

ФИЛОСОФИЯ СОЗНАНИЯ

David J. Chalmers

The Conscious Mind

In Search of a Fundamental Theory

Дэвид ЧАЛМЕРС

Сознающий ум

В поисках
фундаментальной
теории



URSS

David J. Chalmers

THE CONSCIOUS MIND
In Search of a Fundamental Theory

Д. Чалмерс

СОЗНАЮЩИЙ УМ

*В поисках
фундаментальной теории*

Перевод с английского
доктора философских наук, профессора
В. В. Васильева



МОСКВА

Чалмерс Дэвид

Сознающий ум: В поисках фундаментальной теории. Пер. с англ.

М.: УРСС: Книжный дом «ЛИБРОКОМ», 2013. — 512 с. (Философия сознания.)

Эта книга, впервые опубликованная в 1996 году, стала одним из самых заметных философских трактатов конца XX века. В наши дни уже не удастся найти серьезных работ по проблеме сознания, в которых не было бы ссылок на Чалмерса.

«Сознающий ум» — увлекательный философский рассказ о глубочайших парадоксах и тайнах сознания. Это провокативная работа, в которой сделана попытка обосновать «натуралистический дуализм», исходя из тезиса автора о нефизической природе сознания и его зависимости от функциональных схем в мозге. Чалмерс также утверждает, что его теория открывает новые перспективы для интерпретации квантовой механики и позволяет говорить о возможности сознательных роботов.

Ясность изложения, смелость идей, изобретательность мысленных экспериментов, точность рассуждений и широкая эрудиция автора делают эту книгу настоящим подарком для всех, кто интересуется философией.

Издание осуществлено при участии Московского центра исследований сознания

Редактор Д. Б. Волков

*“The Conscious Mind: In Search of a Fundamental Theory”, First Edition,
was originally published in English in 1996.*

This translation is published by arrangement with Oxford University Press.

*“The Conscious Mind: In Search of a Fundamental Theory”, 1-е изд.,
впервые опубликовано на английском языке в 1996 г.*

Перевод на русский язык публикуется по соглашению с Oxford University Press.

Издательство «Книжный дом “ЛИБРОКОМ”». 117335, Москва, Нахимовский пр-т, 56.
Формат 60×90/16. Печ. л. 32. Зак. № 22

Отпечатано в ООО «Чебоксарская типография № 1». 428019, Чебоксары, пр-т И. Яковleva, 15.

ISBN 978-5-453-00045-6

(УРСС)

ISBN 978-5-397-03778-5

(Книжный дом «ЛИБРОКОМ»)

© David J. Chalmers, 1996

© УРСС, 2013

10443 ID 157324



9 785397 037785



Содержание

Благодарности	6
Введение	9
 Часть I. Основы	 19
Глава 1. Два понятия ментального	19
1. Что такое сознание?	19
2. Феноменальное и психологическое понятие ментального	28
3. Двойная жизнь ментальных терминов	35
4. Две проблемы соотношения ментального и телесного	44
5. Два понятия сознания	46
 Глава 2. Супервентность и объяснение	 54
1. Супервентность	54
2. Редуктивное объяснение	66
3. Логическая супервентность и редуктивное объяснение	73
4. Концептуальная истина и необходимая истина*	78
5. Почти все логически супервентно на физическом*	101
 Часть II. Нередуцируемость сознания	 125
Глава 3. Можно ли редуктивно объяснить сознание?	125
1. Является ли сознание логически супервентным на физическом?	125
2. Крах редуктивного объяснения	141
3. Когнитивное моделирование	147
4. Нейробиологическое объяснение	151

5. Апелляция к новой физике	155	
6. Эволюционное объяснение	157	
7. Куда ведет редуктивное объяснение?	158	
Глава 4. Натуралистический дуализм	160	<i>Прим.</i>
1. Аргумент против материализма	160	458
2. Возражения от апостериорной необходимости*	169	
3. Другие аргументы в пользу дуализма*	180	
4. Эпифеноменализм ли это?*	193	
5. Логическая география проблем	207	
6. Размышления о натуралистическом дуализме	216	
Глава 5. Парадокс феноменального суждения	221	<i>Прим.</i>
1. Сознание и познание	221	470
2. Парадокс феноменального суждения	226	
3. Объяснение феноменальных суждений	236	
4. Аргументы против объяснительной нерелевантности	244	
5. Аргумент от самопознания*	245	
6. Аргумент от памяти*	255	
7. Аргумент от референции*	256	
8. Содержание феноменальных убеждений*	258	
Часть III. На пути к теории сознания	267	
Глава 6. Когерентность сознания и познания	267	<i>Прим.</i>
1. На пути к нередуктивной теории	267	474
2. Принципы когерентности	273	
3. Еще о понятии осведомленности	282	
4. Объяснительная роль принципов когерентности	292	
5. Когерентность как психофизический закон	303	
Глава 7. Отсутствующие квалиа, блекнувшие квалиа, скачущие квалиа	309	<i>Прим.</i>
1. Принцип организационной инвариантности	309	480
2. Отсутствующие квалиа	314	
3. Блекнувшие квалиа	316	
4. Инвертированные квалиа	328	
5. Скачущие квалиа	332	
6. Нередуктивный функционализм	342	

Глава 8. Сознание и информация: спекулятивные рассуждения	<i>Прим.</i>
	345
1. На пути к фундаментальной теории	345
2. Аспекты информации	347
3. Некоторые подкрепляющие аргументы	358
4. Повсеместен ли опыт?	365
5. Метафизика информации	376
6. Открытые вопросы	385
Часть IV. Прикладные аспекты	<i>Прим.</i>
Глава 9. Сильный искусственный интеллект	389
1. Машинаное сознание	389
2. Об имплементации вычисления	392
3. В защиту сильного ИИ	398
4. Китайская комната и другие возражения	400
5. Внешние возражения	408
6. Заключение	411
Глава 10. Интерпретация квантовой механики	<i>Прим.</i>
	485
1. Две тайны	413
2. Каркас квантовой механики	414
3. Интерпретируя квантовую механику	418
4. Интерпретация Эверетта	428
5. Возражения против интерпретации Эверетта	434
6. Заключение	440
Примечания	443
Библиография	487
Именной указатель	499
Предметный указатель	502

Благодарности

Я увлекся сознанием и проблемой отношения ментального и физического в первые годы изучения мной математики в Университете Аделаиды. Разговоры с разными людьми, особенно с Полом Бартером, Джоном Бакстером, Беном Хембли и Полом Макканом, помогли формированию моих идей. Даже тогда этот вопрос казался чуть ли не удивительнейшей из проблем. Я и не мог всерьез подумать, чтобы можно было на постоянной основе заниматься размышлениями о том, что было настолько забавным.

Позже, когда я стал аспирантом в Оксфорде, я заметил, что вместо математики все время думаю о сознании, и я решил сменить специализацию, а в конце концов — и континент. Многие люди проявляли терпение и оказывали мне поддержку в это непростое время, особенно Майкл Аттья, Майкл Даммит и Робин Флетчер. Спасибо также всем тем, кому доводилось выслушивать мои соображения о том, чему случалось быть моей последней теорией сознания; идеи этой книги — их далекий отпрыск.

Мое решение перебраться в Университет Индианы для обучения основам философии, когнитивной науки и искусственного интеллекта было одним из лучших моих решений. Особенно я благодарен Дагу Хофтадтеру; именно его тексты изначально, когда я был еще юношой, послужили для меня введением в тайны сознания, а стимулирующая и уютная атмосфера его исследовательской лаборатории — Центра исследований понятий и познания — позволила этим идеям развиться. Хотя он несогласен со многими идеями этой книги, мне хочется думать, что в каком-то плане написанное мной хранит верность интеллектуальному духу его трудов.

Я создал первый вариант этой работы (известной тогда под называнием «На пути к теории сознания») за шесть горячих месяцев 1992 и 1993 гг. В то время у меня было немало полезных дискуссий с множеством людей в Индиане: со всеми в ЦИПП, особенно с Бобом Френчем и Лиан Габора, а также со многими сотрудниками других подразделений, в том числе с Майклом Данном, Робом Голдстоуном, Анилом Гупта, Джимом Хеттмером, Джерри Селигманом и Тимом ван Гелдером.

Спасибо также участникам дискуссионной группы по сознанию в задней комнате «Никс» за многие приятные часы бесед по понедельникам после полудня.

Двухлетняя Макдоннелская стипендия по философии, нейронауке и психологии в Вашингтонском университете создала новую благотворную среду, а также дала мне шанс ощутить на себе парадокс Зенона при завершении этой книги. Я благодарен Фонду Джеймса С. Макдоннела за поддержку, всем участникам моего аспирантского семинара по сознанию за дискуссии, позволившие мне отшлифовать книгу, и многим людям за беседы и комментарии, в том числе Мортену Христиансену, Энди Кларку, Джейсону Клевенджеру, Пегги Деотель, Пепе Торибио и Теду Завидски.

За последнюю пару лет у меня было громадное множество полезных бесед и обменов письмами в связи с материалами этой книги. Среди многих других благодарности заслуживают Джон Бакстер, Нэд Блок, Алекс Бирн, Фрэнсис Крик, Дэн Деннет, Эрик Дитрих, Ави Элитцур, Мэтью Элтон, Оуэн Фланаган, Стэн Франклин, Лиан Габора, Говен Гузельдере, Крис Хилл, Терри Хорган, Стив Хорст, Фрэнк Джексон, Джегвон Ким, Кристофф Кох, Мартин Лики, Дэйв Леисинг, Керри Левенберг, Джо Левин, Дэвид Льюис, Барри Левер, Билл Лайкан, Пол Маккан, Дейрил Маккалоу, Брайан Маклафлин, Томас Метцингер, Роберт Миллер, Эндрю Милн, Джон О'Лери-Хоусон, Джозеф О'Роурке, Калвин Острум, Ретт Севидж, Арон Сломан, Леопольд Штубенберг и Рэд Уотсон. Я признателен за интересные общие разговоры о сознании такому количеству других людей, что не смогу даже их назвать. Особо хочу упомянуть НORTона Нилкина, вернувшего свой экземпляр рукописи со множеством полезных замечаний незадолго до смерти от лимфомы. Его будет не хватать.

У меня есть немало философских долгов и в более широком смысле. Мои первоначальные взгляды на сознание были по большей части моим собственным изобретением, но они существенно обогатились после прочтения соответствующей литературы. Нетрудно убедиться, что любая идея, скорее всего, уже была высказана кем-то еще. Если говорить о современных мыслителях, то Томас Нагель, Фрэнк Джексон и Джозеф Левин немало сделали для того, чтобы подчеркнуть трудности сознания; их работы покрывают во многом ту же территорию, что и мои первые главы. Мой труд также в ряде пунктов пересекается с трудами Нэда Блока, Роберта Кирка и Майкла Локвуда. Метафизическая модель, развиваемая мной во второй главе, во многом опирается на труды Терри Хоргана, Сола Крипке, Дэвида Льюиса и др., а Фрэнк Джексон независимо от меня разработал сходную модель, представленную в его замечательных Локковских лекциях 1995 г. Постоянным

стимулом и вызовом для меня были идеи Дэниела Деннета, Колина Макгинна, Джона Серла и Сидни Шумейкера.

Больше всего я обязан Грэггу Розенбергу — за памятные беседы и ценные отклики, Лизе Томас — за книгу о зомби и моральную поддержку, Шэрон Уол — за профессиональную редактуру и теплую дружбу и, превыше всего, всем трем моим родителям за их помошь и поддержку. И спасибо всем моим квалиа и среде, где они возникли, за постоянное вдохновение.

Когда я дописывал эту книгу, в ресторане мне досталось печенье с предсказаниями, где были слова: «Ваша жизнь будет полна восхитительных тайн». Пока она именно такой и была, и я очень благодарен за это.

Введение

Принимая сознание всерьез

Сознание — величайшая тайна. Это, возможно, наибольшее из громадных препятствий на нашем пути к научному пониманию мира. Физика еще не завершена, но хорошо понята; биология разгадала многие древние тайны, окружавшие природу жизни. Наше понимание в этих областях не лишено пробелов, но они не выглядят неустранимыми. Мы представляем, как можно было бы решить эти проблемы; нам надо лишь уточнить детали.

Большой прогресс был даже в науке о ментальном. Недавние исследования в когнитивной науке и нейронауке ведут нас к лучшему пониманию человеческого поведения и движущих им процессов. Конечно, мы не располагаем множеством детальных теорий познания, но они не за горами.

Сознание, однако, остается столь же загадочным, как и раньше. По-прежнему кажется совершенно таинственным то обстоятельство, что продуцирование поведения должно сопровождаться субъективной внутренней жизнью. У нас есть серьезное основание полагать, что сознание порождается такими физическими системами, как мозг, но мы плохо понимаем, как это происходит или почему оно вообще существует. Как такая физическая система, как мозг, могла бы быть еще и *субъектом опыта*? Почему *что-то* в ней должно показывать, *каково это* — быть такой системой? Современные научные теории едва ли вообще касаются действительно сложных вопросов о сознании. Дело обстоит не так, что у нас нет детальной теории; мы абсолютно не понимаем, как сознание встроено в природу.

За последние годы появилось немало книг и статей о сознании, и можно было бы подумать, что мы движемся вперед. Но при более внимательном рассмотрении оказывается, что большинство этих работ обходит стороной самые трудные проблемы сознания. Зачастую в них обсуждается то, что можно было бы назвать «легкими» проблемами сознания. Как мозг обрабатывает стимулы, идущие из окружающей среды? Как он интегрирует информацию? Как мы формируем отчеты

о внутренних состояниях? Это важные вопросы, но ответ на них не решает трудной проблемы: почему все эти процессы сопровождаются переживаемой в опыте внутренней жизнью? Иногда этот вопрос полностью игнорируется; иногда откладывается до лучших времен; а иногда попросту декларативно объявляется разрешенным. Но в каждом из этих случаев у нас остается чувство, что главная проблема столь же загадочна, как и прежде.

Эта загадочность, впрочем, не повод для отчаяния; скорее, она делает проблему сознания одним из наиболее волнующих интеллектуальных вызовов нашего времени. Поскольку сознание и столь фундаментально, и так плохо понято, решение этой проблемы может оказать глубокое воздействие на наше представление о мире и о самих себе.

Я оптимист в вопросе о сознании; я считаю, что в конце концов у нас будет теория сознания, и в этой книге я занимаюсь ее поисками. Но сознание — не рутинная вещь; если мы хотим добиться реальных результатов, то вначале мы должны поговорить о том, что делает эту проблему настолько сложной. Затем мы сможем приступить к теории, без шор и с четким представлением о том, какую задачу нам предстоит решить.

В этой книге я не решаю проблему сознания раз и навсегда, а пытаюсь взять ее под контроль. Я стараюсь прояснить существование проблем, доказываю, что стандартные методы нейронауки и когнитивной науки не годятся для их разрешения, а затем я пытаюсь сделать какие-то шаги вперед.

Разрабатывая свою концепцию сознания, я старался придерживаться нескольких ограничений. Первое и наиболее важное из них состоит в том, чтобы *принимать сознание всерьез*. Самый простой путь создания «теории» сознания — отрицать его существование или так переопределить феномен, нуждающийся в объяснении, чтобы он перестал быть самим собой. Обычно это ведет к элегантной теории, но проблема никуда не уходит. На протяжении всей этой книги я исхожу из существования сознания и из неприемлемости такого переопределения проблемы, когда все сводится к объяснению реализации определенных когнитивных или поведенческих функций. Это я и имею в виду под серьезным отношением к сознанию.

Некоторые говорят, что сознание — это «иллюзия», но я даже плохо понимаю, что это могло бы означать. Мне кажется, что наша уверенность в существовании сознательного опыта превышает нашу уверенность в существовании чего-либо еще в мире. Временами я изо всех сил пытался убедить себя в том, что в действительности тут ничего нет, что сознательный опыт пуст, есть всего лишь иллюзия. В таком представлении есть нечто облазнительное, и философы разных эпох

пытались обыгрывать его, но в итоге подобная картина совершенно неудовлетворительна. Я ощущаю оранжевое, погружен в это ощущение, и *что-то при этом происходит*. Здесь есть что-то, что нуждается в объяснении даже после того, как мы объяснили процессы различения и действия, и это что-то — *переживание*.

Конечно, я не могу *доказать*, что здесь скрывается дополнительная проблема, — именно потому, что я не могу доказать существование сознания. Мы знаем о сознании более непосредственно, чем мы знаем о чем-либо еще, так что здесь неуместно говорить о «доказательстве». В лучшем случае я могу приводить аргументы, где это возможно, и опровергать аргументы противоположной стороны. Нельзя отрицать, что на каком-то этапе это предполагает обращение к интуиции; но все аргументы так или иначе предполагают интуицию, и я пытался четко обозначать свои интуиции.

В этой связи можно было бы говорить о Великом Разделении, имеющемся в области исследований сознания. Если вы считаете, что решение «легких» проблем объясняет все, что нуждается в объяснении, то у вас получается один тип теории; если вы считаете, что кроме них есть и дополнительная «трудная» проблема, то у вас будет другой ее тип. В какой-то момент общезначимые *аргументы* становятся затруднительны, и обсуждения зачастую сводятся к ударам по столу. Мне кажется очевидным, что здесь есть некий дополнительный момент, нуждающийся в объяснении; другие же готовы принять, что ничего подобного не существует. (Неформальные опросы показывают, что сторонников первой позиции в два или три раза больше, причем эта пропорция мало меняется в зависимости от того, ученым и студентам из каких областей задается вопрос.) Не исключено, что нам надо по-просту смириться с этим базовым разделением.

Эта книга может представлять интеллектуальный интерес для тех, кто думает, что здесь нет особой проблемы, но в действительности она предназначена тем, кто нутром ощущает эту проблему. В настоящий момент мы неплохо понимаем, какого рода теория будет у нас, если мы допустим отсутствие проблемы. В этой работе я пытался исследовать последствия признания нами *существования* проблемы. Аргументация этой книги, по сути, призвана показать, что если мы всерьез принимаем сознание, то в итоге мы должны будем прийти к позиции, которую я излагаю.

Второе ограничение, которого я придерживался, состоит в том, чтобы *всерьез принимать науку*. Я не пытался оспорить имеющиеся научные теории в тех областях, где они пользуются авторитетом. В то же время я не боялся ходить по краю в тех областях, где мнения ученых столь же далеки от основательности, как и мнения всех остальных.

К примеру, я не оспаривал тезис о том, что физический мир каузально замкнут или что поведение можно объяснить в физических терминах; но если физик или когнитивный ученый считает, что сознание может быть объяснено в физических терминах, то это не более чем упование, не фундированное современными теориями, и вопрос остается открытым. В общем, я пытался делать так, чтобы мои идеи были *совместимы* с современной наукой, но при этом не ограничивался идеями, популярными у современных ученых.

Третье ограничение состоит в том, что я признаю сознание естественным феноменом, подпадающим под действие законов природы. Если это так, то должна существовать *какая-то* корректная теория сознания вне зависимости от того, сможем ли мы найти ее. Трудно оспаривать, что сознание — естественный феномен: это крайне широко распространенный компонент природы, повсеместно возникающий у людей и, очень вероятно, у многих других биологических видов. И у нас есть все основания считать, что естественные феномены подпадают под фундаментальные законы природы; было бы очень странно, если бы сознание было исключением. Но это не значит, что естественные законы, связанные с сознанием, ничем не отличаются от законов в других областях, и тем более не значит, что они будут *физическими* законами. Они могут иметь совершенно иной вид.

Проблема сознания — беспокойная пограничная проблема, находящаяся на стыке науки и философии. Я бы сказал, что, по сути, это научный вопрос: речь идет о естественном феномене вроде движения, жизни, познания, требующем объяснения в том же ключе, что и они. Но его нельзя исследовать обычными научными методами. Привычная научная методология испытывает проблемы при работе с ним — не в последнюю очередь из-за трудностей, сопряженных с наблюдением этого феномена. Если отвлечься от случая с самим собой, то данные о нем получить непросто. Это не значит, что внешние данные не могут быть использованы, но перед тем, как мы сможем обосновать их релевантность, мы должны прийти к когерентной философской трактовке. Так что проблема сознания может оказаться такой научной проблемой, которая требует использования философских методов истолкования, предшествующего ее рассмотрению.

Выводы, к которым я прихожу в этой книге, кто-то может счесть «ненаучными»: я доказываю невозможность редуктивного объяснения сознания и даже привожу аргументы в пользу некоего дуализма. Но это всего лишь часть научного процесса. Некоторые виды объяснения, как выясняется, не работают, и поэтому мы должны использовать вместо них другие. Все, что я говорю здесь, совместимо с результатами современной науки; наша картина природного мира расширяется, а не отбрасывается. И это расширение открывает возможность для создания

натуралистической теории сознания, которая могла бы оказаться невозможной без него. Ненаучным, мне кажется, было бы *игнорировать* проблемы сознания; прямое же обращение к ним соответствует научному духу. Того, кто считает, что наука подразумевает материализм, я прошу подождать и посмотреть.

Должен заметить, что выводы этой книги — это именно *выводы*, в самом сильном смысле слова. По своему темпераменту я очень расположен к материалистическому редуктивному объяснению, и у меня нет выраженных спиритуалистических или религиозных наклонностей. Многие годы я питал надежду на материалистическую теорию; и я лишь с большой неохотой оставил ее. В конце концов мне стало ясно, что этих выводов нельзя избежать никому из тех, кто хочет всерьез принимать сознание. Материализм — это прекрасная и захватывающая картина мира, но для объяснения сознания мы должны использовать дополнительные ресурсы.

В настоящий момент я чувствую почти полную удовлетворенность этими выводами. Я не вижу у них каких-либо опасных следствий, и они дают возможность размышлять и теоретизировать о сознании в таком ключе, который превосходит иные варианты почти по всем параметрам. И расширение научного мировоззрения принесло позитивные результаты, по крайней мере для меня: мир стал казаться более интересным местом.

В этой книге четыре части. В первой я раскрываю проблемы и создаю модель, внутри которой с ними можно будет работать. Глава 1 — это введение в вопрос о сознании, разграничение ряда смежных понятий, указание того смысла, в котором сознание является по-настоящему интересным, и предварительное объяснение его нетривиального отношения к иным аспектам ментального. Глава 2 развивает метафизическую и объяснительную модель, внутри которой разворачивается значительная часть последующего обсуждения. Что означает редуктивное объяснение феномена, что означает его физичность? В этой главе разъясняются подобные вещи с акцентом на понятии супервентности. Я доказываю, что у нас есть серьезное основание считать, что *почти* всё в мире может получить редуктивное объяснение; но сознание может оказаться исключением.

После этих предварительных работ вторая часть фокусируется на нередуцируемости сознания. В главе 3 доказывается, что стандартные методы редуктивного объяснения не могут объяснить сознание. Я также критикую различные редуктивные концепции, предлагавшиеся в нейронауке, когнитивной науке и в других областях. Речь не идет о всего лишь негативном выводе: из него следует, что удовлетвори-

тельная теория сознания должна быть новой разновидностью *нередуктивной* теории. В главе 4 эта линия продолжается на основе доводов о ложности материализма и истинности одного из видов дуализма; здесь также обозначаются общие контуры возможной нередуктивной теории сознания. Глава 5 имеет по большей части оборонительный характер: здесь рассматриваются проблемы, которые, как кажется, связаны с моей позицией и касаются отношения между сознанием и нашими *суждениями* о сознании; и здесь доказывается, что эти проблемы не создают каких-либо фатальных трудностей для нее.

