences* 3:417-57.

---

<sup>2</sup> Аргументы, приведенные в этой работе, более глубоко представлены в моей книге *"Сознательный ум"*. Я благодарен Френсису Крику, Peggy DesAutels, Matthew Elton, Liane Gabora, Кристоферу Коху, Paul Rhodes, Gregg Rosenberg, and Sharon Wahl за их комментарии

- Shallice, T. 1972. Dual functions of consciousness. *Psychological Review* 79:383-93.
- Shannon, C.E. 1948. A mathematical theory of communication. *Bell Systems Technical Journal* 27: 379-423.
- Sturgeon, S. 1994. The epistemic basis of subjectivity. *Journal of Philosophy* 91:221-35.
- Wheeler, J.A. 1990. Information, physics, quantum: The search for links. In (W. Zurek, ed.) *Complexity, Entropy, and the Physics of Information*. Addison-Wesley.
- Wilkes, K.V. 1988. - , Yishi, Duh, Um and consciousness. In (A. Marcel and E. Bisiach, eds.) *Consciousness in Contemporary Science*. Oxford University Press.