В третьей части я продвигаюсь к позитивной теории сознания. В каждой из трех глав здесь разрабатывается один из компонентов позитивной теории. Глава 6 фокусируется на «когерентности» между сознанием и когнитивными процессами, очерчивая ряд систематических связей между ними. Я использую эти связи для анализа и обоснования центральной роли, которую играют нейронаука и когнитивная наука при объяснении человеческого сознания. В главе 7 обсуждается отношение между сознанием и функциональной организацией; используются мысленные эксперименты для доказательства того, что сознание является «организационным инвариантом», то есть что всякая система с надлежащей функциональной организацией будет наделена однотипным сознательным опытом вне зависимости от того, из чего она сделана. В главе 8 рассматривается вопрос о том, какой могла бы быть *фундаментальная* теория сознания, и тут высказывается предположение, что в такой теории могло бы содержаться указание на тесную связь между сознанием и информацией. Эта глава, несомненно, наиболее спекулятивна, но на данном этапе, возможно, не обойтись без спекулятивных размышлений, если мы хотим двигаться вперед.

Последние две главы — это десерт. Здесь я применяю полученные ранее результаты к центральным вопросам, касающимся основоположений искусственного интеллекта и квантовой механики. В главе 9 доказывается тезис «сильного искусственного интеллекта», а именно то, что имплементация надлежащей компьютерной программы будет порождать сознающий ум. В главе 10 рассматривается запутанный вопрос о том, каким образом надо интерпретировать квантовую механику; развитые в предшествующих главах идеи о сознании используются здесь для того, чтобы поддержать «бесколлапсную» интерпретацию квантовой механики.

Не исключено, что наибольший отклик вызовет негативный материал, но моя подлинная цель позитивна: я хочу увидеть работающую *теорию* сознания. Когда я только пришел в философию, я был удивлен тем, что большинство дискуссий о сознании было сосредоточено на том, есть ли здесь проблема или нет, или на том, является ли оно

физическим, тогда как задача построения теорий, казалось, была отставлена в сторону. «Теории», похоже, выдвигали лишь те, кто (по моему мнению) не принимал сознание всерьез. Сейчас я уже тоже полюбил тонкости онтологических дебатов, но детальная теория по-прежнему остается моей главной целью. Если какие-то из идей этой книги будут полезны для других при построении ими лучшей теории, то мою попытку можно будет считать успешной.

Эта книга задумана как серьезная философская работа, но я старался сделать ее доступной для нефилософов. В моей концептуальной аудитории всегда находилось мое студенческое Я десятилетней давности: надеюсь, что я написал книгу, которая понравилась бы ему. Несколько параграфов книги имеют технический философский характер. Они отмечены звездочкой (*), и читатели вполне могут пропустить их. Наиболее технический материал содержится во второй и четвертой главах. В четвертом параграфе второй главы, в параграфах 2 и 3 четвертой главы, а также в финальном параграфе пятой главы обсуждаются запутанные проблемы философской семантики. Другие отмеченные звездочкой параграфы можно хотя бы пробежать глазами, чтобы представлять, о чем идет речь. Зачастую я переношу очень уж технический материал и соображения по философской литературе в примечания, находящиеся в конце книги. Единственным техническим понятием, имеющим ключевое значение для данной книги, является понятие супервентности, которое вводится в начале второй главы. Это понятие имеет устрашающее название, но выражает вполне естественную идею, и его хорошее уяснение поможет упорядочить главные проблемы. Значительную часть материала, расположенного в последующих частях этой главы, можно пропустить при первом прочтении, хотя впоследствии может возникнуть желание вернуться к нему для прояснения возникающих по мере дальнейшего продвижения по тексту вопросов.

Краткий тур, избегающий технических деталей, начните с главы 1, затем пробегитесь по начальным частям главы 2 для задания фона, потом прочтайте всю третью главу (при необходимости проскакивая какие-то части первого параграфа), где содержатся главные аргументы против редуктивного объяснения, а также первый и последний параграфы главы 4, где представлены главные соображения относительно дуализма. Стоит прочитать начало шестой главы, чтобы в общих чертах представлять очертания позитивного подхода. Если говорить о позитивном материале, то, возможно, наиболее самодостаточной и увлекательной является глава 7 с ее легкими для понимания мысленными экспериментами с кремниевыми мозгами; а любители смелых и туманных спекуляций могут получить удовольствие от главы 8. Наконец, главы 9 и 10 должны быть предназначены для тех, кто интересуется обсуждаемыми в них вопросами.

Пара философских замечаний. Философская литература о сознании весьма несистематична; мы видим разные и, как кажется, независимые течения, внутри которых обсуждаются сходные проблемы без установления контакта друг с другом. Я попытался привнести некую структуру в этот хаос посредством унифицирующей модели, проясняющей различные метафизические и объяснительные проблемы. Большинство дискуссий, представленных в литературе, можно без потерь перевести на язык данной модели, и я надеюсь, что эта структура выявит глубокие отношения, которые существуют между различными проблемными моментами.

Данная работа, возможно, необычна в том плане, что здесь мало обсуждается философское понятие тождества (к примеру, между ментальными и физическими состояниями), и предпочтение отдано понятию супервентности. Я нахожу, что обсуждения, выстраиваемые в терминах тождества, обычно скорее запутывают, чем проясняют ключевые проблемы, и зачастую приводят к упщению главных трудностей. Супервентность, напротив, создает, похоже, идеальную модель для отработки ключевых проблем. Для избежания расплывчатого философствования мы, однако, должны сосредоточиться на *силе* супервентностной связи: заверена ли она логической необходимостью, естественной необходимостью или чем-то еще? Широко признано, что сознание в каком-то смысле супервентно на физическом; реальный вопрос в том, насколько прочна эта связь. Обсуждения, игнорирующие эти модальные вопросы, обычно проходят мимо труднейших вопросов о сознании. Скептики в вопросе о модальных понятиях скептически отнесутся и ко всему моему обсуждению, но, на мой взгляд, нет другого хорошего способа упорядочить эти проблемы.

Одно из радостных обстоятельств, связанных с работой над этой книгой, состояло для меня в том, что проблема сознания оказалась сопряжена с глубокими вопросами, существующими во многих других областях науки и философии. Однако охват и глубина этой проблемы делают ее и слишком большой. Я остро сознаю, что чуть ли не каждое место в этой книге можно было бы развернуть в более детальное обсуждение и что во многих случаях я лишь коснулся поверхности. Но надеюсь, что мне удалось хотя бы показать достижимость прогресса по проблеме сознания без отрицания его существования или редуктирования его к тому, чем оно не является. Проблема — удивительна, и у нее захватывающее будущее.

— Нет. — Хия остановилась и медленно повернулась к нему. Ее зеленоватые глаза округлились.

— Здесь тебе небезопасно. — Она в ужасе взглянула на дом и побелела. — Сейчас же отправляйся к себе. Пока еще не поздно. И найди мне антидот.

— Какой еще антидот?

Хия скрылась за можжевельником, но ее последние слова рванули в голове Джоуи, как петарда:

— Антидот от зомбийного яда.

*Дайан Кертис Риган.
Моя подруга зомби*

Часть I

ОСНОВЫ

Глава 1

Два понятия ментального

1. Что такое сознание?

Сознательный опыт — самая обычная, но и самая таинственная вещь в мире. Нет ничего, что было бы более непосредственно известно нам, чем сознание, но далеко не ясно, как сочетать это с прочими нашими знаниями. Почему оно существует? Что оно делает? Как оно может возникать из комка серого вещества? Мы знаем сознание гораздо ближе, чем все остальное, но мы понимаем все остальное гораздо лучше, чем сознание.

Сознание может быть на удивление интенсивным. Это самый яркий феномен; нет ничего реальнее его. И при этом оно может раздражать своей эфемерностью: хорошо известно, как трудно, говоря о сознательном опыте, указать, о чем, собственно, идет речь. «Международный психологический словарь» даже не пытается напрямую охарактеризовать его:

Сознание: наличие восприятий, мыслей и чувств; осведомленность. Нельзя определить этот термин, не прибегая к терминам, непостижимым без уяснения того, что имеется в виду под сознанием. Многие попадают в ловушку смешения сознания с самосознанием — для сознания нужно лишь осознание внешнего мира. Сознание — удивительный, но ускользающий феномен: невозможно уточнить, что оно такое, что оно делает и почему эволюционировало. Ничего стоящего о нем не написано (Sutherland 1989).

Почти всякий, кто напряженно размышлял о сознании, поймет эти настроения. Сознание до такой степени неуловимо, что даже эта роб-

кая попытка дефиниции может быть оспорена: вполне могли бы существовать неосознаваемые восприятия и мысли, о чем свидетельствуют понятия подсознательного восприятия и бессознательного мышления. Ключевым для сознания, по крайней мере в самом его интересном смысле, является *опыт*. Это, однако, не дефиниция, а в лучшем случае разъяснение.

Попытка определить сознательный опыт в терминах более элементарных понятий не дает никаких результатов. С тем же успехом можно было бы попробовать определить *материю* или *пространство* в терминах чего-то более фундаментального. Все, на что мы способны, — проиллюстрировать и охарактеризовать его на его же уровне. Такие характеристики нельзя считать подлинными дефинициями из-за их неявной круговой природы, но они помогают пришипить предмет обсуждения. Я допускаю, что у всех читателей имеется сознательный опыт. И если все пойдет хорошо, то эти характеристики помогут уяснить, что мы говорим именно об *этом*.

Наш предмет, возможно, лучше всего охарактеризовать как «субъективное качество опыта». При восприятии, мышлении и действии запускается целый рой каузальных и информационных процессов, но обычно они не протекают в темноте. Тут есть и какой-то внутренний аспект; что-то, выражющее то, каково это — быть когнитивным агентом. Этот внутренний аспект и есть сознательный опыт. Сознательный опыт варьируется от ярких визуальных ощущений цветов до чувства едва уловимых привычных ароматов; от резких болей до ускользающих мыслей, которые мы вот-вот должны припомнить; от повседневных звуков и запахов до потрясающей музыки; от банального зуда до глубинного экзистенциального ужаса; от специфичности вкуса мяты конфеты до всеобщности опыта самости. Всем этим вещам присущи особые переживаемые в опыте качества. Все они играют видную роль во внутренней ментальной жизни.

Мы можем сказать, что некое существо сознательно тогда, когда есть *нечто*, выражющее то, каково это — быть этим существом, если воспользоваться фразой, прославленной Томасом Нагелем¹. Аналогично, ментальное состояние сознательно, если можно говорить о том, каково это — находиться в таком ментальном состоянии. То есть мы можем сказать, что ментальное состояние сознательно, если с ним связано квалитативное *чувство*, ассоциировано опытное качество. Эти квалитативные чувства известны также как феноменальные качества, или, кратко, *квалиа*². Проблема объяснения этих феноменальных качеств и есть проблема объяснения сознания. Это действительно трудная часть проблемы «сознание — тело».

Почему вообще должен существовать сознательный опыт? Он имеет основополагающее значение для субъективной точки зрения, но с

объективной точки зрения он кажется чем-то совершенно непредвиденным. Находясь на объективной позиции, мы можем рассуждать о тонких взаимодействиях полей, волн и частиц пространственно-временного многообразия, приводящих к возникновению таких сложных систем, как мозг. В принципе, в том факте, что подобные системы могут комплексно обрабатывать информацию, реагировать на стимулы хитроумным поведением и даже обнаруживать у себя такие сложные способности, как способность к обучению, памяти и языку, не заключено никакой глубокой философской тайны. Все это впечатляет, но не поражает в метафизическом смысле. А вот существование сознательного опыта кажется с этой позиции какой-то *новой* характеристикой. Мы не предсказали бы его только из тех параметров.

В общем, в сознании есть что-то *неожиданное*. Если бы мы знали лишь физические факты или даже факты, имеющие отношение к динамике сложных систем и обработке информации в них, у нас не было бы убедительных оснований постулировать существование сознательного опыта. Если бы это не было прямо засвидетельствовано с позиции первого лица, такая гипотеза казалась бы необоснованной, а возможно, и мистической. И все же мы напрямую знаем о существовании сознательного опыта. Вопрос в том, как мы согласуем это со всеми остальными нашими познаниями?

Сознательный опыт есть часть природного мира, и, подобно другим естественным феноменам, он требует объяснения. Здесь можно говорить по меньшей мере о двух главных вещах, которые надо объяснить. Первой и наиболее важной из них является само *существование* сознания. Почему существует сознательный опыт? Если — что выглядит правдоподобным — он порождается физическими системами, то как это происходит? Это подталкивает к более конкретным вопросам. Является ли само сознание физическим, или же оно попросту сопутствует физическим системам? Насколько распространено сознание? Есть ли сознательный опыт, к примеру, у мышей?

Вторая вещь, нуждающаяся в объяснении, — это специфический *характер* сознательного опыта. Если допустить существование сознательного опыта, то почему конкретные переживания устроены именно так? Когда я открываю глаза и осматриваю свой кабинет, почему передо мной раскрывается многообразный опыт именно *такого* рода? Или более фундаментально: почему красное видится *так*, а не *иначе*? Кажется представимым, что, глядя на что-то красное, к примеру на розу, можно было бы испытывать то, что фактически испытывается при взгляде на синее. Почему опыт именно *таков*? И почему, кстати, наше ощущение красновато³, а не выглядит как-то совершенно иначе, скажем, как звук трубы?

Когда кто-то нажимает среднюю клавишу До на фортепиано, это дает начало сложной цепи событий. В воздухе возникают звуковые вибрации, и волна попадает в мое ухо. Эта волна обрабатывается, разбивается на частоты в моем ухе, и сигнал отправляется в слуховую кору. Здесь происходит дальнейшая обработка: выделение различных аспектов этого сигнала, категоризация и, в конечном счете, реакция. Все это в принципе несложно понять. Но почему это должно сопровождаться *опытом*? И почему, в частности, это должно сопровождаться *именно таким* опытом, с характерным богатым тоном и тембром? Это два центральных вопроса, на которые мы хотели бы получить ответ в теории сознания.

В идеале мы хотели бы, чтобы теория сознания делала по меньшей мере следующее: она должна была бы указывать условия, при которых физические процессы порождают сознание, и по отношению к процессам, порождающим сознание, она должна была бы специфицировать, какой именно тип опыта оказывается связан с ними. И мы хотели бы, чтобы эта теория объясняла, *как* оно порождается, чтобы возникновение сознания казалось понятным, а не магическим. В итоге мы хотели бы, чтобы она позволила нам увидеть сознание в качестве интегральной части природного мира. Возможно, пока нам трудно уяснить, какого рода была бы эта теория, но без нее нельзя будет сказать, что мы полностью понимаем сознание.

Прежде чем идти дальше, одно терминологическое замечание. Термин «сознание» неоднозначен и отсылает к множеству феноменов. Иногда он используется для обозначения когнитивных способностей, таких как способность интроспекции или способность давать отчет о ментальных состояниях. Иногда он используется как синоним «бодрствования». Иногда он тесно увязывается с нашей способностью фокусировать внимание или произвольно контролировать наше поведение. Иногда «сознание чего-то» отождествляется с «знанием о чем-то». Всё это принятые способы употребления данного термина, но все они говорят о феноменах, отличных от того предмета, о котором я веду речь, и эти феномены гораздо менее трудны для объяснения. Позже я еще вернусь к этим альтернативным понятиям сознания, пока же, говоря о сознании, я имею в виду исключительно субъективное качество опыта: каково это — быть когнитивным агентом.

Множество альтернативных терминов и выражений обозначает примерно тот же класс феноменов, что и «сознание» в его главном смысле. Среди них «опыт», «квалиа», «феноменология», «феноменальный», «субъективный опыт» и «каково это». Помимо грамматических отличий, различия между этими терминами сводятся по большей части к смысловым оттенкам. «Осознавать» в этом смысле — примерно

то же самое, что и «обладать квалиа», «иметь субъективный опыт» и т. д. Любые отличия в классе обозначаемых феноменов являются несущественными. Подобно «сознанию», многие из этих терминов несколько неоднозначны, но я никогда не буду использовать их в альтернативных смыслах. Говоря о главном предмете этой книги, я буду использовать все эти выражения, но четче всего к нему отсылают «сознание» и «опыт», и они будут преобладать.

Наблюдение за сознательным опытом может быть увлекательным делом. Опыт исключительно разнообразен, и каждой его разновидности присущ свой собственный характер. В нижеследующем списке, импрессионистичном и не претендующем на теоричность, представлен далекий от полноты каталог аспектов сознательного опыта. Ничего из этого не надо принимать слишком уж всерьез, как философию, и этот список нужен для того, чтобы помочь сфокусировать внимание на предмете обсуждения.

Каталог сознательного опыта

Визуальный опыт. Среди множества разновидностей визуального опыта ощущения цвета выступают в качестве эталонов сознательного опыта — из-за их чистой и, как кажется, невыразимой квалитативной природы. Некоторые колористические переживания могут казаться особенно поразительными и поэтому могут быть особенно хороши для того, чтобы сфокусировать наше внимание на тайне сознания. Вот кричащий багровый оттенок книги на моей полке, почти сюрреалистичная зелень папоротника на фотографии на стене и мигание множества ярких огоньков, красных, зеленых, оранжевых и голубых, на рождественской елке, которая видна из моего окна. Но изумление может вызвать любой цвет — стоит лишь отметить его и задуматься о его природе. Почему он должен переживаться *именно так*? Почему он вообще должен переживаться? Как я мог бы передать природу этого колористического опыта тому, кто был бы лишен его?

К другим аспектам визуального опыта относится опыт формы, размера, яркости и затемненности. Особенно утонченным аспектом является опыт глубины. Когда я был маленьким, я замечательно видел одним глазом, а другим — очень плохо. Поскольку один мой глаз был хорошим, мир выглядел четким и резким и, несомненно, трехмерным. Но вот мне дали очки, и разница была очевидна. Мир не стал намного более резким, чем раньше, но неожиданно стал казаться *более* трехмерным: вещи, имевшие глубину, каким-то образом стали еще более глубокими, и мир выглядел более насыщенным. Если вы прикроете один глаз, а затем раскроете его, вы сможете понять, что я имею в виду. В своем прежнем состоянии я бы сказал, что глу-

бина моего зрения не подлежит улучшению; мир уже казался таким трехмерным, каким он только и мог бы быть. Изменение было тонким, почти невыразимым, но удивительно впечатляющим. Разумеется, можно рационально объяснить, каким образом бинокулярное зрение позволяет объединять информацию от каждого глаза в информацию о расстояниях, позволяя тем самым более эффективно контролировать действия, но этому каузальному объяснению не удается передать то, как *переживается* этот опыт. Почему то изменение в процессах обработки информации должно сопровождаться подобной перекомпоновкой моего опыта, было тайной для меня в десятилетнем возрасте и до сих пор изумляет меня.

Слуховой опыт. В некоторых отношениях звуки еще более необычны, чем визуальные образы. Структура образов, как правило, напрямую соответствует структуре мира, но звуки могут казаться вполне независимыми. На мой телефон приходит вызов, внутренний механизм вибрирует, в воздухе возникает сложная волна, достигающая в итоге моей барабанной перепонки, — и каким-то образом, почти магическим, я слышу звонок. Кажется, что квалитативность звонка никак напрямую не соответствуют какой-либо структуре в мире, хотя я точно знаю, что он идет от динамика и что он определяется формой волны. Но почему эта форма волны или даже эти нейронные импульсы породили *именно такое* аудиальное качество?

Быть может, богатейшим аспектом слухового опыта является музыкальный опыт, хотя речевой опыт не должен сильно отставать от него. Музыка может литься и полностью поглощать нас, окружая нас, подобно тому как может окружать нас визуальное поле, тогда как со слуховыми переживаниями дело обычно обстоит иначе. Можно проанализировать аспекты музыкального опыта, разбивая воспринимаемые нами звуки на ноты и тона, находящиеся в сложных взаимных отношениях, но музыкальный опыт каким-то образом не сводится к этому. Единый квалитативный опыт порождается аккордом, но не случайно отобранными нотами. Старенькое фортепьяно и далекий гобой могут совместно производить неожиданно сильное впечатление. И, как всегда, раздумывая над этим, мы задаем вопрос: почему все *это* должно чувствоваться *именно так*?

Тактильный опыт. Еще одним богатейшим квалитативным пространством нашего опыта являются текстуры: сравните ощущение бархата с текстурой холодного металла, потной руки или щетинистого подбородка. У всех у них имеется свое особое уникальное качество. Тактильные ощущения воды, сахарной ваты или губ другого человека столь же различны.

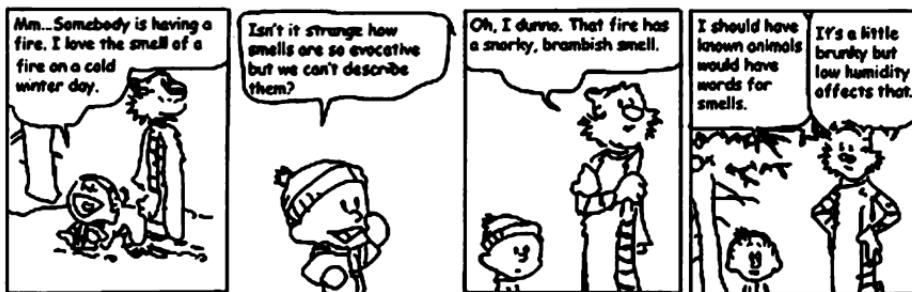


Рис. 1.1. Выразимость и невыразимость обонятельного опыта. (Calvin and Hobbes © Watterson. Distributed by Universal Press Syndicate. Перепечатка разрешена. Все права сохранены.)

- (1) — Ммм... Кто-то развел костер. Люблю запах костра в холодный зимний день.
- (2) — Не странно ли, что запахи так бередят память, но мы не можем описать их?
- (3) — Ох, не знаю. Тот костер пахнет так огневито и хворостисто.
- (4) — Я мог бы и догадаться, что у животных есть слова для запахов. — Немного брунковат, из-за сухого воздуха.

Обонятельный опыт. Вспомните заплесневелый запах старого гардероба, вонь гниющих отходов, дух свежескошенной травы, теплый аромат только что испеченного хлеба. В каком-то смысле обоняние — самое таинственное из всех чувств из-за богатой, неосязаемой и невыразимой природы ощущений запаха. Экерман (Ackermann 1990) называет его «молчаливым и бессловесным чувством». Хотя любое ощущение содержит в себе неуловимое, другие чувства обладают определенными свойствами, облегчающими их описание. Визуальный и слуховой опыт имеют сложную комбинаторную структуру, которая может быть описана. Тактильные и вкусовые переживания обычно возникают вследствие прямого контакта с каким-то объектом, и через отсылки к такого рода объектам был создан богатый дескриптивный словарь. Запахи кажутся малоструктурированными и зачастую лишены очевидной связи с каким-либо объектом, примитивным образом присутствуя в многообразии наших ощущений. (Возможно, животные справлялись бы с ними лучше нас (рис. 1.1).) Быть может, эта примитивность отчасти связана с тем, что процессы, благодаря которым наши обонятельные рецепторы оказываются чувствительными к различным видам молекул, протекают по принципу ключа и замочной скважины. Кажется случайным, что данная разновидность молекул должна вызвать ощущение *именно такого* рода, но это и происходит.

Вкусовой опыт. Психофизические изыскания показывают, что существует всего лишь четыре самостоятельных измерения вкусового

восприятия: сладкое, кислое, горькое и соленое. Но это четырехмерное пространство в сочетании с нашим обонянием производит громадное множество всевозможных ощущений: от вкуса лукума, салата из спаржевой фасоли с карри⁴ и мятного «Лайфсейвера» до вкуса спелого персика.

Опыт теплого и холодного. Угнетающе жаркая духота и морозный зимний день вызывают совершенно разные квалитативные ощущения. Представьте также ощущения жара на коже вблизи огня и ощущение обжигающего холода при прикосновении к очень холодному льду.

Боль. Боль — эталонный пример сознательного опыта, любимый философами. Быть может, это сопряжено с тем, что болевые ощущения составляют совершенно особый класс квалитативных переживаний, и эти ощущения непросто напрямую соотнести с какой-либо структурой в мире или в теле, хотя они обычно и связаны с определенными частями тела. Поэтому боль может казаться даже более субъективной, чем большинство чувственных переживаний. Существует огромное множество болевых ощущений, от пристрелов и острых приступов до уколов и ноющей боли.

Другие телесные ощущения. Боль — всего лишь наиболее заметный вид ощущений, связанных с определенными частями тела. Среди других подобных ощущений — нытье в голове (которое, возможно, является разновидностью боли), приступы голода, зуд, щекотание и позывы к мочеиспусканию. Многие телесные ощущения обладают совершенно уникальными качествами, отличными по типу от всего нашего остального опыта: подумайте, к примеру, об оргазмах или о чувстве, возникающем, когда вы задеваете локтевой нерв. А есть еще и ощущения, связанные с проприоцепцией — чувством расположения в пространстве нашего собственного тела.

Ментальные образы. Продолжая погружение в опыт, не связанный с конкретными объектами вокруг нас или с телом, но в известном смысле являющийся внутренним порождением, мы приходим к ментальным образам. Зачастую визуальные образы, созданные нашим воображением, обладают богатой феноменологией, хотя они и не так детализированы, как те, что извлечены из прямого визуального восприятия. Кроме того, если закрыть глаза или прищуриться, возникают любопытные цветовые паттерны, а после взгляда на что-то яркое — сильные остаточные образы. Воображение может создавать и аналогичные слуховые «образы», и даже тактильные, обонятельные и вкусовые, хотя их не так просто удержать, и связанное с ними квалитативное чувство обычно не столь заметно.

Осознанное мышление. Некоторые из объектов наших мыслей и убеждений не сопряжены с каким-то особым квалитативным чувством, но многие не лишены его. В особенности это справедливо по отношению к детальным размышлениям, разворачивающимся перед нами, и к разного рода мыслям, оказывающим влияние на наш поток сознания. Зачастую непросто указать, какое именно квалитативное чувство присуще той или иной возникшей у нас мысли, но оно несомненно присутствует. Здесь есть *что-то*, что выражает то, каково это — обладать подобными мыслями.

Когда я, к примеру, думаю о льве, в мою феноменологию проникает некое качество львиности: характер мысли о льве неуловимым образом отличается от характера мысли об Эйфелевой башне. Феноменологический аромат — нередко сильный — более очевиден у таких когнитивных установок, как желания. Желания словно тянут за собой феноменологический «канат»; квалитативный компонент зачастую имеется и у памяти, в частности когда мы имеем дело с переживаниями ностальгии или раскаяния.

Эмоции. Эмоции нередко связаны с особыми переживаниями. Оживленность приподнятого настроения, томление глубокой депрессии, накал вспышки гнева, меланхолия раскаяния — все это может глубоко влиять на сознательный опыт, хотя и гораздо менее определенным образом, чем при локализованном опыте, таком как ощущения. Эти эмоции при их наличии пронизывают и окрашивают весь наш сознательный опыт.

Другие, более мимолетные чувства занимают промежуточное место между эмоциями и более выраженным когнитивными аспектами ментального. Возьмем удовольствие от шутки. Другой пример — напряжение при просмотре триллера или при ожидании важного события. Мандраж, иногда сопровождающий обеспокоенность, тоже относится к этому классу.

Чувство Я. Иногда мы чувствуем, что в сознательном опыте кроме всех этих специфических элементов есть еще что-то: скажем, фоновый шум, неким образом существенный для сознания и имеющийся даже при отсутствии других компонентов. Эта феноменология Я настолько глубока и неуловима, что подчас кажется иллюзорной, на деле сводясь исключительно к специфическим элементам, подобным тем, что были перечислены выше. И все же кажется, что в этой феноменологии Я *что-то* есть, хотя и очень непросто со всей точностью указать, что именно.

Этот каталог охватывает ряд базовых случаев, но не меньше остается за его рамками. К примеру, я ничего не сказал о сновидениях,

возбужденности и усталости, интоксикации или о других, необычных переживаниях, вызванных химическими препаратами. Имеются и сложные переживания, происходящие из сочетания двух или большего количества вышеописанных компонентов. Я упоминал совместные действия обоняния и вкуса, но не менее распространенным примером является опыт сочетания музыки и эмоций, тесно сплетающихся в своем утонченном взаимодействии. Я также оставил в стороне опыт единства сознания — кажущуюся увязку всех указанных переживаний в опыт одного субъекта. Подобно чувству Я, это единство порой кажется иллюзорным — и совершенно точно, что выявить его труднее, чем какое-либо из конкретных переживаний — но интуиция того, что это единство существует, сильна.

К сожалению, мы больше не будем детализировать богатое разнообразие сознательного опыта. При рассмотрении философских тайн, связанных с сознательным опытом, простое ощущение цвета поднимает столь же глубокие проблемы, что и опыт прослушивания хорала Баха. Глубокие вопросы так соотносятся с этим разнообразием, что рассмотрение природы конкретных переживаний оказывается не столь уж важным. И все же этот краткий обзор богатого разнообразия сознательного опыта должен помочь сфокусировать внимание на предмете обсуждения и дать набор примеров, о которых надо помнить при более абстрактном рассмотрении⁵.

2. Феноменальное и психологическое понятие ментального

Ментальное не сводится к сознательному опыту. Чтобы убедиться в этом, заметим, что, хотя современная когнитивная наука почти ничего не говорит о сознании, она может немало сказать о ментальном вообще. Ее интересуют другие аспекты ментального. Когнитивная наука имеет дело главным образом с объяснением поведения, и в той мере, в какой ее вообще интересует ментальное, ее интересует ментальное, истолкованное как внутренняя основа поведения, и ментальные состояния, понятые как состояния, ответственные за продуцирование и объяснение поведения. Подобные состояния могут быть, но могут и не быть сознательными. С точки зрения когнитивной науки внутреннее состояние, ответственное за продуцирование поведения, ментально вне зависимости от его осознанности.

В основе всего этого лежат два совершенно разных понятия ментального. Первое — *феноменальное* понятие ментального. Это понятие ментального, понятого в качестве сознательного опыта, и ментального состояния как осознанно переживаемого ментального состояния. Это

наиболее загадочный аспект ментального, и именно на этом аспекте я сосредоточу свое внимание. Однако он не исчерпывает ментального. Второе — *психологическое* понятие ментального. Это понятие ментального как каузальной, или объяснительной, основы поведения. Некое состояние является ментальным в этом смысле, если оно играет надлежащую каузальную роль в продуцировании поведения или, по крайней мере, играет надлежащую роль в объяснении поведения. Согласно психологическому понятию, вопрос о том, наделено ли ментальное состояние качеством сознания, не имеет большого значения. Значима лишь его роль в когнитивной экономии.

В соответствии с феноменальным понятием ментальное характеризуется тем, как оно *чувствуется*; в соответствии с психологическим понятием ментальное характеризуется тем, что оно *делает*. И речь здесь не идет о столкновении двух этих понятий ментального. Ни одно из них не является *единственно верным* анализом ментального. Они покрывают различные феномены, но обе эти группы феноменов совершенно реальны.

Иногда я буду говорить о феноменальном и психологическом «аспектах» ментального, иногда — о «феноменальном ментальном» и «психологическом ментальном». На этой ранней стадии я не хочу предрешать каких-либо вопросов о том, не окажутся ли они одним и тем же. Не исключено, что всякое феноменальное состояние является психологическим состоянием, играя важную роль в продуцировании и объяснении поведения. Возможно также, что всякое психологическое состояние тесно соотнесено с феноменальным. Пока для нас важно лишь концептуальное различие этих двух понятий: быть феноменальным состоянием — значит переживаться определенным образом, быть психологическим состоянием — значит играть надлежащую каузальную роль. Эти различные понятия не должны объединяться, по крайней мере на начальных этапах рассмотрения.

Конкретное ментальное понятие обычно может быть проанализировано или как феноменальное понятие, или как психологическое понятие, или как комбинация того и другого. К примеру, ощущение в его главном смысле лучше всего трактовать как феноменальное понятие: ощущать — значит находиться в состоянии определенного чувствования. А вот понятия обучения и памяти могли бы лучше всего толковаться как психологические. Обучение чего-либо, при первом приближении, — это надлежащая адаптация поведенческих способностей этого нечто в ответ на определенные воздействия, идущие из его окружения. В общем, феноменальные черты ментального характеризуются тем, каково это — обладать данными чертами для какого-то субъекта, а психологические черты характеризуются сопряженной с ними ролью в порождении и/или объяснении поведения.

Разумеется, такое использование термина «психологический» — условность: оно идет от отождествления психологии с когнитивной наукой в вышеуказанном смысле. Обыденное понятие «психологического состояния», возможно, выходит за эти границы и вполне может включать элементы феноменального. Но от этого словоупотребления ничего не зависит.

Краткая история

Объединение феноменальных и психологических аспектов ментального имеет давнюю историю. Возможно, Рене Декарт несет частичную ответственность за это. Его знаменитое учение о самопрозрачности ума приблизило его к отождествлению ментального и феноменального. Декарт считал, что всякое ментальное событие является *cogitatio*, то есть является собой содержание некоего опыта. К этому классу он присоединил воления, намерения и любые мысли. Отвечая на «Четвертые возражения», он писал:

Что же касается моего утверждения, что в нашем уме, поскольку он — вещь мыслящая, не может оставаться ничего не осознанного, то оно представляется мне само собой разумеющимся, ибо в уме, рассматриваемом таким образом, не может быть ничего, что не было бы мыслью или чем-то зависящим от мысли; в противном случае оно не имело бы никакого отношения к уму — вещи мыслящей; и у нас не может быть ни одной мысли, которую мы бы не осознавали в тот самый момент, когда она у нас появляется. (Перевод С. Я. Шейнман-Топштейн.)

И даже если Декарт на деле не отождествлял психологическое с феноменальным, он по крайней мере допускал, что всё психологическое, достойное именоваться ментальным, имеет сознательный аспект⁶. Для Декарта понятие бессознательного ментального состояния — это противоречие.

Разделение этих аспектов ментального было обязано прогрессу в области скорее не философских, а психологических теорий. Всего лишь сто лет назад такие психологи, как Вильгельм Вундт и Уильям Джемс, оставались картезианцами в том, что они использовали интроспекцию для исследования причин поведения и строили психологические теории на основе интроспективных данных. Соответственно, феноменология была сделана арбитром психологии. Но появившиеся вскоре после этого разработки привели к утверждению психологического как автономной области.

Прежде всего речь идет о развитой Фрейдом и его современниками идеи о том, что многие ментальные акты имеют бессознательный характер и что могут существовать бессознательные убеждения и желания. Сам факт того, что это понятие казалось когерентным, свидетельствует об использовании нефеноменального анализа мышления. Фрейд, похоже, *каузально* конструировал понятия. В самом общем

виде желание неявным образом конструировалось как состояние такого рода, которое продуцирует определенное поведение, связанное с объектом желания. Аналогично, в соответствии с его каузальной ролью, конструировалось и убеждение. Конечно, Фрейд не эксплицировал эти анализы, но ясно, что нечто подобное им лежит в основе его использования этих понятий. Эксплицитно же он признавал, что сознательный доступ к какому-либо состоянию несуществен для объяснения им поведения и что сознательность не конституирует бытие чего-то в качестве убеждения или желания. Эти выводы опираются на независимое от феноменальных понятий понятие ментального.

Примерно в то же время бихевиористское движение в психологии полностью отвергло интроспекционистскую традицию. Был разработан новый, «объективный» вид психологического объяснения, в котором не оставалось места для сознания. Этот вид объяснений имел лишь ограниченный успех, но он утвердил идею о том, что психологическое объяснение может игнорировать феноменальное. В теоретических взглядах бихевиористов не было единства: одни признавали существование сознания, но считали, что оно не имеет отношения к психологическим объяснениям, другие вовсе отрицали его существование. Многие из них пошли еще дальше, отрицая существование *любых* ментальных состояний. Официальной причиной этого было предположение о том, что внутренние состояния не имеют отношения к объяснению поведения, которое может осуществляться исключительно на основе отсылки к внешним данным. Но, возможно, более глубокой причиной было то, что все ментальные понятия были подернуты дымкой уже дискредитированного феноменального.

В любом случае две эти разработки сделали ортодоксальной мысль о том, что объяснение поведения никоим образом не зависит от феноменальных понятий. Сдвиг от бихевиоризма к вычислительной когнитивной науке по большей части сохранил эту ортодоксию. Хотя этот сдвиг вернул определенную роль внутренним состояниям, которые могли даже называться «ментальными» состояниями, в них не было ничего особенно феноменального. Эти состояния допускались именно потому, что они были нужны для объяснения поведения; любые же связанные с ними феноменальные качества оказывались в лучшем случае нерелевантными. В фокусе тем самым оказалось понятие ментального как психологического.

В философии смещение акцента с феноменального на психологическое получило систематическое выражение у Гилберта Райла (Ryle 1949), доказывавшего, что все наши ментальные понятия могут быть проанализированы в терминах определенных видов связанного с ними поведения или в терминах предрасположенности вести себя

определенным образом⁷. Эта позиция — логический бихевиоризм — явным образом предвосхитила многое из того, что считается ортодоксией в современной философии психологии. В частности, она являла собой наиболее детальную кодификацию связи между ментальными понятиями и продуцированием поведения.

Райл подавал свою теорию не в качестве анализа лишь *некоторых* ментальных понятий. Он намеревался охватить ею все такие понятия. Многие полагают, как полагаю и я, что эта установка заведомо непригодна для анализа наших феноменальных понятий, таких как ощущение и само сознание. Многим казалось очевидным, что, когда мы говорим о феноменальных состояниях, мы уж точно не говорим о нашем поведении или о какой-то поведенческой диспозиции. Но в любом случае анализ Райла указал перспективный подход ко многим другим ментальным понятиям, таким как убеждение, радость, желание, притворство и припоминание.

Помимо проблем с феноменальными состояниями, позиция Райла не была лишена и некоторых технических трудностей. Во-первых, естественно допустить, что ментальные состояния *продуцируют* поведение, но если ментальные состояния сами относятся к поведению или являются поведенческими диспозициями, в противоположность внутренним состояниям, то трудно понять, как они могли бы выполнять эту задачу. Во-вторых, был выдвинут аргумент (Chisholm 1957, Geach 1957), что никакое ментальное состояние не может быть определено одним набором поведенческих диспозиций, независимых от других ментальных состояний. Так, если кто-то убежден, что идет дождь, его поведенческие диспозиции будут меняться в зависимости от того, хочет ли он промокнуть. Поэтому при характеристике поведенческих диспозиций, связанных с данным ментальным состоянием, надо обращаться к другим ментальным состояниям.

Эти проблемы были хитроумно разрешены в так называемом *функционализме*, разработанном Дэвидом Льюисом (Lewis 1966) и, детальнее всего, Дэвидом Армстронгом (Armstrong 1968)⁸. Согласно этому воззрению, ментальное состояние определяется исключительно посредством его *каузальной роли*, то есть в терминах различных типов стимулов, которые обычно производят его, типа поведения, которое оно обычно производит, и того, как оно взаимодействует с другими ментальными состояниями. Эта концепция полностью интериоризировала ментальные состояния, наделив их возможностью находиться в надлежащем каузальном отношении к поведению, отвечая тем самым на первое возражение, и она допустила возможность определения ментальных состояний в терминах их взаимодействия, давая тем самым ответ на второе возражение.

Согласно этой концепции, наши ментальные понятия могут быть подвергнуты *функциональному анализу* — в терминах их фактических или типичных причин и действий. Осуществление подобного анализа для какого-либо конкретного ментального понятия — крайне непростое дело. Армстронг (Armstrong 1968) проводит ряд таких анализов, но они оказываются очень далеки от завершенности. Но в принципе функционализм может давать приемлемое толкование многих наших ментальных понятий, во всяком случае в той мере, в какой они играют какую-то роль в объяснении поведения. К примеру, понятие обучения могло бы быть проанализировано как адаптация чьих-либо поведенческих способностей в ответ на стимулы, идущие из окружающей среды. Если взять более сложное состояние, убеждение в том, что идет дождь, то в самом общем виде оно могло бы быть проанализировано как состояние такого рода, которое обычно порождается, когда идет дождь, которое вызывает поведение, уместное при дожде, определенным образом инференциально взаимодействует с другими убеждениями и желаниями и т. д. Детали нуждаются в значительной проработке, но многие сочли общую идею в целом верной.

Подобно Райлу, Армстронг и Льюис, однако, преподносили это не в качестве анализа *некоторых* ментальных понятий. Скорее, он был задуман как анализ *всех* ментальных понятий. В частности, они доказывали, что подобным образом могут быть проанализированы понятия опыта, ощущения, сознания и т. п. Эта ассоциация феноменального в психологическое представляется мне не меньшей ошибкой, чем декартовская ассоциация психологического в феноменальное. Это попросту неверный анализ значения феноменального. Когда мы задаемся вопросом, имеется ли у кого-то опыт цвета, мы не спрашиваем о том, получает ли он стимулы из окружающей среды и обрабатывает ли он их определенным образом. Нас интересует, *переживает* ли он ощущение цвета, и это — другой вопрос. Концептуально непротиворечивой возможностью является то, что нечто могло бы играть данную каузальную роль, не будучи связанным с опытом.

Иначе говоря, этот анализ феноменальных понятий оставляет неясным то, почему кого-то вообще должна была беспокоить эта проблема⁹. В том, как некое состояние могло бы играть определенную каузальную роль, нет особой тайны, хотя тут, конечно, есть технические научные проблемы. Таинственно то, почему такое состояние должно как-то *переживаться*; почему оно должно иметь некое феноменальное качество. Вопросы о том, почему исполняется каузальная роль и почему имеется феноменальное качество, совершенно различны. Функционалистский анализ отрицает раздельность этих вопросов и поэтому представляется неудовлетворительным.

Позже я гораздо более подробно буду говорить об этом, пока же мы можем отметить, что даже если функционалистский подход не позволяет удовлетворительно проанализировать феноменальные понятия, он может хорошо справляться с анализом других ментальных понятий, таких как обучение и память, а возможно, и убеждение. В этих случаях не возникает тех проблем, о которых шла речь выше. В том, что некая система должна быть способна к обучению, не больше таинственного, чем в том, что некая система должна быть способна адаптировать свое поведение в ответ на стимулы, идущие из окружающей среды; и впрямь речь здесь идет примерно об одном и том же. Аналогичным образом, когда мы спрашиваем, обучился ли некто чему-то, кажется вполне приемлемым сказать, что, задаваясь этим вопросом, мы интересуемся, претерпел ли он такое изменение, которое улучшит его способность справляться с определенными ситуациями в будущем. Разумеется, исчерпывающий анализ понятия обучения будет более утонченным, чем этот, данный в первом приближении, но его дальнейшая детализация будет проводиться в тех же самых рамках.

В самом деле, функционалистский подход точно соответствует данной мной дефиниции *психологических* свойств. Большинство не-феноменальных ментальных свойств попадает в этот класс и поэтому может быть подвергнуто функциональному анализу. Разумеется, всегда можно спорить о деталях того или иного функционалистского анализа. Остаются и важные базовые вопросы о роли окружающей среды в характеристике психологических свойств и о том, что является определяющим связующим звеном между психологическими свойствами и поведением — причинность, объяснение или и то и другое вместе. Но в нашем случае эти детали не очень существенны. Важно, что не-феноменальные ментальные состояния во многом характеризуются их ролью в нашей когнитивной экономии.

Мораль, которую можно извлечь из этого обсуждения, состоит в том, что как психологическое, так и феноменальное — реальные и различные аспекты ментального. В первом приближении можно сказать, что феноменальные понятия имеют дело с теми аспектами ментального, которые связаны с перспективой от первого лица, тогда как психологические понятия соотнесены с аспектами, связанными с перспективой от третьего лица. Наш подход к ментальному будет весьма разниться в зависимости от того, какие аспекты ментального нас интересуют. Если нас интересует роль ментального в продуцировании поведения, мы сфокусируемся на психологических свойствах. Если нас интересует сознательный опыт ментальных состояний, мы сфокусируемся на феноменальных свойствах. Ни феноменальное, ни психологическое не должны переопределяться в терминах друг друга. Нельзя ис-

ключить, что глубокий анализ мог бы обнаружить некое фундаментальное звено, соединяющее феноменальное и психологическое, но отыскание такого звена было бы нетривиальной задачей, а не тем, что могло бы предрешаться в декларативном порядке. Ассимиляция феноменального в психологическое без всякого основательного разъяснения означала бы тривиализацию проблемы сознательного опыта; а ассимиляция психологического в феноменальное означала бы резкое ограничение той роли, которую играет ментальное в объяснении поведения.

3. Двойная жизнь ментальных терминов

Небезосновательным кажется утверждение, что психологическое и феноменальное исчерпывают ментальное. Иными словами, всякое ментальное свойство есть или феноменальное свойство, или психологическое свойство, или же некое сочетание того и другого. И правда, заинтересовавшись теми очевидными свойствами ментального, которые требуют объяснения, мы сначала наткнемся на разного рода сознательный опыт, а затем на продуцирование поведения. Ничего третьего, явно нуждающегося в объяснении, не дано, а два первых источника данных — опыт и поведение — не дают оснований верить в существование какого-либо третьего вида нефеноменальных, нефункциональных свойств (небольшим исключением, возможно, являются реляционные свойства, о которых мы вскоре поговорим). Разумеется, есть и другие классы ментальных состояний, которые нередко становятся предметом обсуждения — интенциональные состояния, эмоциональные состояния и т. п., но они, похоже, могут быть сведены к психологическим, феноменальным или же к их комбинации.

Дело осложняется тем, что многие обыденные ментальные понятия не соблюдают границ и содержат как феноменальные, так и психологические компоненты. Ярким примером может быть боль. Этот термин нередко используется для именования особого неприятного феноменального качества, и здесь на первый план выходит феноменальное понятие. Но с данным термином связано и психологическое понятие: а именно, в общем виде, понятие о такого рода состоянии, которое обычно вызывается повреждением организма, ведет к реакциям уклонения и т. п. Оба этих аспекта существенны для обыденного понятия боли. Мы могли бы сказать, что понятие боли раздвоено на феноменальное и психологическое понятия или же что оба они являются собой компоненты одного богатого понятия.

Задав вопрос о том, что более важно для боли — феноменальное качество или функциональная роль, мы можем столкнуться с необходимостью распутывания множества узлов. К примеру, можно ли ска-

зать, что гипотетической системе, в которой удовлетворены все функциональные критерии, но в которой отсутствует сознательный опыт, действительно больно? Может возникнуть искушение дать отрицательный ответ, но как быть с фактом наших рассуждений о боли, которая продолжалась весь день, хотя в какие-то его промежутки и не осознавалась? Едва ли, впрочем, стоит пытаться как-то разрешить этот вопрос. Ничего важного не зависит от семантического решения о том, является ли некое феноменальное качество *действительно существенным* для того, чтобы нечто могло считаться болью. Вместо этого можно признать, что с понятием связаны различные компоненты, — и четко различить их, говоря, к примеру, о «феноменальной боли» и «психологической боли». Не исключено, что наше обыденное понятие боли сочетает то и другое в некоей тонкой пропорции, но философское обсуждение будет более ясным, если развести их.

Причина того, что феноменальные и психологические свойства зачастую увязываются, очевидна — потому что они обычно сосуществуют. При протекании процессов, вызываемых повреждением тканей и ведущих к реакциям уклонения, обычно реализуется некое феноменальное качество. Иными словами, при наличии психологической боли, как правило, имеется и феноменальная боль. То, что данный процесс должен сопровождаться указанным феноменальным качеством, не является *концептуальной* истиной, но это тем не менее один из фактов нашего мира. При подобном сосуществовании данных свойств в обыденных ситуациях естественно то, что наши обыденные понятия будут связывать их друг с другом.

Многие ментальные понятия ведут двойную жизнь такого рода. К примеру, понятие *восприятия* может трактоваться сугубо психологически, обозначая процессы, благодаря которым когнитивные системы обнаруживают такую чувствительность к стимулам, идущим из окружающей среды, что возникающие в результате состояния играют определенную роль в управлении когнитивными процессами. Но его можно трактовать и феноменально, включая в него сознательный опыт воспринятого. Возможность подсознательного восприятия противоречит такой трактовке, но можно было бы сказать, что подсознательное восприятие может считаться восприятием только в ослабленном смысле соответствующего термина. Впрочем, и здесь проблема имеет терминологический характер. Если мы стремимся к ясности, мы можем просто оговорить, интересует ли нас психологическое свойство, феноменальное свойство или же их сочетание.

И все же некоторые двойные понятия оказываются ближе к феноменальным, тогда как другие — к психологическим. Возьмем понятие *ощущения*, находящееся в близком родстве с понятием восприятия и тоже имеющее как феноменальные, так и психологические компоненты.

Феноменальный компонент гораздо более заметен в «ощущении», чем в «восприятии», о чем свидетельствует тот факт, что идея бессознательного восприятия кажется более осмысленной, чем идея бессознательного ощущения. Размытость все же сохраняется: один из смыслов «восприятия» подразумевает сознательный опыт, а один из смыслов «ощущения» — нет, но эти смыслы кажутся менее важными, чем их альтернативы. Возможно, естественней всего было бы использовать «восприятие» как психологический термин, а «ощущение» — как феноменальный термин. Тогда ощущение может трактоваться как некий феноменальный коррелят восприятия.

Хорошим тестом на то, является ли ментальное понятие *M* по преимуществу психологическим, оказывается вопрос: могло бы нечто быть примером *M* без какого-либо связанного с ним феноменального качества? Если да, то *M*, скорее всего, психологическое понятие. Если нет, то *M* — феноменальное понятие или по меньшей мере комбинированное понятие, центральное место в котором занимает феноменология. Последнее нельзя исключить, так как для некоторых понятий может требоваться как надлежащее феноменальное качество, так и надлежащая когнитивная роль; не исключено, что подобный комбинированный характер присущ, к примеру, главному смыслу понятия ощущения. Но мы хотя бы можем отделить понятия, подразумевающие феноменологию, от тех, которые обходятся без нее.

Этот тест показывает, что, к примеру, понятие *обучения* является по большей части психологическим. В первом приближении обучаться чему-то — значит определенным образом адаптировать когнитивные способности к разнообразным новым ситуациям и стимулам. Для того чтобы когнитивный процесс был примером обучения, не требуется никакого особого феноменального качества; подобное качество может присутствовать, но данный процесс считается процессом обучения вовсе не из-за этого. Небольшой феноменальный привкус может порождаться связью с такими понятиями, как убеждение, о котором пойдет речь ниже, но он будет очень слаб. При объяснении обучения мы должны главным образом объяснить, как данной системе удается адаптироваться надлежащим образом. Нечто подобное мы видим и в понятиях категоризации и памяти, которые кажутся по большей части психологическими понятиями, поскольку главным в отношении них является исполнение надлежащей когнитивной роли.

У эмоций феноменальный аспект гораздо более очевиден. Думая о радости или огорчении, мы мыслим особые осознанные переживания. Не вполне ясно, однако, существует ли феноменальный аспект для того, чтобы некое состояние было эмоцией; сюда явным образом присоединяется и сильное психологическое свойство. Но, как обычно,

нам нет нужды принимать какое-либо решение по этому вопросу. Мы можем просто говорить о психологическом и феноменальном аспектах эмоции и отмечать, что они исчерпывают те аспекты эмоции, которые требуют объяснения.

Сложнее всего обстоит дело в случае таких ментальных состояний, как *убеждение*, нередко именуемых «пропозициональными установками», потому что они являются собой установки по отношению к пропозициям, сообщающим что-то о мире. Если я, к примеру, убежден, что Боб Дилан совершил тур по Австралии, я одобряю пропозицию, сообщающую что-то о Дилане; если я *надеюсь*, что Дилан совершил тур по Австралии, у меня имеется другая установка относительно той же пропозиции. Главной чертой этих ментальных состояний оказывается их семантический аспект, или *интенциональность*: тот факт, что они говорят *о* вещах в мире. Иными словами, убеждение имеет семантическое содержание: содержание моего вышеупомянутого убеждения состоит в пропозиции «Дилан совершил тур по Австралии» (хотя на этот счет могут быть и другие мнения).

Убеждение обычно трактуется как психологическое свойство. Согласно этой позиции, быть убежденным в истинности некоей пропозиции, в самом первом приближении, — значит находиться в таком состоянии, при котором субъект осуществляет такие действия, которые были бы уместными в случае ее истинности; которое, как правило, порождается выражаемым пропозицией положением дел, и таком, при котором когнитивная динамика размышлений субъекта отражает надлежащее взаимодействие данного убеждения с другими убеждениями и желаниями. Функциональные критерии убеждения, однако, трудно-уловимы, и никто еще не осуществил даже подобия полного анализа этих критериев. И все же есть основание считать, что эта позиция схватывает многое из того, что является существенным для убеждения. Она родственна взгляду, что убеждение есть что-то вроде *объяснительного конструкта*: мы приписываем убеждения другим людям главным образом для объяснения их поведения.

Кто-нибудь мог бы сказать, что эта картина неполна и что для убеждения нужно что-то еще помимо соответствующего психологического процесса. В частности, эта картина оставляет в стороне *опытные* аспекты убеждения, которые, как доказывали некоторые авторы, существенны для того, чтобы нечто могло считаться убеждением. К примеру, Серл (Searle 1990a) доказывал, что интенциональное содержание убеждения целиком зависит от сопряженного с ним состояния сознания или же от состояния сознания, которое может порождаться этим убеждением. В отсутствии сознания мы имеем дело исключительно с «как бы» интенциональностью¹⁰.

Ясно, что сознательный опыт нередко сопровождает убеждение: при наличии явного (то есть осознанного) убеждения всегда можно говорить о том, каково это — иметь его, а большинство неявных убеждений, по крайней мере, могут порождать осознанные убеждения. Ключевой вопрос, однако, в том, является ли это сознательное качество тем, что *создает* состояние убеждения, а также в том, придает ли оно ему его содержание. В случае некоторых убеждений это может казаться более правдоподобным, чем в случае других: так, можно было бы попробовать доказать, что сознательное качество требуется для того, чтобы по-настоящему обладать убеждениями о собственном *опыте*. Возможно, какие-то опытные переживания требуются и для наличия некоторых перцептивных убеждений о внешнем мире (не исключено, что для убежденности в красноте какого-то объекта необходимо обладать опытом красного). В других случаях это кажется более проблематичным. К примеру, когда я думаю о том, что Дон Брэдман — величайший игрок в крикет за всю историю этой игры, кажется вполне вероятным, что я обладал бы тем же самым убеждением, даже если бы с ним был связан совершенно другой сознательный опыт. Феноменология этого убеждения довольно размыта, и трудно понять, как данное феноменальное качество могло бы делать это убеждение убеждением о Брэдмане. Более существенной для содержания этого убеждения кажется связь этого убеждения и Брэдмана и та роль, которую оно играет в моей когнитивной системе.

Ослабленная позиция могла бы состоять в предположении, что, хотя для наличия конкретного убеждения не требуется какого-либо конкретного феноменального качества, для убежденности в чем-то требуется хотя бы способность существа обладать сознательным опытом¹¹. В идее о том, что существо, лишенное сознательной внутренней жизни, не может обладать убеждениями в подлинном смысле слова, есть некое правдоподобие; в лучшем случае оно могло бы обладать псевдоубеждениями. Тем не менее это делало бы роль феноменального в интенциональных понятиях весьма незначительной. Наиболее важные условия обладания конкретным убеждением будут находиться за пределами феноменального. Можно было бы даже вычесть все феноменальные компоненты, оставив понятие псевдоубеждения, во многих важных отношениях сходное с понятием убеждения и отличающееся от него лишь исключением понятия сознания. И в самом деле, похоже, что псевдоубеждение могло бы выполнять большую часть той объяснительной работы, которая выполняется понятием убеждения.

В любом случае здесь я не буду пытаться разрешить эти непростые вопросы об отношении интенциональности и сознания. Можно отметить, что по крайней мере существует дефляционное понятие убеждения, имеющее чисто психологический характер и не включающее

сознательный опыт; если существо находится в соответствующем психологическом состоянии, то оно находится в состоянии, во многих важных отношениях напоминающем убеждение, если не принимать во внимание феноменальные аспекты. И существует *инфляционное* понятие убеждения, согласно которому для подлинного обладания убеждениями требуется сознательный опыт, а возможно, и определенный сознательный опыт — для подлинной убежденности в какой-то конкретной пропозиции. Какое из этих понятий убеждения является «истинным», не имеет большого значения для моих нынешних целей.

Важно то, что не существует таких характеристик убеждения, которые выходили бы за пределы феноменального и психологического. Впрочем, здесь, возможно, надо сделать небольшую оговорку, добавив *реляционный* элемент для объяснения того факта, что некоторые убеждения могут зависеть не только от внутреннего состояния мыслящего существа, но и от состояния окружающей среды. К примеру, были попытки доказать, что убежденность субъекта в том, что вода — мокрая, предполагает его определенное отношение к окружающей воде. Такое отношение обычно трактуется как каузальное отношение, так что не исключено, что его можно встроить в характеристики соответствующего психологического свойства, каузальные силы которого будут простираться за пределы головы в окружающую среду. Если так, то не понадобится никакого дополнительного компонента. И в любом случае добавление, что в некоторых ментальных состояниях помимо психологического и феноменального мог бы присутствовать еще и реляционный компонент, не так уж обременительно. Так или иначе, тут не возникает какой-либо новой глубокой тайны.

Для уяснения того, что помимо феноменального и психологического/реляционного аспектов интенциональных состояний не существует никаких дополнительных глубинных аспектов, заметим, что очевидные для нас феномены, нуждающиеся в объяснении, распадаются на два класса, а именно: феномены, к которым мы имеем доступ с позиции третьего лица, и феномены, к которым мы имеем доступ с позиции первого лица. Те, что относятся к первому классу, в конечном счете сводятся к поведению, отношениям к окружающей среде и т. п. и могут быть отнесены к классу психологического и реляционного. Те, что относятся ко второму классу, сводятся к *опыту*, связанному с убеждениями, — к примеру, к тому, как наши понятия проявляются в феноменальном мире, — и таким образом составляют часть проблемы сознания, а не какую-то особую тайну. Основания для убежденности в каком-то конкретном аспекте убеждения (включая семантическое содержание, соотнесенность с чем-то и т. п.) будут происходить из одного из этих двух классов; не существует самостоятельного третьего класса феноменов, требующего от нас какого-то объяснения.

Другой способ убедиться в этом состоит в наблюдении, что после фиксации психологических, феноменальных и реляционных свойств индивида, похоже, не остается независимых ментальных переменных. Мы не можем даже *вообразить* субъекта, идентичного мне в упомянутых выше трех отношениях, но убежденного в чем-то ином, подобно тому как мы, похоже, можем *вообразить* психологически идентичного мне субъекта, который переживает что-то иное. В концептуальном пространстве попросту недостает места для этой возможности. Интенциональные понятия в каком-то смысле менее фундаментальны, чем психологические и феноменальные понятия, поскольку они не могут меняться независимо от них¹².

Все сказанное мной здесь по поводу убеждения в равной степени применимо и к другим интенциональным состояниям, таким как желание, надежда и т. п. Все эти состояния имеют психологический и феноменальный аспект, и у нас нет никакой необходимости решать, какой из них первичен, несмотря на то что можно привести сильные аргументы в пользу их психологического анализа. Важно то, что ни один аспект этих состояний не выходит за пределы психологического и феноменального (за исключением, возможно, реляционного компонента). Психология вместе с феноменологией составляют центральные аспекты ментального.

Одним из фактов, касающихся нашего ума, является то, что при реализации какого-либо феноменального свойства всякий раз реализуется какое-то психологическое свойство. Сознательный опыт не протекает в вакууме. Он всегда связан с когнитивными процессами и, похоже, в известном смысле порождается ими. Так, при ощущении происходит некая обработка информации — если угодно, соответствующее восприятие. Сходным образом при осознанном переживании радости некое внутреннее состояние, как правило, играет функциональную роль, связанную с радостью. Не исключена логическая возможность обладания данным опытом без подобной причинности, но их сопутствие кажется эмпирическим фактом.

Сосуществование феноменальных и психологических свойств

Ввиду этого сосуществования у кого-то может возникнуть малодушное искушение озадачиться тем, а было ли проведено реальное различие. Ясно, однако, что здесь есть хотя бы концептуальное различие, даже если объемы указанных понятий, как кажется, совпадают. Можно задаться вопросом об объяснении феноменального качества и можно задаться вопросом об объяснении исполнения каузальной роли, и это разные вопросы.

Тем не менее это сосуществование феноменальных и психологических свойств отражает некие важные черты наших феноменальных понятий. У нас нет независимого языка для описания феноменальных качеств. Мы видели, что в них есть что-то неуловимое. Хотя зеленоватость — особое ощущение с богатыми внутренними характеристиками, очень мало можно сказать о ней помимо того, что она зеленовата. Рассуждая о феноменальных качествах, мы обычно вынуждены конкретизировать их в терминах связанных с ними внешних свойств или же в терминах сопряженных с ними каузальных ролей. Язык, с помощью которого мы рассуждаем о феноменальных качествах, произведен от нашего нефеноменального языка. Как выразился Райл, не существует «чистых» слов об ощущении.

Если взглянуть на приведенный мной выше каталог сознательного опыта, то мы увидим, что соответствующие переживания никогда не описываются в терминах их внутренних качеств. Скорее я использовал такие выражения, как «аромат только что испеченного хлеба», «паттерны, возникающие, если закрыть глаза» и т. п. Даже если речь идет о таком термине, как «ощущение зеленого», его референт может быть по-настоящему зафиксирован лишь по отношению к чему-то внешнему. Ведь мы по-настоящему обретаем понимание выражения «ощущение зеленого» остенсивным путем — мы научаемся применять его к опытным данным, порождаемым травой, деревьями и т. д. В общем случае, в той мере, в какой мы вообще обладаем феноменальными категориями, о которых можно сообщить кому-то, они определяются отношением либо к их типичным внешним коррелятам, либо к связанным с ними психологическим состояниям. К примеру, если мы говорим о феноменальном качестве счастья, то референт слова «счастье» неявно фиксируется с помощью некоей каузальной роли — состояние, при котором человек рассуждает о том, что все хорошо, радостно прыгает и т. п. Возможно, так можно истолковать знаменные слова Витгенштейна: «Внутренний процесс нуждается во внешних критериях».

Эта зависимость феноменальных понятий от каузальных критериев подтолкнула некоторых авторов (в том числе Витгенштейна и Райла в какие-то моменты их размышлений) к предположению, что значение наших ментальных понятий сводится к связанным с ними каузальным критериям. Это предположение имеет видимость правдоподобия: если феноменальное свойство всегда вычленяется с помощью психологического, то почему бы не допустить, что речь идет об одном-единственном свойстве? Но с этим искушением надо бороться. Если мы говорим об ощущении зеленого, наши рассуждения не эквивалентны рассуждению о «состоянии, порождаемом травой, деревьями и т. д.».

Мы говорим о *феноменальном качестве*, обычно возникающем, когда некое состояние порождается травой и деревьями. Если здесь вообще можно провести какой-то каузальный анализ, то он будет примерно таким: «некое *феноменальное состояние*, порождаемое травой, деревьями и т. д.»¹³. Феноменальный элемент данного понятия не позволяет провести его анализ только в функциональных терминах.

В общем случае, если какое-то феноменальное свойство выделяется с помощью психологического свойства *P*, это феноменальное понятие не сводится к «*P*». Оно — «сознательный опыт, обычно сопровождающий *P*». И важно то, что само понятие «феноменального качества» или «сознательного опыта» не определяется в психологических терминах. Скорее, как мы видели ранее, понятие сознательного опыта есть нечто изначальное. Если бы был возможен функциональный анализ понятия опыта или феноменального качества, то он позволял бы функционально проанализировать конкретные феноменальные свойства, однако при отсутствии такого анализа мы не можем делать подобного вывода.

Мы не можем отождествлять понятие «феноменальное *P*» с понятием «психологическое *P*» по известным нам причинам: это два разных понятия, о чем свидетельствует факт наличия двух разных вещей, нуждающихся в объяснении. Хотя «феноменальное *P*» определяется как «опыт, обычно сопровождающий психологическое *P*», мы вполне можем вообразить ситуацию, в которой феноменальное *P* существует без психологического *P* и наоборот. Логотип «Роллс-Ройса» можно определить как логотип, обычно встречающийся на машинах марки «Роллс-Ройс», но это не значит, что быть таким логотипом — значит быть машиной марки «Роллс-Ройс».

Это не только позволяет нам ощутить бедность нашего специфически феноменального словаря в сравнении с нашим психологическим словарем, но и помогает понять, почему феноменальные и психологические свойства так часто смешивались. Для большинства повседневных дел это смешение не имеет большого значения: когда кто-то говорит, что он счастлив, он не обязательно имеет в виду либо феноменальное качество, либо функциональную роль — ведь они, как правило, существуют. Однако при решении философских задач, и в частности при попытках объяснения, смешение этих свойств оказывается фатальным. Этот шаг может быть заманчивым, так как разрушение границ между ними неожиданно делает проблему объяснения сознательного опыта на удивление прозрачной; но именно по этой причине он совершенно неудовлетворителен. От проблемы сознания нельзя отделаться одними лишь вербальными ухищрениями.

4. Две проблемы соотношения ментального и телесного

Следствием разделения свойств на феноменальные и психологические оказывается разделение проблемы соотношения ментального и телесного на две части: трудную часть и легкую часть. Психологические аспекты ментального ставят немало технических проблем перед когнитивной наукой и содержат много интересных философских головоломок, но они не заключают в себе каких-либо глубоких метафизических тайн. Вопрос «как физическая система могла бы быть способна к обучению или запоминанию?» не столь остр, как аналогичный вопрос об ощущениях или о сознании в целом. Причина этого ясна. Согласно нашему анализу в параграфе 3 настоящей главы, обучение и память — функциональные свойства, характеризуемые каузальными ролями, и поэтому вопрос «как физическая система могла бы обладать психологическим свойством P ?» тождествен вопросу «как некое состояние физической системы могло бы играть такую-то каузальную роль?». Это вопрос для наук о физических системах. Надо просто рассказать об организации физической системы, которая позволяет ей надлежащим образом реагировать на стимулы, идущие из окружающей среды, и продуцировать надлежащее поведение. Хотя технические проблемы, связанные с подобного рода объяснениями, громадны, их решение может быть найдено в рамках строго определенной исследовательской программы. Метафизические проблемы в этой сфере относительно немногочисленны.

Это не значит, что с психологическими свойствами не связано никаких философских затруднений. К примеру, корректный анализ этих понятий может представлять собой весьма значительную проблему. Даже если широко признано, что они являются функциональными понятиями, могут иметься серьезные разногласия в вопросе о том, как именно должен осуществляться требуемый для них функциональный анализ. Богатый материал для споров поставляют, к примеру, такие интенциональные свойства, как убеждение и желание. В частности, до сих пор плохо понято то, что же именно является содержанием конкретных интенциональных состояний. Технические проблемы сопряжены также и с тем, как именно подобные высокоуровневые конструкты могут играть реальную каузальную роль в продуцировании поведения, особенно если они частично конституируются свойствами окружающей среды или если не существует строгих законов, связывающих психологические состояния с поведением. В когнитивной науке мы сталкиваемся и с базовыми для нее частично эмпирическими проблемами относительно того, как эти свойства могли бы реализовываться в наличных когнитивных системах и реализованы ли они вообще.

Все эти проблемы серьезны, но они скорее головоломки, чем тайны. Ситуация здесь аналогична ситуации в философии биологии, где нет насущной проблемы соотношения жизненного и телесного начала, но есть множество технических проблем, касающихся эволюции, отбора, адаптации, приспособляемости и биологических видов. Подобно тому как давно рассеялось большинство мнимых метафизических тайн биологии, мы вправе сказать, что, по сути, рассеялась и проблема соотношения ментального и телесного для психологических свойств. Остался лишь набор не столь серьезных технических проблем, с которыми можно будет справиться в ходе обычного научного и философского анализа.

Совершенно иначе обстоит дело с феноменальными аспектами ментального. Здесь проблема соотношения ментального и телесного столь же тяжела, как и раньше. Впечатляющий прогресс физических и когнитивных наук не пролил большого света на вопрос о том, как и почему когнитивные функциональные процессы сопровождаются сознательным опытом. Прогресс в понимании ментального сводился почти исключительно к объяснению поведения. Он не коснулся вопроса о сознательном опыте.

При желании мы можем трактовать дистинкцию психологического и феноменального не столько в качестве разбиения проблемы соотношения ментального и телесного, сколько в качестве выделения в ней двух особых частей. Труднейшая часть проблемы соотношения ментального и телесного состоит в следующем вопросе: каким образом физическая система могла бы порождать сознательный опыт? Мы могли бы вычленить в звене, связующем физическое и сознательный опыт, две части: одной из них было бы звено между физическим и психологическим, другой — звено между психологическим и феноменальным. Ранее мы видели, что сейчас мы уже достаточно хорошо представляем, каким образом физическая система может обладать психологическими свойствами: *психологическая* проблема соотношения ментального и телесного была разрешена. Остается вопрос, почему и как эти психологические свойства сопровождаются феноменальными свойствами: почему, к примеру, все эти стимулы и реакции, связанные с болью, сопровождаются *переживанием боли*? Следуя Джекендоффу (Jackendoff 1987), мы можем назвать этот остаток *проблемой соотношения ментального и ментального*. Имеющиеся у нас физические объяснения не выводят нас за пределы психологического ментального. Но связь между психологическим и феноменальным ментальным остается плохо понятой¹⁴.

Не исключено, что звено, связующее феноменальное и физическое, могло бы оказаться независимым от звена, связующего психологическое и физическое. В таком случае подобная фрагментация была бы

невозможной. Но этот вариант кажется маловероятным. Отмеченная нами тесная корреляция между феноменальными и психологическими свойствами наводит на мысль о глубоком сочленении между ними. В ряде последующих глав я буду доказывать, что это сочленение оченьочно и что стратегия указанной фрагментации важна для подхода к решению проблемы соотношения ментального и телесного. И если это так, то уяснение связующего звена между психологическим и феноменальным имеет решающее значение для понимания самого сознательного опыта.

5. Два понятия сознания

Исходя из того, что очень многие ментальные термины имеют двойную природу, не удивительным будет узнать, что даже термин «сознание» имеет как феноменальный, так и психологический смысл. До сих пор я говорил о сознании в феноменальном смысле, вбирающем в себя все ранее упомянутые феноменальные аспекты ментального. Быть сознательным в этом смысле — значит реализовывать какое-то феноменальное качество. Это основной смысл «сознания» или, по крайней мере, тот смысл, с которым связаны главные объяснительные проблемы. Но это не единственный смысл данного термина. Он может использоваться также и для отсылки к множеству психологических свойств, таких как способность отчитаться о чем-то или иметь интроспективный доступ к информации. Мы можем сгруппировать психологические свойства такого рода под рубрикой *психологического сознания*, противопоставляемого *феноменальному сознанию*, на котором я до сих пор фокусировал свое внимание.

Эта двусмысленность может приводить к немалой путанице при обсуждении сознания. Нередко бывает так, что кто-то, претендующий на объяснение сознания, ставит соответствующую проблему со всей тяжестью проблемы феноменального сознания, но в итоге объясняет какой-либо из аспектов психологического сознания, к примеру способность к интроспекции. Подобное объяснение может быть ценным само по себе, но у нас остается чувство, что мы получили меньше, чем было обещано.

Разновидности психологического сознания

Существует множество психологических понятий, по отношению к которым подчас используется термин «сознание». Среди них:

Бодрствование. Иногда мы говорим, что кто-то находится в сознании, имея в виду, что он не спит. Логично предположить, что у нас имеются какие-то переживания во сне, так что это понятие явным образом

не совпадает с феноменальным сознанием. Вполне вероятно, что бодрствование может быть проанализировано в функциональных терминах — скажем, в первом приближении, в терминах способности обрабатывать информацию о мире и рациональным образом использовать ее.

Интроспекция. Это процесс, благодаря которому мы можем распознать содержательную сторону наших внутренних состояний. Если вы спрашиваете меня о моих ментальных состояниях, я найду ответ при помощи интроспекции. Этот доступ к своим ментальным состояниям — важный компонент обыденного понятия сознания, и, по крайней мере отчасти, здесь идет речь о функциональном понятии. Его можно было бы проанализировать в терминах рациональных процессов субъекта, обнаруживающих надлежащую чувствительность к информации, касающейся его внутренних состояний, и способности надлежащего использования последней.

Способность давать отчет. Имеется в виду способность отчитываться о содержании наших ментальных состояний. Она предполагает способность к интроспекции, но более ограничена, так как предполагает и языковую способность. Это понятие сознания зачастую было главным объектом философов и психологов операционистского толка.

Самосознание. Этот термин отсылает к нашей способности мыслить о самих себе, к нашей осведомленности о собственном существовании в качестве индивидов и о нашем отличии от других. Мое самосознание могло бы быть проанализировано в терминах моего доступа к моей самостной модели или же в терминах наличия у меня определенного рода представления, неким образом связанного со мной самим. Вполне может оказаться, что самосознание свойственно только людям и некоторым видам животных.

Внимание. Мы часто говорим, что некто осознает что-то, именно тогда, когда он обращает внимание на это что-то; иными словами, когда значительная часть его когнитивных ресурсов направлена на обработку соответствующей информации. Мы можем феноменально осознавать и не будучи внимательными, о чем свидетельствует периферическое зрение.

Произвольный контроль. Еще один смысл: мы говорим, что какой-то поведенческий *акт* сознателен, если он совершается намеренно, то есть если он надлежащим образом порождается некими предшествующими мыслями.

Знание. В другом обыденном смысле мы говорим, что некто знает какой-то факт, именно тогда, когда он знает этот факт, и что он сознает вещь — именно тогда, когда он знает о ней. Это понятие

редко оказывается в центре технических дискуссий о сознании, но похоже, что оно не менее значимо для обыденного использования этого термина, чем и все другое.

То, что все эти понятия по большей части функциональны, можно понять, взглянув на то, как можно было бы объяснить данные феномены. Если бы нам надо было попытаться объяснить внимание, можно было бы придумать модель когнитивных процессов, позволяющую концентрировать ресурсы на каком-то одном аспекте доступной информации. Если бы нам надо было попытаться объяснить интроспекцию, мы попробовали бы объяснить процессы, благодаря которым субъект оказывается надлежащим образом чувствителен к его внутренним состояниям. Подобного можно было бы ожидать и от объяснения других свойств. В каждом из этих случаев функциональное объяснение, похоже, схватывает самую суть дела.

Но хотя эти понятия имеют психологическое ядро, многие из них связаны с феноменальными состояниями. К примеру, с самосознанием связано определенного рода феноменальное состояние. Это же справедливо и для интроспекции, внимания и произвольного контроля за поведением. Подобно другим обсуждавшимся выше двуаспектным терминам, такие термины, как «интроспекция» и «самосознание», иногда используются для обозначения феноменальных состояний, что может приводить к путанице. В самом деле, кто-нибудь мог бы попытаться доказать, что некий процесс может считаться подлинной «интроспекцией», подлинным «вниманием» и т. п., лишь если ему свойствен феноменальный аспект. Но и в этом случае, как и раньше, спор по большей части идет о словах. Ясно, что с каждым из этих понятий сопряжены как феноменальные, так и психологические свойства. Тот, кто не хочет награждать психологическое свойство ментальным термином вроде «внимания», может использовать термин «псевдовнимание». Но как бы мы ни назвали эти свойства, сущностные философские проблемы останутся теми же самыми, что и прежде.

Феноменальные и психологические свойства, сопряженные с этими понятиями, обычно сосуществуют, но, как и в случае с другими ментальными понятиями, их нельзя смешивать. Мы не должны смешивать и феноменальные смыслы этих терминов с феноменальным сознанием вообще.

Сознание и осведомленность

Мы видели, что с эмоциональным опытом, опытом самосознания, ощущения и т. д. связаны какие-то психологические свойства. Естественно предположить, что какое-то психологическое свойство могло бы быть связано и с самим опытом, или феноменальным сознанием.

И я полагаю, что такое свойство действительно есть; мы можем назвать его «осведомленностью». Это наиболее общее обозначение психологического сознания.

Осведомленность может быть в общем виде определена как состояние, при котором мы обладаем доступом к определенной информации и можем использовать эту информацию для контроля за поведением. В частности, мы можем быть осведомлены о каком-то объекте в нашем окружении, о состоянии нашего тела и о наших ментальных состояниях. Осведомленность относительно какой-то информации, как правило, сопряжена со способностью осмысленно контролировать поведение в зависимости от этой информации. Ясно, что это функциональное понятие. В обыденной речи термин «осведомленность» часто используется как синоним «сознания», но я зарезервирую его для вышеписанного функционального понятия.

В целом, похоже, что при наличии феноменального сознания всегда имеет место осведомленность. Мой феноменальный опыт лежащей proximity желтой книги сопровождается моей функциональной осведомленностью об этой книге, а также осведомленностью о желтом цвете. Мое переживание боли сопровождается осведомленностью о наличии чего-то плохого, обычно приводящей к уклонению, если оно возможно, и т. п. Факт того, что любой сознательный опыт сопровождается осведомленностью, становится очевидным вследствие того факта, что мы можем *отчитываться* о сознательном опыте. Если я переживаю что-то, я могу говорить о факте этого переживания. Я могу не обращать на него внимания, но я, по крайней мере, способен сосредоточиться на нем и сказать о нем, если захочу. Из этой возможности давать отчет непосредственно следует моя осведомленность в интересующем нас смысле этого слова. Разумеется, животные и еще безгласные люди могли бы иметь сознательный опыт и при отсутствии способности отчитываться о нем, но вполне вероятно, что подобные существа все же обладали бы некоей степенью осведомленности. Из осведомленности не вытекает способность давать отчет, хотя они обычно совмещаются у тех существ, которые наделены речью.

Сознание всегда сопровождается осведомленностью, но осведомленность в указанном смысле не обязана сопровождаться сознанием. Можно, к примеру, быть осведомленным о факте без того, чтобы с этим был связан какой-то конкретный феноменальный опыт. Не исключено, однако, что можно так ограничить понятие осведомленности, чтобы его объем стал совпадать или почти совпадать с феноменальным сознанием. Пока я оставлю эту идею в стороне, но вернусь к ее обсуждению в главе 6.

Понятие осведомленности вбирает в себя большинство вышеперечисленных психологических понятий сознания или даже все эти

понятия. Интроспекцию можно определить как осведомленность о каком-то внутреннем состоянии. Внимание можно определить как очень высокую степень осведомленности об объекте или о событии. Само-сознание может быть определено как осведомленность о самом себе. С произвольным контролем дело обстоит не так просто, хотя его можно частично определить через внимание, обращаемое нами на то, как мы ведем себя. Бодрствование может быть схематично охарактеризовано как состояние, в котором мы способны в известной степени рационально взаимодействовать со своим окружением, что предполагает осведомленность определенного рода.

Идея о том, что существует функциональное понятие сознания, которое может быть эксплицировано в терминах доступа, была сформулирована Блоком (Block 1995), рассуждавшим о различии «феноменального сознания» и «сознания доступа». Блоковское понятие сознания доступа почти в точности соответствует описанному мной понятию осведомленности (подробнее я обсуждаю отношение между ними в главе 6). Детальное различие между «осведомленностью» и «сознанием» в сходном ключе проводит и Ньюэлл (Newell 1992). Он характеризует осведомленность как «способность субъекта обуславливать свое поведение неким знанием» и проговаривает различие между этим понятием и сознанием, которое, по его мнению, является нефункциональным феноменом. Похожие дистинкции проводились и другими философами и когнитивными учеными¹⁵.

Объяснение сознания и объяснение осведомленности

Подобно другим психологическим свойствам, осведомленность не порождает многих метафизических проблем. Проблемы, связанные с психологическими разновидностями сознания, не более трудны, чем те, что связаны с памятью, обучением и убеждением. Понятие осведомленности, разумеется, не является совершенно прозрачным, и для его прояснения может потребоваться серьезный философский анализ. Кроме того, здесь предстоит еще очень много сделать когнитивной науке, изучающей то, каким образом естественные и искусственные когнитивные системы могли бы функционировать таким образом, чтобы они оказывались в состоянии осведомленности. Но контуры этих исследовательских программ вполне ясны. И едва ли можно сомневаться в том, что эти задачи будут в конце концов решены в ходе естественного развития когнитивной науки, опирающейся на надлежащего рода философский анализ.

В той мере, в какой сознание является действительно трудной проблемой для науки о ментальном, речь идет прежде всего о феноменальном сознании. С ним связаны проблемы совсем другого порядка. Даже после того как мы объяснили физическое и вычислительное

функционирование сознательной системы, нам все еще предстоит объяснить, почему эта система обладает сознательным опытом. Конечно, некоторые оспаривают этот тезис, и вскоре я обсуджу его подробнее. Пока же можно просто обратить внимание на *prima facie* различие проблем, связанных с феноменальной и психологической разновидностями сознания. Именно феноменальное сознание заставляет беспокоиться о проблеме сознания.

Если психологическое и феноменальное понятия сознания различны, то, к сожалению, они нередко смешиваются в литературе. В повседневной речи это смешение не столь значимо, поскольку осведомленность и феноменальное сознание обычно сопутствуют друг другу. Но когда речь идет об объяснении, это концептуальное различие становится решающим. И если посмотреть на хотя бы отчасти убедительные объяснения «сознания», мы увидим, что обычно объясняется именно психологический аспект. Феноменальный аспект, как правило, остается без рассмотрения.

Многие недавние попытки философского анализа сознания были связаны главным образом с его нефеноменальными аспектами. Розенталь (Rosenthal 1996) доказывает, что ментальное состояние оказывается сознательным только при наличии высокоуровневой мысли об этом ментальном состоянии. Этот анализ мог бы быть полезным для интроспективного сознания, а возможно, и для других аспектов осведомленности, но он, похоже, не объясняет феноменальный опыт¹⁶. Аналогичным образом Деннет (Dennett 1991) посвящает большую часть своей книги очерчиванию детальной когнитивной модели, которую он преподносит как объяснение сознания. Представляется, однако, что эта модель является прежде всего моделью способности субъекта давать вербальный отчет о ментальных состояниях. Поэтому она могла бы служить для объяснения способности давать отчет, интроспективного сознания, а быть может, и других аспектов осведомленности. Однако она не дает никакого объяснения феноменального сознания (хотя Деннет и не согласился бы с такой оценкой).

Армстронг (Armstrong 1968), наталкиваясь на сознание как на препятствие на пути построения его функционалистской теории ментального, анализирует это понятие в терминах устройства, сканирующего самого себя. Эти шаги могли бы быть полезными для объяснения самосознания и интроспективного сознания, но они оставляют в стороне проблему феноменального опыта. В другой работе Армстронг говорит как о перцептивном сознании, так и об интроспективном сознании (Armstrong 1981), но они интересуют его лишь в качестве разновидностей осведомленности, и он не обсуждает проблемы, связанные с феноменальными качествами опыта. Поэтому он обходит стороной тот

смысл сознания, который действительно является *проблематичным* для его функционалистской теории, и это оказывается возможным благодаря неоднозначности понятия сознания.

Другие авторы, пишущие на тему «сознания», занимались главным образом самосознанием или интроспективным сознанием. Ван Гулик (van Gulick 1988), указывая, что сознание должно быть определено как обладание «рефлексивной метапсихологической информацией», в лучшем случае может претендовать на анализ этих психологических понятий и в действительности сам признает, что феноменальные аспекты могут остаться за его бортом. Сходным образом, детальная теория сознания Джейнис (Jaynes 1976) имеет дело только с нашей осведомленностью о наших мыслях. Она не касается феноменов, связанных с восприятием, и поэтому не может рассматриваться как теория осведомленности вообще, не говоря уже о теории феноменального сознания. Хофштадтер (Hofstadter 1979) высказывает ряд интересных наблюдений о сознании, но его больше интересует интроспекция, свободная воля и осмысление нашей самости, чем опыт *per se*.

Если говорить об обсуждении сознания психологами, то можно отметить, что их редко заботило тщательное разведение феноменальных и психологических понятий. Психологические исследования, как правило, посвящены какому-то из аспектов осведомленности, таких как интроспекция, внимание или самосознание. Даже психологические аспекты сознания имели дурную репутацию в психологии, по крайней мере до недавнего времени. Возможно, это связано с некоторой неясностью этих понятий, а также с трудностями, сопряженными с такими высокоуровневыми феноменами, как интроспекция. Можно предположить, что во многом эта дурная репутация была обязана самому термину «сознание», обозначающему также и феноменальное сознание. Использование этого общего имени создавало ощущение некоего соучастия в преступлении.

Иногда можно услышать, что в последние годы психологи в своих исследованиях «вновь обратились к сознанию». Реальная картина, как представляется, состоит в том, что предметом активных исследований стали психологические аспекты сознания и что исследователи перестали бояться использовать термин «сознание» для обозначения этих явлений. Однако феноменальное сознание по большей части остается в стороне. И это можно понять. Хотя очевидно, что методы экспериментальной психологии могли бы способствовать пониманию различных видов осведомленности, сложно понять, как они могли бы объяснить феноменальный опыт¹⁷.

Когнитивные модели хорошо приспособлены для объяснения психологических аспектов сознания. В идее того, что физическая система

должна быть способна к интроспективному постижению своих внутренних состояний или к рациональной обработке информации, идущей из ее окружения, или что она должна быть способна фокусировать свое внимание сначала на каком-то одном месте, а затем на другом, не заключается никакой великой метафизической проблемы. Вполне очевидно, что надлежащая функциональная теория сможет объяснить эти способности, даже если на отыскание подобной теории уйдут десятилетия или столетия. Действительно сложной проблемой является проблема феноменального сознания, но она-то и оставлялась без рассмотрения в тех объяснениях психологического сознания, которые выдвигались до настоящего времени.

В дальнейшем я вернусь к использованию термина «сознание» исключительно для обозначения феноменального сознания. Когда мне потребуется использовать психологические понятия, я буду говорить о «психологическом сознании» или об «осведомленности». И меня будет интересовать по большей части именно феноменальное сознание.

Глава 2

Супервентность и объяснение

Какое место занимает сознание в природе? Является ли сознание физическим? Можно ли объяснить сознание в физических терминах? Чтобы разобраться с этими вопросами, мы должны задать некие рамки для их рассмотрения; и в этой главе я задам их. Главным их элементом будет понятие *супервентности*: я раскрою это понятие и применю его для прояснения идеи редуктивного объяснения. Используя эту концепцию, я нарисую картину отношения, существующего между большинством высокоуровневых феноменов и физическими фактами, которая, похоже, будет охватывать собой всё, за исключением, возможно, сознательного опыта.

1. Супервентность

Широко распространено представление, что большинство фундаментальных фактов относительно нашего мира — это физические факты и что все остальные факты зависят от них. Если брать слово «зависят» в достаточно слабом смысле, это может быть чуть ли не тривиальной истиной; однако если брать его в сильном смысле, тезис становится спорным. В целом существует большое множество отношений зависимости между высокоуровневыми и низкоуровневыми фактами, и тот тип отношений зависимости, который имеет место в какой-то одной области, к примеру в биологии, может отсутствовать в другой, к примеру в сфере сознательного опыта. Философское понятие супервентности создает унифицированные рамки, в которых могут обсуждаться эти отношения зависимости.

Понятие супервентности формализует интуитивную идею о том, что одно множество фактов может полностью определять другое множество фактов¹. К примеру, физические факты о мире, по-видимому, определяют биологические факты в том плане, что, если зафиксировать все физические факты о мире, биологические факты уже не смогут меняться. (Фиксация всех физических фактов одновременно фиксирует, какие объекты являются живыми.) Из этого становится более или

менее понятен смысл, в котором биологические свойства супервентны на физических. В общем, супервентность — это отношение между двумя множествами свойств: *B*-свойствами — интуитивно, *высокоуровневыми* свойствами и *A*-свойствами — более фундаментальными, *низкоуровневыми* свойствами.

В нашем контексте *A*-свойства обычно будут физическими свойствами, точнее, фундаментальными свойствами, которые были бы выявлены совершенной физической теорией. Среди них, возможно, окажется масса, заряд, положение в пространстве и времени; свойства, характеризующие распределение различных пространственно-временных полей, проявления различных сил, форму различных волн и т. п. Конкретные параметры этих свойств не имеют большого значения. Если физика претерпит радикальные изменения, не исключено, что соответствующий класс свойств будет сильно отличаться от упомянутого мной, но это не затронет аргументы, о которых пойдет речь. Я исключу из рассмотрения такие высокоуровневые свойства, как сочность, комковатость, жирафовость и т. п., хотя в известном смысле они и являются физическими. В дальнейшем рассуждения о физических свойствах будут неявным образом ограничены классом фундаментальных свойств, если иное не будет оговорено мной. Для большей ясности я иногда буду говорить о «микрофизических» или «низкоуровневых физических» свойствах.

A-факты и *B*-факты о мире — это факты, связанные с реализацией и распределением *A*-свойств и *B*-свойств². Так что физические факты о мире охватывают все факты, связанные с реализацией физических свойств в пространственно-временном многообразии. Полезно также условиться, что физические факты о мире включают его базовые физические законы. Согласно некоторым концепциям, эти законы детерминированы совокупностью отдельных физических факторов, но это нельзя принимать за аксиому.

Основой для дефиниции супервентности является следующая формула:

B-свойства *супервентны* на *A*-свойствах, если невозможны две ситуации, тождественные в плане *A*-свойств, но различные в их *B*-свойствах.

К примеру, биологические свойства супервентны на физических свойствах, если любые две возможные физически тождественные ситуации биологически тождественны. (Здесь я использую слово «тождественные» в смысле неразличимости, а не нумерического тождества. В этом смысле два разных стола могут быть физически тождественными.) Более определенные понятия супервентности могут быть получены при конкретизации этой основной формулы. В зависимости от того, понимаем ли мы под «ситуациями», о которых идет речь,

индивидуов или целые миры, мы приходим к понятиям соответственно локальной и глобальной супервентности. И в зависимости от того, как мы конструируем понятие возможности, мы получаем понятия логической супервентности, естественной супервентности, а возможно, и другие. Сейчас я разъясню эти дистинкции.

Локальная и глобальная супервентность

В-свойства локально супервентны на А-свойствах, если А-свойства индивида детерминируют В-свойства этого индивида, то есть если два любых возможных индивида, реализующих те же самые А-свойства, реализуют и те же самые В-свойства. К примеру, форма локально супервентна на физических свойствах: любые два объекта с одинаковыми физическими свойствами с необходимостью имеют одну и ту же форму. А вот ценность не является локально супервентной на физических свойствах: точная физическая копия «Моны Лизы» не столь ценна, как оригинал «Моны Лизы». В целом локальная супервентность свойства на физическом отсутствует, если это свойство является зависимым от контекста, то есть если обладание объектом этим свойством зависит не только от физического устройства данного объекта, но также от его окружения и истории. Оригинал «Моны Лизы» более ценен, чем его копия, из-за различия их исторического контекста: оригинал этой картины был создан Леонардо, а копия — кем-то другим³.

Глобально же В-свойства оказываются супервентными на А-свойствах в том случае, когда А-факты о *мире* в целом детерминируют В-факты, то есть если не существует двух возможных миров, тождественных относительно их А-свойств, но различных в плане их В-свойств⁴. Мир здесь понимается как весь универсум в целом; различные возможные миры соответствуют тому, каким мог бы оказаться универсум.

Локальная супервентность влечет глобальную супервентность, но не наоборот. К примеру, вполне вероятно, что биологические свойства глобально супервентны на физических свойствах в том смысле, что любой мир, физически тождественный нашему, был бы также и биологически тождествен ему. (Здесь надо сделать небольшую оговорку, о которой я вскоре скажу.) Но они, по-видимому, не являются локально супервентными. Два физически тождественных организма, судя по всему, могут отличаться по некоторым биологическим характеристикам. К примеру, один из них может быть *более приспособленным*, чем другой, вследствие различий в окружающей их среде. Можно даже вообразить себе ситуацию, при которой физически тождественные организмы оказывались бы представителями разных видов — при наличии у них разной эволюционной истории.

Различие между глобальной и локальной супервентностью не слишком принципиально, когда мы говорим о сознательном опыте, так как вполне вероятно, что в той мере, в какой сознание вообще супервентно на физическом, оно является локально супервентным на нем. Если два существа физически тождественны, то различия в их окружении и истории не помешают им иметь тождественный опыт. Разумеется, контекст может косвенно влиять на опыт, но только через воздействие на внутреннюю структуру, как это имеет место в случае восприятия. Такие феномены, как галлюцинации и иллюзии, иллюстрируют тот факт, что *прямую* ответственность за опыт несет внутренняя структура, а не контекст.

Более важно для нас различие между *логической* (или концептуальной) и *естественной* (номической или эмпирической) супервентностью.

**Логическая
и естественная
супервентность**

В-свойства логически супервентны на А-свойствах, если не существует двух логически возможных ситуаций, тождественных относительно их А-свойств, которые были бы различны относительно их В-свойств. Далее в этой главе я еще поговорю о логической возможности. Пока же можно в общем помыслить ее как возможность в самом широком смысле, приблизительно соответствующей представимости, не ограниченной законами нашего мира. Полезно мыслить логически возможный мир как мир, который был бы в силах при желании создать Бог (гипотетически)⁵. Бог не мог бы создать мир с [женатыми холостяками], но он мог бы создать мир с летающими телефонами. При определении логической возможности истинности какого-то утверждения ограничения имеют по большей части концептуальный характер. Понятие [женатого холостяка] содержит противоречие, и поэтому [женатые холостяки] логически невозможны; а вот понятие летающего телефона концептуально когерентно, пусть и несколько необычно, и поэтому летающие телефоны являются логически возможными.

Следует подчеркнуть, что логическая супервентность не определяется в терминах выводимости в рамках какой-либо системы формальной логики. Скорее, логическая супервентность определяется в терминах логически возможных миров (и индивидов), где понятие логически возможного мира оказывается независимым от этих формальных соображений. Подобная возможность нередко именуется в философской литературе «широкой логической» возможностью в противоположность «узкой логической» возможности, которая определяется формальными системами⁶.

На глобальном уровне биологические свойства логически супервентны на физических свойствах. Даже Бог не мог бы создать мир,

физически идентичный, но биологически отличный от нашего. Тут просто нет логического места для независимого изменения биологических фактов. Фиксируя все физические факты о мире — включая факты о распределении всех частиц в пространстве и времени — мы в результате также фиксируем макроскопическую форму всех объектов в мире, то, как они перемещаются и функционируют, и то, как они физически взаимодействуют друг с другом. Если в этом мире есть живой кенгуру, то в *любом* мире, физически тождественном с данным миром, будет существовать физически идентичный кенгуру, и этот кенгуру автоматически будет живым кенгуру.

Можно вообразить, что некое гипотетическое высшее существо — к примеру, демон Лапласа, знающий расположение всех частиц в универсуме, — смог бы непосредственно «вычитать» все биологические факты из данного ему набора всех микрофизических фактов. Микрофизических фактов будет достаточно, чтобы подобное существо построило модель микроскопической структуры и динамики мира в пространстве и времени, из которой оно может непосредственно вывести макроскопическую структуру и динамику. При наличии всей этой информации у него будет все необходимое для определения того, какая из систем является живой, какая из систем относится к тому же самому виду и т. п. Если оно обладает биологическими понятиями и имеет полную спецификацию микрофизических фактов, ему не потребуется никакой дополнительной информации.

В общем, если В-свойства логически супервентны на А-свойствах, то мы можем сказать, что А-факты *влекут* В-факты, где один факт влечет другой, если логически невозможным является наличие одного факта без другого. В таких случаях демон Лапласа мог бы вычитывать В-факты из спецификации А-фактов, при условии наличия у него В-понятий. (Далее в этой главе я гораздо подробнее остановлюсь на связи между этими различными способами понимания логической супервентности; данное обсуждение имеет по большей части иллюстративный характер.) В известном смысле можно сказать, что при логической супервентности *всё*, что нужно для наличия В-фактов, — это наличие таких же А-фактов.

Супервентность может иметь место, однако, и без логической супервентности. Более слабая разновидность супервентности обнаруживается при систематической и совершенной *корреляции* двух множеств свойств в природе. К примеру, давление, оказываемое одним молем газа, систематическим образом зависит от его температуры и объема в соответствии с законом

$$pV = KT,$$

где K — константа. (В иллюстративных целях я допускаю, что все газы являются идеальными газами.) В актуальном мире данная температура и объем моля газа определяют его давление: эмпирически невозможна ситуация, при которой два различных моля газа имели бы одинаковую температуру и объем, но оказывали разное давление. Из этого следует, что давление моля газа в известном смысле супервентно на его температуре и объеме. (В этом примере класс А-свойств берется мной как нечто гораздо более узкое, чем класс физических свойств, по причинам, которые вскоре прояснятся.) Но эта супервентность не столь сильна, как логическая супервентность. Логически возможна ситуация, при которой моль газа, имеющий определенную температуру и объем, оказывал бы другое давление; вообразим, к примеру, мир с большей или меньшей константой K . Эта корреляция является, скорее, неким *естественным* фактом.

Это пример *естественной* супервентности одного свойства на других. Здесь давление естественно супервентно на температуре, объеме и свойстве быть молем газа. В общем, В-свойства оказываются естественно супервентными на А-свойствах, если в любых *естественно возможных* ситуациях при одинаковых А-свойствах имеются одинаковые В-свойства.

Естественно возможная ситуация — это ситуация, которая может реально возникать в природе без нарушения каких-либо законов природы. Это гораздо более сильный ограничитель, чем простая логическая возможность. Сценарий с различными константами газа, к примеру, логически возможен, но он не мог бы быть реализован в реальном мире, и поэтому он не является естественно возможным. Если говорить о естественно возможных ситуациях, то любые два моля газа с одинаковой температурой и одного и того же объема будут оказывать одинаковое давление.

Интуитивно, естественная возможность соответствует тому, что мы считаем реальной *эмпирической* возможностью: естественно возможная ситуация — это ситуация, которая при надлежащих условиях могла бы возникнуть в реальном мире. К ним относятся не только актуальные ситуации, но и контрафактические ситуации, которые могли бы возникнуть в истории мира при наличии других граничных условий или которые могли бы возникнуть в будущем при определенном стечении обстоятельств. Скажем, небоскреб высотой в милю почти наверняка естественно возможен, даже если в реальном мире он еще и не был построен. Естественно возможно (хотя и крайне маловероятно) даже, что обезьяна наберет текст «Гамлета». Естественно возможную ситуацию можно также представлять как ситуацию, которая соответствует природным законам нашего мира⁷. Поэтому естественную возмож-

ность иногда называют *номической*, или *номологической*, возможностью⁸ от греческого слова *nomos* — «закон».

Имеется громадное множество логически возможных ситуаций, которые не являются естественно возможными. К этому классу относится любая ситуация, в которой происходит нарушение природных законов нашего мира. Пример: универсум без гравитации или с другими значениями фундаментальных констант. В научной фантастике можно найти немало примеров подобных ситуаций, связанных с анти-гравитационными устройствами, вечными двигателями и т. п. Их легко вообразить, но можно быть почти уверенным, что их не могло бы существовать в нашем мире.

Если говорить об обратном отношении, то следует отметить, что любая естественно возможная ситуация будет логически возможной. Класс естественных возможностей является, таким образом, подклассом логических возможностей. Это различие можно проиллюстрировать следующим образом: как кубическая миля золота, так и кубическая миля урана-235 кажутся логически возможными, но, насколько нам известно, только первая является естественно возможной — (стабильная) кубическая миля урана-235 не могла бы существовать в нашем мире.

Естественная супервентность имеет место, когда во всех естественно возможных ситуациях ситуации с одинаковым распределением А-свойств имеют одинаковое распределение В-свойств, то есть когда А-факты о ситуации *естественным образом вынуждают* В-факты. Это происходит, если одинаковые кластеры А-свойств в нашем мире всегда сопровождаются одинаковыми В-свойствами и если эта корреляция является не случайной, а закономерной, то есть если реализация А-свойств будет всегда порождать В-свойства, когда бы и где бы это ни происходило. (Говоря философским языком, зависимость должна поддерживать контрфактуалы.) Это сопутствование не обязано иметься в любой логически возможной ситуации, но оно должно быть в любой естественно возможной ситуации.

Ясно, что логическая супервентность влечет за собой естественную супервентность. Если любые две логически возможные ситуации с одинаковыми А-свойствами имеют одинаковые В-свойства, то это будет верным и для любых двух естественно возможных ситуаций. Обратное же, однако, не проходит, как показывает закон состояния газа. Температура и объем моля газа определяют давление в естественно возможных, но не в логически возможных ситуациях, так что давление естественно, но не логически зависит от температуры и объема. Когда мы будем сталкиваться с естественной супервентностью без логической супервентности, я буду говорить, что мы имеем дело с сугубо естественной супервентностью.

По причинам, которые еще проясняются, примеры естественной супервентности на множестве *физических* свойств, которые не были бы одновременно примерами логической супервентности, найти непросто, но хорошей иллюстрацией может быть само сознание. Кажется очень вероятным, что сознание естественным образом, локально или глобально, супервентно на физических свойствах, если в естественном мире два любых физически идентичных существа будут иметь квалитативно тождественный опыт. Но далеко не очевидно, что сознание логически супервентно на физических свойствах. И по крайней мере, многие считают логически возможным, что существо, физически тождественное сознательному существу, могло бы быть вовсе лишено сознательного опыта, или что оно могло бы иметь иной сознательный опыт. (Некоторые оспаривают это, но пока я говорю об этом только для иллюстрации.) Если это так, то сознательный опыт естественно, но не логически, супервентен на физическом. Необходимая связь между физической структурой и опытом обеспечивается только законами природы, а не логической или концептуальной силой.

Различение логической и естественной супервентности очень важно для наших целей⁹. Мы можем интуитивно схватить его следующим образом. Если В-свойства логически супервентны на А-свойствах, то при (гипотетическом) создании Богом мира с определенными А-фактами В-факты оказываются бесплатным и автоматическим следствием этого. Если же В-свойства всего лишь естественно супервентны на А-свойствах, то после завершения дел с А-фактами Бог должен еще поработать над В-фактами: он должен обеспечить существование закона, соотносящего А-факты и В-факты. (Я заимствую этот образ у Крипке (Kripke 1972).) При наличии этого закона соответствующие А-факты будут автоматически порождать В-факты; но в принципе они могли бы и не делать этого.

Иногда говорят также о *метафизической* супервентности, основанной ни на логической, ни на естественной необходимости, а на «необходимости tout court», или «метафизической необходимости», как ее иногда называют (отталкиваясь от обсуждения Крипке (Kripke 1972) *апостериорной* необходимости). Позже я попытаюсь показать, что метафизически возможные миры — не что иное, как логически возможные миры (и что метафизическая возможность тех или иных утверждений — это логическая возможность с апостериорным семантическим оттенком), но пока мы вполне можем допустить понятие метафизической супервентности, определяемое по аналогии с вышеуказанными понятиями логической и естественной супервентности. Иногда заходит речь и о понятии «слабой» супервентности, но оно представляется слишком слабым для того, чтобы выражать какое бы то ни было интересное отношение между свойствами¹⁰.

Различие логической и естественной супервентности пересекается с различием глобальной и локальной супервентности. Мы вправе говорить как о глобальной логической супервентности, так и о локальной логической супервентности, хотя меня чаще будет интересовать первая. Когда я говорю о логической супервентности без дополнительных уточнений, я имею в виду глобальную логическую супервентность. Допустимы и рассуждения о глобальной естественной супервентности и локальной естественной супервентности, хотя отношения естественной супервентности, с которыми мы имеем дело, как правило, локальны или хотя бы локализуемы по той простой причине, что свидетельства, касающиеся отношения естественной супервентности, обычно сводятся к локальным регулярностям, имеющим место между теми или иными кластерами свойств¹¹.

Проблема, связанная с логической супервентностью*

Нам надо рассмотреть одну техническую проблему, связанную с понятием логической супервентности. Эта проблема возникает в силу логической возможности мира, физи-

чески идентичного нашему, но имеющего в своем составе добавочный нефизический материал, отсутствующий в нашем мире: к примеру, ангелов, эктоплазму или привидения. Можно представить мир в точности как наш, за исключением того, что в нем есть еще и ангелы, порхающие в нефизической сфере из эктоплазмы. Если бы эти ангелы размножались и эволюционировали, то они могли бы обладать собственными биологическими свойствами. У них вполне могли бы иметься разного рода убеждения и развитые социальные структуры.

Проблема, которую ставят эти примеры, очевидна. Мир с ангелами физически идентичен нашему, но биологически отличен от него. Если мир с ангелами логически возможен, то, в соответствии с нашей дефиницией, биологические свойства не являются супервентными на физических свойствах. Но мы, несомненно, хотим сказать, что биологические свойства супервентны на физических свойствах, если и не в мире с ангелами, то по крайней мере в *этом* мире (допуская, что в актуальном мире нет ангелов!). Интуитивно кажется нежелательным то, что одна лишь логическая возможность мира с ангелами будет препятствовать детерминации биологических свойств физическими свойствами нашего собственного мира.

Подобная проблема подтолкнула некоторых авторов (напр., Hugeland 1982, Petrie 1987) к предположению, что логическая возможность и необходимость слишком сильны для обслуживания соответствующих им возможности и необходимости, имеющихся в отношениях супервентности, и что вместо них надо использовать их более слабые разновидности, такие как естественная возможность и естественная

необходимость. Но это обесценило бы очень полезное различение логической и естественной супервентности, о котором шла речь выше, и привело бы к игнорированию того факта, что в определенном и вполне реальном смысле биологические факты, касающиеся нашего мира, логически определяются физическими фактами. Другие авторы (напр., Teller 1989) решались сказать, что, несмотря на видимость противоположного, миры с добавочным нефизическим материалом не являются логически или метафизически возможными — но это приводит к тому, что логическая и метафизическая возможность начинает казаться чем-то весьма и весьма произвольным. К счастью, в подобных шагах нет необходимости. Вполне реально сохранить полезное понятие логической супервентности, совместимое с возможностью подобных миров, — при надлежащем уточнении соответствующей дефиниции¹².

Ключ к решению в том, чтобы трактовать супервентность как утверждение о *нашем* мире (или, в более общем плане, о некоторых мирах). Это отдает должное интуиции, согласно которой биологические факты логически детерминируются физическими фактами в нашем мире, несмотря на существование необычных миров, где этого не происходит. В соответствии с уточненной дефиницией, В-свойства оказываются логически супервентными на А-свойствах, если В-свойства в нашем мире логически детерминируются А-свойствами в следующем смысле: в любом возможном мире с теми же А-фактами будут иметься и те же В-факты¹³. Наличие возможных миров с *добавочными* В-фактами, таким образом, не будет свидетельствовать против логической супервентности в нашем мире, если *по крайней мере* В-факты, истинные в нашем мире, будут истинными и во всех физически идентичных мирах. Так, в общем, и будет (за исключением тех случаев, которые будут обсуждаться ниже). Если в нашем мире какой-то коала жует эвкалиптовые листья, то в любом физически идентичном мире будет существовать идентичный коала, жующий эвкалиптовые листья, вне зависимости от того, будут ли там над ним виться ангелы или нет.

Тут есть небольшое затруднение. Биологические факты определенного рода, касающиеся нашего мира, не являются фактами в мире с ангелами: к примеру, факт того, что в нашем мире нет живой эктоплазмы, или тот факт, что все живые организмы основаны на ДНК. Не исключено даже, что мир с ангелами мог бы быть устроен так, что его эктоплазма была бы казуально зависима от физических процессов, и совокупление вомбатов на физическом уровне иногда приводило бы к рождению эктоплазмических малышей-вомбатов в нефизической реальности. В таком случае в нашем мире мог бы существовать бездетный (в определенном смысле) вомbat, а его двойник в физически идентичном мире с ангелами не был бы бездетным. Из этого следует, что

свойство бездействия, в соответствии с нашей дефиницией, не является супервентным, равно как и свойства мирового масштаба, такие как отсутствие живой эктоплазмы. Не все факты относительно нашего мира вытекают из одних лишь физических фактов.

Чтобы разобраться с этой проблемой, отметим, что все эти факты заключают в себе негативные экзистенциальные положения и поэтому зависят не только от того, что происходит в нашем мире, но и от того, что в нем не происходит. Мы не можем ожидать, что эти факты будут определяться какими-либо локализованными фактами, поскольку они зависят не только от локальных событий в мире, но и от ограничений этого мира. Тезисы о супервентности должны применяться только к *позитивным* фактам и свойствам, таким, которые не могут быть подвергнуты отрицанию вследствие простого расширения мира. Мы можем определить позитивный факт в *W* как такой факт, который имеется в любом мире, содержащем *W* в качестве своей собственной части¹⁴; позитивное же свойство есть такое свойство, которое, будучи реализованным в мире *W*, реализуется соответствующим индивидом также и во всех мирах, содержащих *W* в качестве их собственной части¹⁵. Большинство обыденных фактов и свойств позитивны — такие, к примеру, как свойство быть кенгуру, быть высотой в шесть футов или иметь ребенка. Негативные факты и свойства всегда так или иначе будут содержать негативные экзистенциальные положения. Они включают эксплицитно негативные экзистенциальные факты, такие как несуществование эктоплазмы, всеобще квантифицированные факты, такие как факт, что все живые существа созданы из ДНК, негативные реляционные свойства, такие как бездействие, и суперлативы вроде свойства быть самым плодовитым из всех существующих организмов.

Отношения супервентности, которыми мы будем заниматься, должны мыслиться ограниченными позитивными фактами и свойствами. Тезис, что биологические свойства супервентны на физических свойствах, относится только к позитивным биологическим свойствам. Поскольку все интересующие нас свойства позитивны — к примеру, локальные физические и феноменальные свойства, — это не столь уж сильное ограничение.

Таким образом, дефиниция глобальной логической супервентности *B*-свойств на *A*-свойствах сводится к следующему: в любом логически возможном мире *W*, *A*-неотличимом от нашего мира, *B*-факты о нашем мире имеются и в *W*. Нам необязательно делать оговорку по поводу позитивности, но обычно будет подразумеваться, что релевантными *B*-фактами и свойствами являются лишь позитивные факты и свойства. Аналогичным образом, *B*-свойства локально и логически супервентны на *A*-свойствах, если для каждого актуального индивида *x*

и для каждого логически возможного индивида y верно, что если y А-неотличим от x , то В-свойства, реализованные посредством x , реализованы посредством y . Более кратко и общо: В-свойства логически супервентны на А-свойствах, если В-факты, касающиеся актуальных ситуаций, вытекают из А-фактов, где под ситуациями имеются в виду миры и индивиды соответственно в глобальных и локальных случаях. Эта дефиниция схватывает идею о том, что тезисы о супервентности, как правило, являются тезисами о нашем мире, сохраняя ключевую роль за логической необходимостью¹⁶.

Логическая и естественная супервентность имеют весьма различные последствия для онтологии, то есть в вопросе о том, что именно существует в мире. Если В-свойства логически супервентны на А-свойствах, то в определенном смысле при данности А-фактов В-факты достаются даром. Если Бог (гипотетически) обеспечил все физические факты нашего мира, то биологические факты автоматически следуют за ними. В-факты просто по-новому описывают то, что уже было описано А-фактами. Они могут быть *другими* фактами (факт о слонах — не микрофизический факт), но они не являются фактами, *дополнительными* к тем, что имеются до них.

Супервентность и материализм

В случае сугубо естественной супервентности онтология оказывается не столь прямолинейной. Контингентные законосообразные связи связывают различные черты мира. В общем, если В-свойства лишь естественно супервентны на А-свойствах в нашем мире, то *мог бы* существовать мир, в котором наши А-факты не сопровождались В-фактами. Как мы видели раньше, после того как Бог зафиксировал все А-факты, он должен был бы поработать еще, чтобы зафиксировать В-факты. В-факты дополнительны к А-фактам, и их наличие означает появление в мире чего-то нового.

Помня об этом, мы можем дать точную дефиницию широко распространенного учения *материализма* (или *физикализа*ма), согласно утверждениям которого, как обычно считается, все в мире имеет физический характер, или в нем нет ничего сверх физического, или физические факты в известном смысле исчерпывают все факты относительно этого мира. На нашем языке, материализм верен, если все позитивные факты об этом мире глобально логически супервентны на физических фактах. Это отдает должное интуитивной идеи о том, что если материализм верен, то фиксация Богом всех физических фактов об этом мире фиксирует все факты о нем.

(Или хотя бы фиксирует все позитивные факты. Ограничение позитивными фактами требуется для гарантии того, чтобы миры с добавочными эктоплазмическими фактами не препятствовали матери-

ализму в нашем мире. Негативные экзистенциальные факты, такие как «ангелов не существует», не являются строго логически супервентными на физическом, но их несупервентность вполне совместима с материализмом. В известном смысле для фиксации негативных фактов Бог должен не только зафиксировать физические факты, но также провозгласить: «И это все». При желании мы могли бы добавить второ-порядковый факт «и это все» к супервентностной основе в дефиниции материализма; в этом случае можно было бы избавиться от ограничите-ля, указующего на позитивные факты.)

Согласно этой дефиниции, материализм верен, если все позитив-ные факты, касающиеся нашего мира, выводимы из физических фак-тов¹⁷. Иными словами, материализм верен, если в любом логически возможном мире W , физически неотличимом от нашего мира, все по-зитивные факты, имеющиеся в нашем мире, имеются также и в W . Это, в свою очередь, эквивалентно тезису, что любой мир, физически не-отличимый от нашего, содержит копию нашего мира в качестве своей (собственной или несобственной) части, — эта дефиниция интуитивно кажется верной¹⁸. (Сказанное соответствует дефиниции физикализма, данной Джексоном (Jackson 1994), критерий которого состоит в том, чтобы всякая минимальная физическая копия нашего мира являлась копией *simpliciter* нашего мира¹⁹.)

В главе 4 я обсужу эти вопросы гораздо более подробно, и там мы получим новое подкрепление этой дефиниции материализма. Кто-то мог бы возразить против использования идеи логической возможно-сти, а не возможности *tout court*, или «метафизической возможности». Эти люди могут заменить в данной выше дефиниции логическую воз-можность на метафизическую возможность. Позже я покажу, что это одно и то же.

2. Редуктивное объяснение

Выдающийся прогресс науки на протяжении последних несколь-ких столетий дал нам серьезное основание считать, что в мире имеется лишь очень немного глубоких тайн. Кажется, что практически любой естественный феномен, возвышающийся над уровнем физики мик-ромира, в принципе можно объяснить редуктивным способом, то есть объяснить его исключительно в терминах более простых сущностей. В этих случаях, при наличии надлежащей характеристики низкоуров-невых процессов, мы получаем объяснение и соответствующего высо-коуровневого феномена.

Хорошей иллюстрацией этого являются биологические феномены. Размножение можно объяснить через описание генетических и кле-

точных механизмов, позволяющих организмам порождать другие организмы. Адаптацию можно объяснить через описание механизмов, приводящих к надлежащим изменениям внешней функции в ответ на стимулы, идущие из окружающей среды. Сама жизнь объясняется объяснением различных механизмов, вызывающих размножение, адаптацию и т. п. Как только наши низкоуровневые истории становятся в достаточной степени детализированными, ощущение глубокой тайны исчезает: феномены, которые должны были быть объяснены, получают свое объяснение.

Сказанное применимо к большинству естественных феноменов. В физике мы объясняем тепло, рассуждая об энергии и возбуждении молекул. В астрономии мы объясняем фазы луны, детализируя процессы орбитального движения и отражения света. В геофизике землетрясения объясняются описанием взаимодействия подземных масс. В когнитивной науке для объяснения такого феномена, как обучение, нам, в первом приближении, надо лишь объяснить различные функциональные механизмы — механизмы, порождающие надлежащие изменения в поведении в ответ на стимулы из окружающей среды (если отвлечься от проблем, связанных с *переживанием* процесса обучения). Многие детали подобных объяснений в настоящее время ускользают от нас и, скорее всего, окажутся весьма запутанными, но мы знаем, что при достаточной проработке низкоуровневых концепций мы в итоге поймем, что происходит и на высоком уровне.

Пока я не буду давать точного определения понятия редуктивного объяснения; его характеристики останутся данными на уровне примеров. Однако я могу сделать ряд пояснений относительно того, чем не является редуктивное объяснение. Редуктивное объяснение феномена не обязано требовать *редукции* данного феномена, по крайней мере в некоторых смыслах этого неоднозначного термина. В известном смысле феномены, которые могут быть реализованы на многих различных физических носителях, к примеру обучение, могли бы и не быть редуцируемыми в том плане, что мы не можем *отождествить* обучение с каким-либо конкретным низкоуровневым феноменом. Но эта множественная реализуемость не препятствует редуктивному *объяснению* любых примеров обучения в терминах низкоуровневых феноменов²⁰. Нельзя также смешивать редуктивное объяснение феномена с редуцированием высокоуровневой *теории*. Редуктивное объяснение феномена иногда будет сопровождаться редуцированием существовавшей до этого высокоуровневой теории, однако в других случаях оно будет показывать ошибочность подобных теорий. А зачастую в таких случаях мы вообще не нашли бы высокоуровневых теорий.

Редуктивное объяснение не исчерпывает и не завершает всякое объяснение. Имеется множество других видов объяснения, и некоторые

из них могут лучше высвечивать феномен, чем редуктивное объяснение. Бывают, к примеру, *исторические* объяснения, объясняющие генезис таких феноменов, как жизнь, — редуктивное объяснение дает лишь синхронистский отчет о функционировании живых систем. Существуют также разнообразные *высокоуровневые* объяснения вроде объяснения различных аспектов поведения в терминах убеждений и желаний. Хотя в принципе можно было бы редуктивно объяснить это поведение, высокоуровневые объяснения нередко оказываются более полными и ясными. Не надо думать, что редуктивные объяснения должны вытеснить эти другие виды объяснений. У каждого из них есть свое собственное место.

Редуктивное объяснение через функциональный анализ

Что позволяет редуктивно объяснить такие разные феномены, как размножение, обучение и тепло? Во всех этих случаях ключевое значение имеет природа понятий, требующихся для характеристики этих феноменов. Если бы кто-то стал возражать против клеточного объяснения размножения, сказав: «Это объясняет, каким образом процессы в клетках могут приводить к рождению сложного физического объекта, сходного с изначальным объектом, но не объясняет *размножение*», мы были бы возмущены: ведь именно это и *означает* «размножение». В общем, редуктивное объяснение феномена сопровождается схематичным *анализом* данного феномена, имплицитным или эксплицитным. Понятие размножения может быть схематично проанализировано в терминах способности организма определенным образом порождать другой организм. Из этого следует, что, объяснив процессы, благодаря которым один организм порождает другой, мы объясняем этот случай размножения.

Этот момент может показаться тривиальным, но возможность подобного анализа фундирует возможность редуктивного объяснения в целом. Без такого анализа мы не смогли бы построить объяснительный мост от низкоуровневых физических фактов к интересующему нас феномену. При наличии же упомянутого анализа нам остается лишь показать, каким образом определенные низкоуровневые физические механизмы соответствуют этому анализу, — и в итоге мы получаем то или иное объяснение.

В случае самых интересных феноменов, требующих объяснения, включая такие феномены, как размножение и обучение, соответствующие понятия обычно допускают *функциональный* анализ. Ядро подобных понятий может быть охарактеризовано в терминах осуществления определенной функции или функций (при этом «функция» понимается скорее каузально, чем телескопически) или же в терминах способности осуществлять эти функции. Из этого следует, что, объяснив, как

именно осуществляются эти функции, мы объясняем интересующий нас феномен. Объяснив, каким образом организм осуществляет функцию порождения другого организма, мы объясняем размножение, поскольку размножение и означает осуществление этой функции. Это же верно и для объяснения обучения. Обучение организма в общем сводится к надлежащей адаптации его поведенческих способностей в ответ на стимулы из окружающей среды. Если мы объясним, каким образом организм способен осуществлять соответствующие функции, мы объясним обучение.

(В худшем случае нам не удастся объяснить *феноменальные* аспекты обучения, которые я здесь не рассматриваю по очевидным причинам. Если понятие обучения содержит феноменальный элемент, то он может остаться необъясненным; но здесь я фокусируюсь на психологических аспектах обучения, составляющих, по всей видимости, ядро этого понятия.)

В принципе, объяснение осуществления этих функций идет достаточно прямым путем. Если результаты осуществления подобных функций сами могут быть охарактеризованы физически, и если все физические события имеют физические причины, то должно существовать физическое объяснение осуществления любой функции такого рода. Нужно лишь показать, каким образом определенные состояния отвечают за продуцирование надлежащих итоговых состояний через некий каузальный процесс в соответствии с законами природы. Детали таких физических объяснений могут быть, конечно, нетривиальными. И именно детали составляют львиную долю любого редуктивного объяснения, тогда как аналитический компонент зачастую тривиален. Но при наличии релевантных деталей описание низкоуровневой физической каузальности объяснит осуществление соответствующих функций, а значит, и интересующий нас феномен.

Даже физические понятия, такие как тепло, допускают функциональное конструирование: тепло в общем — это нечто расширяющее металлы, порождаемое огнем, вызывающее определенного рода ощущение и т. п. Объяснив, каким образом реализуются эти каузальные отношения, мы объясним тепло. Тепло есть *каузально-ролевое понятие*, характеризуемое в терминах того, чем оно обычно порождается и что оно порождает при надлежащих условиях. Узнав в результате эмпирических изысканий, как исполняется эта каузальная роль, мы объясняем этот феномен.

Здесь имеются некоторые технические проблемы, но они несущественны. Кripке (Kripke 1980) указал на различие, имеющееся между термином, таким как тепло, и связанным с ним описанием каузальной роли: если принять, что тепло реализуется движением молекул, то это

движение молекул могло бы считаться теплом в контрафактическом мире независимо от того, играли ли они бы там соответствующую каузальную роль. Тем не менее *объяснение* тепла по-прежнему подразумевает объяснение осуществления каузальной роли, а не объяснение движения молекул. Чтобы убедиться в этом, заметим, что идентичность тепла с движением молекул известна апостериори: мы узнаем об этом *в результате* объяснения тепла. Понятие тепла, имевшееся у нас априори — до объяснения этого феномена, — было приблизительно таким: «нечто, играющее данную каузальную роль в нашем действительном мире». Обнаружив, как именно играется эта роль, мы объясняем этот феномен. В качестве бонуса мы обретаем знание о том, что *такое* тепло. Оно есть движение молекул, поскольку движение молекул играет соответствующую каузальную роль в актуальном мире.

Второй небольшой проблемой является то, что многие каузально-ролевые понятия несколько неоднозначны в плане того, идет ли речь о состоянии, играющем определенную каузальную роль, или о реальном исполнении этой роли. «Тепло» может трактоваться либо в качестве обозначения молекул, выполняющих каузальную работу, либо в качестве обозначения самого каузального процесса (нагревания). Сходным образом «восприятие» может отсылать либо к перцептивному акту, либо к внутреннему состоянию, возникающему в его результате. Впрочем, эта двусмысленность не имеет каких-либо важных последствий. Объяснение исполнения каузальной роли объяснит тепло или восприятие в обоих смыслах.

Третья проблема связана с тем, что многие каузально-ролевые понятия частично характеризуются в терминах их воздействия на *опыт*. К примеру, тепло естественно конструируется как причина ощущений тепла. Значит ли это, что, прежде чем мы сможем объяснить тепло, мы должны объяснить ощущения тепла? Разумеется, у нас нет хорошего объяснения ощущений тепла (как и опыта вообще), так что на практике оказывается, что эта часть данного феномена остается без объяснения. Если мы можем объяснить, как происходит движение молекул при определенных условиях, как оно приводит к расширению металлов и определенным образом воздействует на нашу кожу, то вполне достаточно будет ограничиться наблюдением, что это движение *коррелировано* с ощущениями тепла. Из этой корреляции мы заключаем, что здесь почти наверняка имеется каузальная связь. Конечно, ни одно объяснение тепла не будет полным до тех пор, пока мы не покажем, каким образом реализуется эта каузальная связь, но и неполного объяснения будет вполне достаточно для большинства целей. Несколько парадоксально, что мы можем объяснить почти всё в том или ином *феномене*, за исключением деталей его влияния на нашу феноменологию,

но на практике это не порождает проблем. Не стоило бы останавливать развитие остальной науки до появления у нас теории сознания.

Парадигма редуктивного объяснения через функциональный анализ прекрасно работает в большинстве областей когнитивной науки, по крайней мере в принципе. Как мы видели в предыдущей главе, большинство нефеноменальных ментальных понятий может быть подвергнуто функциональному анализу. Психологические состояния могут быть охарактеризованы в терминах играемых ими каузальных ролей. Для объяснения этих состояний мы объясняем, как реализуется соответствующая каузальность.

Редуктивные объяснения в когнитивной науке

В принципе, это можно делать на уровне фундирующей их нейрофизиологии. Объяснив, каким образом определенные нейрофизиологические состояния отвечают за осуществление интересующих нас функций, мы дадим объяснение того или иного психологического состояния. Мы, однако, не обязаны всякий раз опускаться на нейрофизиологический уровень. Зачастую мы можем объяснить какой-то аспект ментального, указав надлежащую *когнитивную модель*, то есть указав детали абстрактной каузальной организации системы, механизмов которой достаточно для осуществления соответствующих функций, но не специфицируя физико-химический субстрат, в котором реализована эта каузальная организация. Подобным способом мы даем *принципиально возможное объяснение* данного аспекта психологии, показывая, каким образом некие каузальные механизмы *могли бы* быть основой соответствующих ментальных процессов. Если нас интересует объяснение ментальных состояний *реального организма* или типа организмов (к примеру, человеческого обучения, в отличие от возможности обучения вообще), то подобное объяснение должно дополняться демонстрацией того, что каузальная организация этой модели отражает каузальную организацию организма, о котором идет речь.

Для объяснения возможности обучения мы можем указать модель, механизмы которой ведут к надлежащим изменениям в поведенческой способности в ответ на разного рода стимулы, идущие из окружающей среды, — к примеру, коннекционистскую модель обучения. Для объяснения человеческого обучения мы должны также продемонстрировать, что подобная модель отражает каузальную организацию, ответственную за осуществление данных функций у людей. Этот второй шаг, как правило, оказывается непростым: мы не можем напрямую указать на это соответствие из-за пробелов в наших нейрофизиологических знаниях и поэтому мы обычно ищем косвенные свидетельства вроде качественных сходств схем наших реакций, временных параметров и т. п. Это одна из причин того, что когнитивная наука пока еще недостаточно

развита. Но опять-таки из функциональной природы психологических понятий можно напрямую заключить к принципиальной возможности подобного объяснения.

К сожалению, функциональное объяснение, так хорошо работающее с психологическими состояниями, похоже, не работает при объяснении феноменальных состояний. Причина этого очевидна. При любом функциональном объяснении человеческого познания возникает *новый вопрос*: почему подобное функционирование сопровождается сознанием? В случае психологических состояний таких новых вопросов не возникает. Если бы кто-то спросил о функциональной модели обучения: «Почему функционирование такого рода сопровождается обучением?», то правильный ответ имел бы семантический характер: «Потому что под обучением *имеется в виду* исключительно функционирование такого рода». Но для понятия сознания не существует подобного анализа. В отличие от психологических состояний, феноменальные состояния не определяются исполняемыми ими каузальными ролями. Из этого следует, что объяснения исполнения какой-то каузальной роли недостаточно для объяснения сознания. После объяснения осуществления некоей функции остается, по сути, необъясненным тот факт, что оно сопровождается сознанием (если это так).

Этот момент можно выразить следующим образом. Если мы располагаем надлежащим функциональным объяснением обучения, то по-просту логически невозможной является ситуация, при которой нечто могло бы быть примером реализации этой объяснительной модели, но не обучаться (во всяком случае в той мере, в какой обучение не требует сознания). Однако, как бы мы функционально ни объясняли познание, кажется логически возможным, что наша объяснительная модель могла бы быть реализована и без какого бы то ни было сопровождающего ее сознания. Это может быть невозможным по естественным причинам — сознание может фактически возникать из этой функциональной организации в актуальном мире, — но важно, что такое представление является логически когерентным.

Если это и в самом деле логически возможно, то любое функциональное и, более того, физическое толкование ментальных феноменов окажется, по сути, незавершенным. Говоря словами Левина (Levine 1983), между подобными толкованиями и сознанием существует *пропад в объяснении*. Даже если на практике соответствующая функциональная организация всегда порождает сознание, вопрос, *почему* она порождает сознание, остается без ответа. В дальнейшем мы подробно обсудим этот момент.

Если так, то отсюда следует, что любому ментальному понятию, содержащему в себе какой-то феноменальный элемент, будет соответ-

ствовать частичный провал в объяснении. К примеру, если для убеждения или обучения требуется сознательный опыт, то мы не сможем дать полное редуктивное объяснение убеждения или обучения. Тем не менее у нас по крайней мере есть основание считать, что *психологические* аспекты этих ментальных характеристик — составляющие, судя по всему, ядро соответствующих понятий — в принципе будут допускать редуктивное объяснение. Если мы отвлечемся от проблем, связанных с феноменологией, то ситуация выглядит таким образом, что у когнитивной науки имеются ресурсы, позволяющие ей успешно справляться с объяснением ментального.

3. Логическая супервентность и редуктивное объяснение

Эпистемология редуктивного объяснения напрямую сочетается с метафизикой супервентности. Естественный феномен может быть редуктивно объясним в терминах каких-то низкоуровневых свойств именно тогда, когда он логически супервентен на этих свойствах. Он редуктивно объясним в терминах физических свойств — или просто «редуктивно объясним» — когда он является логически супервентным на физическом.

Более точно: естественный феномен редуктивно объясним в терминах низкоуровневых свойств, если свойство, реализующее этот феномен, глобально логически супервентно на этих низкоуровневых свойствах. Феномен редуктивно объясним *simpliciter*, если свойство, реализующее этот феномен, глобально логически супервентно на физических свойствах.

Это можно рассматривать в качестве *разъяснения* понятия редуктивного объяснения с возможным элементом условности. То, что из того понятия редуктивного объяснения, которым мы пользовались ранее, следует (глобальная) логическая супервентность, должно быть ясно из предшествующего обсуждения. Если свойство, реализующее феномен, не является логически супервентным на каких-то низкоуровневых свойствах, то какое бы низкоуровневое объяснение этих свойств мы ни дали, всегда будет возникать новый безответный вопрос: почему этот низкоуровневый процесс сопровождается данным феноменом? Редуктивное объяснение предполагает такой анализ интересующего нас феномена, при котором низкоуровневые факты влекут за собой реализацию этого анализа. Поэтому редуктивное объяснение предполагает отношение логической супервентности. Так, именно потому, что размножение логически супервентно на низкоуровневых фактах, оно редуктивно объяснимо в их терминах.

Не столь ясно, является ли логическая супервентность *достаточной* для редуктивной объяснимости. Если феномен *P* логически супервенчен на каких-то низкоуровневых свойствах, то при наличии характеристики низкоуровневых фактов, связанных с наличием *P*, реализация *P* оказывается логическим следствием. Характеристика низкоуровневых фактов будет, таким образом, автоматически приводить к объяснению *P*. Тем не менее подобное объяснение подчас может казаться неудовлетворительным — по двум причинам. Во-первых, низкоуровневые факты могли бы оказаться большой мешаниной на первый взгляд случайных деталей, лишенных явного объяснительного единства. Чем-то подобным могла бы, к примеру, быть характеристика всех молекулярных движений, фундирующих какой-либо процесс обучения. Во-вторых, не исключено, что различные примеры *P* могли бы сопровождаться совершенно разными наборами низкоуровневых фактов, так что объяснения конкретных примеров не давали бы объяснения данного феномена, рассматривавшегося в его родовом аспекте.

Одно из решений состоит в том, чтобы трактовать логическую супервентность как нечто всего лишь *необходимое*, а не достаточное для редуктивного объяснения. Этого хватит для моих аргументов о сознании в следующей главе. Более полезным, однако, будет отметить, что работоспособное понятие редуктивного объяснения, при котором логическая супервентность оказывается как необходимой, так и достаточной, все же существует. Вместо того чтобы рассматривать обозначенные выше проблемы в качестве свидетельства того, что упомянутые характеристики не являются *объяснениями*, мы можем трактовать их как указания на то, что редуктивное объяснение не всегда бывает *поясняющим* объяснением. Редуктивное объяснение — это, скорее, *демистифицирующее* объяснение.

Как я отмечал раньше, редуктивное объяснение не исчерпывает и не завершает всякое объяснение. Его главная роль состоит в том, чтобы устраниТЬ глубинное ощущение таинственности, окружающей высокоуровневый феномен. Оно достигает этой цели, редуцируя данности и произвольность интересующего нас феномена к данности и произвольности низкоуровневых процессов. Поскольку сами низкоуровневые процессы могут быть просто даны и произвольны, редуктивное объяснение может и не приводить нас к глубокому пониманию данного феномена, но оно хотя бы устраняет ощущение, что здесь происходит и еще что-то «сверх того».

Провал между редуктивным объяснением и поясняющим объяснением зачастую может быть, однако, прикрыт гораздо лучше, чем мы только что видели. Это связано с двумя основополагающими фактами, касающимися физики нашего мира: *автономностью* и *простотой*. Микрофизическая каузальность и микрофизические объяснения кажутся

автономными в том смысле, что каждое физическое событие имеет физическое объяснение; законы физики достаточны для объяснения физических событий в их собственных терминах. Кроме того, законы, о которых идет речь, весьма прости, так что такие объяснения довольно компактны. Но ведь все это могло бы выглядеть иначе. Мы могли бы жить в мире с сугубо эмерджентными фундаментальными законами, управляющими поведением высокουровневых комплексов, таких как организмы, и с сопряженной с этим ниспадающей каузальностью, преодолевающей действие любых микрофизических законов, имеющих отношение к таким комплексам. (Британские эмерджентисты, такие как Александр (Alexander 1920) и Броуд (Broad 1925), полагали, что мир устроен примерно таким образом.) Или наш мир мог бы оказаться миром, в котором поведение микрофизических частиц регулируется только лишь бесчисленными причудливыми законами, или же миром, где микрофизическoe поведение хаотично и не подчиняется никаким законам. В подобных мирах оставалось бы мало надежды на получение поясняющих редуктивных объяснений, поскольку элементарность низкоуровневых концепций могла бы не допускать упрощения.

Но актуальный мир с его низкоуровневой автономией и простотой, похоже, не исключает постепенного осмыслиения даже сложных процессов. Низкоуровневые факты, лежащие в основе высокуюровневого феномена, нередко наделены глубоким единством, позволяющим давать исчерпывающее объяснение. Отталкиваясь от высокуюровневой каузальности, такой как выстрел из пистолета, вызванный нажатием курка, мы не только можем изолировать связку устанавливающих ее низкоуровневых фактов, но и достаточно просто изложить, как реализуется эта каузальность, подводя эти факты под те или иные простые принципы. Этот путь может не всегда приводить к успеху. Может оказаться так, что некоторые области, такие как социология и экономика, настолько далеки от простоты низкоуровневых процессов, что поясняющее редуктивное объяснение будет невозможным, даже если соответствующие феномены являются логически супервентными. Если так, пусть так: мы можем удовлетвориться высокуюровневыми объяснениями этих областей, отмечая, что из логической супервентности следует, что в принципе здесь имеется редуктивное объяснение, хотя его, возможно, могло бы понять только высшее существо.

Отметим также, что согласно данной концепции редуктивное объяснение оказывается, по сути, *конкретным*, объясняющим конкретные примеры какого-либо феномена, но не обязательно все эти примеры вместе. Этого нам и следовало ожидать. Если свойство может быть реализовано множеством различных способов, мы не можем ожидать, что одно-единственное объяснение будет применимо ко всем случаям его реализации. К примеру, температура по-разному реализуется

в различных средах, и для каждого случая здесь существует свое объяснение. И крайне маловероятно, что можно было бы найти объяснение, покрывающее все случаи убийства, — если взять пример гораздо более высокого уровня. Тем не менее нередко бывает, что объяснение конкретных случаев отличается известной унифицированностью, так что удачное объяснение одного зачастую объясняет и многие другие. Это опять-таки скорее следствие фундаментальной простоты нашего мира, чем необходимая черта объяснения. В нашем мире простые унифицированные истории, которые можно рассказать о низкоуровневых процессах, нередко обладают универсальной значимостью или по крайней мере значимостью для самых разных индивидов. Часто бывает и так, особенно в биологических науках, что индивиды имеют общее происхождение, что ведет к сходству их низкоуровневых процессов. Поэтому вторая из упомянутых проблем, а именно унификация объяснений конкретных примеров феномена, не является настолько большой проблемой, какой она могла бы быть. В любом случае главным оказывается объяснение индивидуальных случаев.

Можно было бы еще долго рассуждать о том, как закрыть провал между редуктивным объяснением и поясняющим объяснением, но этот вопрос заслуживает отдельного обстоятельного рассмотрения, и он не слишком важен для моих целей. Ключевое значение имеет то, что при отсутствии логической супервентности (как в случае сознания, что я попытаюсь показать в дальнейшем) нельзя будет осуществить никакого редуктивного объяснения, даже если мы будем щедро раздавать этот титул. Важно и то, что логическая супервентность устраниет все остатки *метафизических* тайн в высокоуровневых феноменах, редуцируя их данности к данностям низкоуровневых фактов. Важным, пусть и в меньшей степени, следует признать и то, что при наличии логической супервентности будет возможно редуктивное объяснение. Хотя эти объяснения могут быть туманными и бесполезными, эти недостатки далеко не так существенны, как объяснительные пробелы там, где нет логической супервентности.

Дальнейшие замечания о редуктивном объяснении

Еще несколько замечаний. Во-первых, редуктивное объяснение феномена на практике обычно не доходит до микромирофизического уровня. Это было бы исключительно трудным делом и порождало бы все те проблемы, связанные с непроясненными данностями, о которых шла речь выше. Вместо этого высокоуровневые феномены объясняются в терминах чуть более фундаментальных свойств. К примеру, размножение объясняется в терминах клеточных механизмов, а фазы луны — в терминах орбитального движения. И мы надеемся, что сами эти более фундаментальные феномены будут редуктивно

объяснимы в терминах чего-то еще более фундаментального. Если все пойдет хорошо, то биологические феномены можно будет объяснить в терминах клеточных феноменов, объяснимых в терминах биохимических феноменов, объяснимых в терминах химических феноменов, объяснимых в терминах физических феноменов. Что же касается физических феноменов, то надо пытаться унифицировать их, насколько это возможно. Однако на каком-то уровне физику надо принять просто как данность: возможно, нельзя найти объяснение тому, почему фундаментальные законы и ограничительные условия таковы, каковы они есть. Об этой объяснительной лестнице пока можно только мечтать, но в движении к ней мы все же видим значительный прогресс. Исходя из логической супервентности, а также из простоты и автономности самого базового уровня, мы можем допустить принципиальную возможность подобного объяснительного соединения наук. Но пока неясно, не помешают ли различные сложности мира практической реализуемости этой цели.

Во-вторых, по крайней мере представимо, что феномен будет редуктивно объясним в терминах низкоуровневых свойств, но не *simplifier*. Это могло бы случиться в ситуации, где С-свойства логически супервентны на В-свойствах и поэтому объяснимы в терминах В-свойств, притом что сами В-свойства не являются логически супервентными на физическом. Очевидно, что в одном смысле подобное объяснение будет редуктивным, в другом — нет. По большей части меня будет интересовать редуктивное объяснение в терминах физического, в терминах свойств, которые могут быть объяснены в терминах физического, и т. д. Даже если наши С-свойства в относительном смысле редуктивно объяснимы, само их наличие говорит о том, что редуктивное объяснение не может быть тотальным.

В-третьих, локальная логическая супервентность — слишком строгое требование для редуктивного объяснения. Можно редуктивно объяснить даже контекстуально обусловленные свойства индивида, указав механизмы реализации соответствующих им отношений окружающей среды. В той мере, в какой феномен является глобально супервентным, он допускает редуктивное объяснение в терминах каких-то низкоуровневых фактов, даже если эти факты далеко разбросаны в пространстве и во времени.

В-четвертых, в принципе можно говорить о двух проектах редуктивного объяснения таких феноменов, как жизнь, обучение или тепло. Первый из них — *экспликация*, когда мы проясняем посредством анализа то, что должно быть объяснено. К примеру, обучение можно было бы проанализировать как процесс адаптации. Второй — *объяснение*, когда мы показываем, как подобный анализ реализуется

низкоуровневыми фактами. Первый проект концептуален, второй же имеет эмпирический характер. Для многих или большинства феноменов концептуальная стадия будет весьма тривиальной. Однако для некоторых феноменов, таких как убеждение, экспликация может представлять собой очень большую проблему. На практике, конечно, эти проекты никогда четко не разводятся, поскольку экспликация и объяснение идут параллельными курсами.

4. Концептуальная истина и необходимая истина*

В своих рассуждениях о супервентности и объяснении я во многом опирался на понятия логической возможности и необходимости. И теперь пришло время поподробнее поговорить об этом. Логическую необходимость какого-либо утверждения можно уяснить прежде всего в терминах его истинности во всех логически возможных мирах. При осмыслении соответствующего класса миров и способов оценки тех или иных утверждений в этих мирах требуется определенная осмотрительность; позже в этом параграфе я еще более или менее подробно обсажу данный вопрос. Логическую необходимость какого-либо утверждения можно также истолковать как истинность по значению: утверждение логически необходимо, если его истинность гарантируется значением используемых в нем понятий. Но опять-таки при истолковании «значений» требуется определенная осмотрительность. Далее в этом параграфе я рассмотрю оба этих варианта интерпретации, а также обсажу вопрос об их отношении друг к другу.

(Как и раньше, понятие логической необходимости не следует брать в узком смысле, подразумевающем выводимость в терминах логики первого порядка или какого-либо другого синтаксического формализма. В самом деле, можно попробовать доказать, что обоснование аксиом и правил этих формальных систем зависит как раз от их логической необходимости в этом более широком и более изначальном смысле.)

Все это требует серьезного отношения, по крайней мере до определенной степени, к понятию *концептуальной истины*, то есть к представлению о том, что некоторые утверждения являются истинными или ложными исключительно вследствие значений используемых в них терминов. Ключевые моменты моего предшествующего обсуждения зависели от характеристик, даваемых различным понятиям. К примеру, я говорил о возможности редуктивного объяснения размножения, доказывая, что из низкоуровневых деталей вытекает осуществление определенных функций и что осуществлением этих функций исчерпывается понятие размножения.

Понятие концептуальной истины стало одиозным в некоторых кругах после критических выпадов Куайна (Quine 1951), доказывавшего, что нельзя провести сколь бы то ни было полезного различия между концептуальными и эмпирическими истинами. Возражения, высказываемые против этих понятий, обычно группируются вокруг следующих моментов.

1. Большинство понятий лишены дефиниций, указывающих необходимые и достаточные условия (это наблюдение делалось много-кратно, но его нередко ассоциируют с Витгенштейном (Wittgenstein 1953)).
2. Большинство истин, кажущихся концептуальными, в действительности может быть пересмотрено, и они могут быть отброшены перед лицом достаточных эмпирических свидетельств (об этом говорил Куайн).
3. Соображения об *апостериорной необходимости*, высказанные Кripке (Kripke 1972), показывают, что условия применения множества терминов в возможных мирах не могут быть установлены априори.

Эти соображения свидетельствуют против чрезмерно упрощенного представления о концептуальной истине, но не против того, как я использую эти понятия. В частности, они не затрагивают класса *супервентностных кондиционалов* — «если А-факты о какой-либо ситуации являются собой *X*, то В-факты оказываются *Y*», где А-факты полностью специфицируют ситуацию на самом базовом уровне. Мой аргумент нуждается только в таких концептуальных истинах, и мы увидим, что ни одно из приведенных выше соображений не свидетельствует против них. Я также более детально проанализирую отношение между концептуальной и необходимой истиной и проясню их роль в понимании логической супервентности.

Отсутствие четких дефиниций — наименее серьезная из трудностей, связанных с концептуальными истинами. Ни один из моих аргументов не зависит от существования подобных дефиниций. Иногда я отсылаю к анализу различных понятий, но эти анализы могут быть лишь приблизительными и не претендующими на указание точных необходимых и достаточных условий. Большинство понятий (к примеру, «жизнь») несколько неопределенны в своем применении и нет большого смысла пытаться устраниć эту неопределенность путем искусственных уточнений. Вместо того чтобы говорить, что «система является живой, если и только если она воспроизводится, адаптируется с коэффициентом полезности 800 или выше и осуществляет метаболизм с эффективностью в 75 процентов

Дефиниции

или проявляет все это в надлежащем сочетании с такими-то и такими-то свойствами», мы можем просто заметить, что, если система проявляет все это в достаточной степени, она будет живой вследствие значения этого термина. Если описание соответствующих низкоуровневых фактов фиксирует данные факты о размножении, полезности, метаболизме и т. д. системы, то оно также фиксирует факты о том, является ли эта система *живой*, — в той мере, в какой этот вопрос вообще имеет фактуальный характер.

Мы можем обобщить это с помощью схематической диаграммы (рис. 2.1), показывающей, как высокогенетическое свойство *P* могло бы зависеть от двух низкоуровневых параметров *A* и *B*, каждый из которых может принимать множество значений. Если бы у нас имелось четкое определение в терминах необходимых и достаточных условий, то ситуация напоминала бы картину слева, где черный прямоугольник изображает область реализации свойства *P*. Зависимость, однако, неизменно напоминает картину справа с размытыми границами и большим участком, в котором вопрос о наличии свойства *P* остается неопределенным, — при наличии, однако же, и участка, где этот вопрос определен. (Неопределенным, возможно, является, живы ли бактерии и компьютерные вирусы, но нет сомнений, что собаки — это живые существа.) При наличии примера, взятого из второго участка, реализующего *A* и *B* в степени, достаточной для реализации *P*, кондиционал «если *x* есть *A* и *B* в этой степени, то *x* есть *P*» оказывается концептуальной истиной, несмотря на отсутствие четкой дефиниции *P*. Любая неопределенность в таких кондиционалах, в серых участках, будет отражать фактуальную неопределенность, как и должно быть. Эту картину можно напрямую расширить до зависимости свойства от любого множества факторов и до супервентностных кондиционалов в целом.

Важно, таким образом, что одно множество фактов может влечь за собой другое множество и при отсутствии четкого определения понятий второго множества в терминах первого. Указанный выше случай может быть примером этого: не существует простой дефиниции *P* в терминах *A* и *B*, но факты относительно *A* и *B* сразу влекут за собой факты о *P*. В качестве другого примера возьмем *округлость* замкнутых кривых в двумерном пространстве (рис. 2.2). Округлость наверняка нельзя точно определить в терминах более простых математических понятий. Тем не менее посмотрим на фигуру слева, характеризуемую уравнением $2x^2 + 3y^2 = 1$. В той мере, в какой округлость вообще фактуальна, фактом является то, что эта фигура округла (сравним с фигурой справа, которая уж точно не округла). Более того, этот факт *вытекает* из базовой характеристики данной фигуры в математических терминах — наличие этой характеристики и понятия округлости определяют факт округлости данной фигуры. Принимая во внимание, что *A*-факты

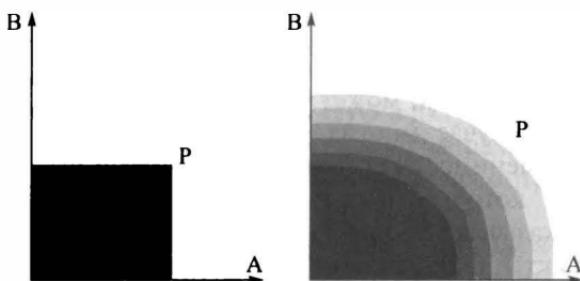


Рис. 2.1. Два варианта возможной зависимости свойства P от свойств A и B

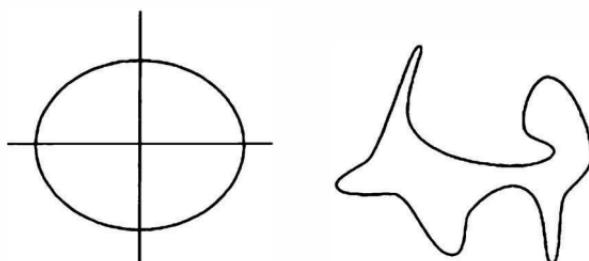


Рис. 2.2. Округлая кривая $2x^2 + 3y^2 = 1$
и ее лишенный окружности товарищ

могут влечь за собой В-факты без определения В-фактов в терминах А-фактов, можно заключить, что понятие логической супервентности не затрагивается отсутствием дефиниций. (Полезно помнить об этом примере, размышляя о более сложных проблемах и возражениях, связанных с логической супервентностью.)

Этот момент можно выразить замечанием, что «значение» понятия, релевантное в большинстве случаев, является не дефиницией, а интенсионалом: функцией, специфицирующей то, каким образом понятие применяется в различных ситуациях. Иногда интенсионал можно выразить в дефиниции, но, как показывают наши случаи, это не обязательно. Но пока можно отсыпать к фактам о том, как понятия применяются в различных ситуациях, у нас будут интенсионалы; и мы вскоре увидим, что этим обычно и будет исчерпываться «значение», нужное для моих аргументов.

Второе возражение, выдвинутое Куайном (Quine 1951), состоит в том, что предполагаемые концептуальные истины всегда могут быть пересмотрены перед лицом достаточных эмпирических свидетельств. К примеру, если какие-то данные

Ревизуемость

вынуждают меня пересмотреть различные базовые положения какой-то теории, то не исключено, что положение, некогда представлявшееся концептуально истинным, может оказаться ложным.

Это верно в случае многих предположительно концептуальных истин, но неприменимо к рассматриваемым нами супервентностным кондиционалам, которые имеют следующий вид: «если низкоуровневые факты оказываются такими-то, высокуюровневые факты будут такими-то». Факты, специфицированные в антецеденте этого кондиционала, на деле включают все релевантные эмпирические факторы. Эмпирические свидетельства могут показать нам ложность антецедента этого кондиционала, но не ложность его самого. В предельном случае можно сделать так, чтобы антецедент полностью специфицировал низкоуровневые факты о мире. Сама полнота антецедента гарантирует нерелевантность эмпирических свидетельств для истинностного значения кондиционала. (Эта картина несколько осложняется существованием апостериорных необходимостей, о которых вскоре пойдет речь. Здесь меня интересуют только эпистемические кондиционалы, касающиеся возможных вариантов актуального мира.)

Хотя соображения о ревизуемости могут быть реальным доводом в пользу отсутствия многих *простых* концептуальных истин, ничто в них не говорит против наличия того лимитированного, комплексного типа концептуальных истин, который был предметом моего рассмотрения. Итогом этих соображений является то, что условия истинности высокоуровневых положений могут и не быть легко *локализуемыми*, так как на них могли бы косвенно влиять самые разные факторы; но это не ставит под угрозу глобальные условия истинности, выражаемые супервентностными кондиционалами. В самом деле, если значение определяет функцию от возможных миров к классам референции (интенсионал), и если возможные миры допускают конечное описание (скажем, в терминах размещения базовых качеств в этих мирах), то в результате мы автоматически получаем громадный класс концептуально истинных кондиционалов.

Апостериорная необходимость

Традиционно считалось, что все концептуальные истины познаются априори, как и все необходимые истины, и что классы априорных истин, необходимых истин и концептуальных истин близкородственны или даже совпадают по своему объему. Книга Сола Крипке «Именование и необходимость» (Kripke 1972) покоробила эту картину. В ней доказывалось существование большого класса необходимо истинных положений, истинность которых не является чем-то, что познается априори. Примером является положение «вода есть H_2O ». Мы не можем знать о его истинности априори; вполне можно предположить (или вполне

можно предположить в начале изысканий), что вода состоит из чего-то другого, возможно XYZ. Кripке доказывает, что, невзирая на это, если исходить из того, что вода есть H_2O в актуальном мире, получается, что вода есть H_2O во всех возможных мирах. Из этого следует, что «вода есть H_2O » — необходимая истина, несмотря на то что она апостериорна по своей природе.

Это создает ряд трудностей для предложенной мной модели. К примеру, согласно некоторым толкованиям, эти необходимые истины оказываются концептуальными истинами, из чего следует, что не все концептуальные истины познаются априори. Согласно альтернативным толкованиям, подобные положения не являются концептуальными истинами, но тогда разрушается связь между концептуальной истиной и необходимостью. В различных частях этой книги я использую априорные методы, чтобы удостовериться в необходимости; но именно такие процедуры, как нередко утверждается, ставятся под вопрос концепцией Кripке.

По размышлении я полагаю, что можно показать, что эти сложности не вносят существенных изменений в мои аргументы; но дело стоит того, чтобы прояснить, что тут, собственно, происходит. Я потрачу некоторое время для создания систематической модели, позволяющей справляться с этими проблемами, которые будут периодически напоминать о себе. В частности, я покажу, как можно естественным путем выразить кripкеанские интуиции в двумерной картине значения и необходимости. Эта модель представляет собой синтез идей, высказанных Кripке, Патнэмом, Капланом, Столнейкером, Льюисом, Эвансом, Дэвисом, Хамберстоуном и другими авторами, рассматривавшими эти двумерные феномены.

Согласно традиционному взгляду на референцию, восходящему к Фреге, но выраженному здесь в современной терминологии, понятие определяет некую функцию $f: W \rightarrow R$ от возможных миров к референтам. Подобная функция зачастую именуется *интенсионалом*; вместе со спецификацией мира w она определяет *экстенсионал* $f(w)$. Сам Фреге считал, что каждое понятие имеет *смысл*, который, как предполагалось, определяет референт данного понятия в зависимости от состояния мира; так что эти смыслы тесно соотносятся с интенсионалами. Смысл нередко трактовали в качестве *значения* соответствующего понятия.

Более современные исследования привели к признанию того, что никакой единичный интенсионал не может выполнить всю ту работу, которую должно выполнять значение. Картина, полученная Кripке, усложнила ситуацию, показав, что референция в актуальном мире и референция в контрфактическом мире определяются совершенно разными механизмами. В известном отношении кripкеанская картина

словно бы разбивает фрегеанскую картину на два отделенных друг от друга уровня.

Эту крипкеанскую интуицию можно выразить, сказав, что на деле с понятием связано *два* интенсионала. Иными словами, существуют две совершенно разные схемы зависимости референта понятия от состояния мира. Во-первых, это зависимость, фиксирующая референцию в *актуальном* мире сообразно тому, каким является этот мир: если он является таким-то, то понятие будет выделять одно, если каким-то другим — другое. Во-вторых, это зависимость, определяющая референцию в *контрфактических* мирах при условии фиксации референции в актуальном мире. Каждой из этих зависимостей соответствует свой интенсионал. Я буду называть эти интенсионалы соответственно *первичным* и *вторичным*.

Первичный интенсионал понятия — это функция от миров к экспенсионалам, отражающая способ фиксации референции в актуальном мире. Она выделяет в мире тот референт понятия, который имелся бы при актуальности этого мира. Возьмем понятие «вода». Если бы океаны и озера актуального мира заполняла XYZ, то «вода» отсылала бы к XYZ²¹, но поскольку фактически их заполняет H₂O, «вода» отсылает к H₂O. Поэтому первичный интенсионал «воды» соотносит XYZ-мир с XYZ, H₂O-мир — с H₂O. В первом приближении можно было бы сказать, что первичный интенсионал указывает в мире на прозрачную, пригодную для питья жидкость, преобладающую в океанах и морях, или, короче, на *водянистую материю*.

Однако поскольку оказывается, что «вода» отсылает к H₂O в актуальном мире, Крипке (как и Патнэм (Putnam 1975)) замечает, что мы вправе сказать, что вода есть H₂O в любом контрфактическом мире. *Вторичный интенсионал* «воды» указывает на воду в любом контрфактическом мире; поэтому, если Крипке и Патнэм правы, он указывает на H₂O во всех мирах²².

Для моих целей наибольшую важность имеет первичный интенсионал: именно первичный интенсионал содержит то, что нуждается в объяснении в понятии естественного феномена. Если бы кто-нибудь попросил объяснить воду задолго до того, как мы узнали, что вода в действительности является H₂O, то они так или иначе спрашивали бы об объяснении прозрачной, пригодной для питья жидкости, окружающей их. Только *после* завершения объяснения мы узнаем, что вода есть H₂O. Первичный интенсионал понятия, в отличие от вторичного интенсионала, независим от эмпирических факторов: он *специфицирует* то, каким образом референция зависит от наличного состояния внешнего мира, а значит, сам по себе не зависит от того, каким именно оказывается этот внешний мир.

Разумеется, любая краткая характеристика первичного интенсионала понятия в духе «преобладающая в окружающей среде прозрачная, пригодная для питья жидкость» будет упрощением. Подлинный интенсионал может быть определен только при детальном рассмотрении конкретных сценариев: что сказали бы мы, если бы мир оказался таким? А что сказали бы мы, если бы он оказался вот таким? К примеру, если бы оказалось, что жидкость в озерах — это H_2O , а жидкость в океанах — XYZ, то мы, вероятно, стали бы говорить, что и то и другое является водой. Если бы материя, заполняющая океаны и озера, оказалась смесью 95 процентов A и 5 процентов B, мы, вероятно, сказали бы, что вода — это A, но не B; если бы оказалось, что некая субстанция, не являющаяся ни прозрачной, ни пригодной для питья, родственна в микрофизическом смысле прозрачной, пригодной для питья жидкости, окружающей нас, мы, вероятно, тоже называли бы ее «водой» (как мы делаем это со льдом или с «грязной водой»). Полный перечень условий того, чтобы считаться «водой», будет весьма размытым по краям и не обязан быть непосредственно очевидным для рефлексии, но ничто из этого существенно не меняет ту картину, которую я описываю. Я буду использовать слова «водянистая материя» как искусственный термин для выражения первичного интенсионала, чем бы он ни был²³.

В некоторых случаях решение о том, к чему отсылает понятие в актуальном мире, предполагает немалый объем размышлений относительно того, что именно было бы наиболее уместным сказать; как, например, в вопросах о референции «массы», если бы актуальный мир оказался миром, где верна общая теория относительности²⁴, или, возможно, в вопросах о том, что считается «убеждением» в актуальном мире. Поэтому рассмотрение того, на что именно указывает первичный интенсионал у разных кандидатов на роль актуального мира, может потребовать соответствующего объема размышлений. Но это не значит, что речь идет не об априорных разысканиях: мы можем размышлять об этом независимо от актуального состояния мира. Допустим, что сообщения об экспериментах, подтверждающих относительность, опровергаются, и, соответственно, мы не уверены, является ли актуальный мир релятивистским миром, но в любом случае мы можем размышлять о том, к чему отсылала бы «масса», если бы такое положение дел оказалось актуальным.

(При анализе первичных интенсионалов понятий, используемых индивидами в лингвистическом сообществе, возникают различные сложности. С ними можно справиться, отметив, что понятие, имеющееся у индивида, может быть наделено первичным интенсионалом, предполагающим признание авторитета окружающего сообщества — к примеру, мое понятие «вяза» могло бы относиться к тому, что окружающие называют «вязом»; но в любом случае подобные проблемы

не имеют отношения к тем вопросам, которыми мы будем заниматься, — при их рассмотрении мы вполне могли бы допустить, что в сообществе есть лишь один человек, или что все индивиды в равной степени хорошо информированы, или даже что это сообщество есть один гигантский индивид. При использовании первичных интенсионалов для построения общей семантической теории мог бы возникнуть и ряд технических проблем, к примеру: является ли референция понятия существенной для этого понятия? Могли ли бы различные люди ассоциировать различные первичные интенсионалы с одним и тем же словом? Но я не пытаюсь построить здесь законченную семантическую теорию, и мы можем отвлечься от такого рода проблем.

Подчас философы с подозрением относятся к таким сущностям, как первичные интенсионалы, так как они напоминают им «дескриптивную» теорию референции. Но дескрипции не играют существенной роли в этой модели. Я использую их лишь для конкретизации специфики соответствующих функций от возможных миров к экстенсионалам. Подлинно существенной оказывается сама функция, а не какая-либо суммирующая ее дескрипция. Эта картина вполне совместима с «каузальной» теорией референции: нам надо лишь заметить, что первичный интенсионал понятия, такого как «вода», может требовать надлежащей каузальной связи между референтом и субъектом. В самом деле, мы приходим к убеждению в верности каузальной теории референции прежде всего именно в результате рассмотрения различных вариантов того, каким мог бы быть актуальный мир, и констатаций, чем был бы референт данного понятия в этих случаях; то есть в результате оценки первичного интенсионала понятия в этих мирах.)

Исходя из того, что референция «воды» в актуальном мире фиксируется через водянистую материю, можно было бы подумать, что вода является водянистой материей во всех возможных мирах. Кripке и Патнэм показали, что это не так: если вода есть H_2O в актуальном мире, то вода есть H_2O во всех возможных мирах. В мире (патнэмовском «Двойнике Земли»), где преобладает прозрачная, пригодная для питья жидкость XYZ, а не H_2O , эта жидкость не является водой; это просто водянистая материя. Все это выражается *вторичным интенсионалом «воды»*, выделяющим воду во всех мирах, то есть выделяющим H_2O во всех мирах.

Вторичный интенсионал понятия, такого как «вода», не определяется априори, так как он зависит от положения дел в актуальном мире. Но он тем не менее тесно связан с первичным интенсионалом. В данном случае для определения вторичного интенсионала нужно сначала оценить первичный интенсионал в актуальном мире, а затем *придать жесткость* этой оценке, чтобы подобная вещь выделялась во всех

возможных мирах. Поскольку первичный интенсионал («водянистая материя») указывает на H_2O в актуальном мире, из придания этому отношению жесткости следует, что вторичный интенсионал указывает на H_2O во всех возможных мирах.

Можно суммировать сказанное, заметив, что «вода» концептуально эквивалентна $\langle dthat \rangle$ (водянистая материя), где $dthat$ — разновидность жесткого оператора Каплана, превращающего интенсионал в жесткий десигнатор по итогам оценки первого в актуальном мире (Kaplan 1979). Единый фрегеанский интенсионал разбивается на две части: первичный интенсионал («водянистая материя»), фиксирующий референцию в актуальном мире, и вторичный интенсионал ($\langle H_2O \rangle$), указующий референцию в контрафактических возможных мирах и зависящий от положения дел в актуальном мире.

(Некоторые склонны считать, что апостериорная необходимость оставляет не у дел априорный концептуальный анализ, но такое предположение безосновательно. Еще до того, как поставить вопрос о жесткой десигнации и т. п., мы уже можем сказать, что именно делает X в актуальном мире таким, что он может *считаться* референтом $\langle X \rangle$. И это можно сказать только произведя анализ первичного интенсионала. И подобное начинание является априорным, так как оно предполагает получение ответов на вопросы о том, к чему отсыпало *бы* наше понятие, если бы актуальный мир был таким-то или таким-то. Если мы можем знать, к чему отсылают наши понятия при знании о состоянии дел в актуальном мире, то мы можем знать, к чему отсылали *бы* наши понятия, если бы актуальный мир был таким-то или таким-то. *Действительно* ли актуальный мир оказывается таким-то или таким-то, мало влияет на ответ на этот вопрос, не считая того, что это обстоятельство позволяет сфокусировать наше внимание.)

Как первичный, так и вторичный интенсионалы могут рассматриваться в качестве функций $f: W \rightarrow R$ от возможных миров к экстенсионалам, однако возможные миры при этом трактуются несколько по-разному. Мы могли бы сказать, что первичный интенсионал указывает на референт понятия в мире при *трактовке последнего в качестве актуального мира* — то есть при его рассмотрении в качестве кандидата на роль актуального мира для того, кто мыслит о нем, — тогда как вторичный интенсионал указывает на референт понятия в мире при *трактовке последнего в качестве контрафактического мира*, если актуальный мир уже зафиксирован для того, кто мыслит о нем. Если XYZ-мир рассматривается как актуальный, то мой термин «вода» указывает на XYZ в нем, но если он рассматривается как контрафактический мир, то «вода» указывает на H_2O .

Данное различие двух способов взгляда на миры близко каплановскому различию (Kaplan 1989) *контекста выражения* мысли

и обстоятельств оценки. Если мы рассматриваем мир *w* как контрафактический, то мы задействуем актуальный мир в качестве контекста выражения, тогда как *w* используется в качестве обстоятельств оценки. Так, если я выражаю мысль «в океане есть вода» в этом мире и *оценываю* ее в XYZ-мире, то «вода» отсылает к H₂O, и мое утверждение должно. Но если мы рассматриваем *w* как актуальный мир, мы представляем его в качестве потенциального контекста выражения мысли и задаемся вопросом о том, как обстояло бы дело в случае, если бы контекстом выражения оказался *w*. Если бы контекстом моей фразы «в океане есть вода» оказался XYZ-мир, это утверждение было бы истинным при его оценке в этом мире. Первичный интенсионал, таким образом, близок тому, что Каплан называет *характером* термина, хотя здесь есть и ряд отличий²⁵, а вторичный интенсионал соответствует тому, что он называет *содержанием* термина.

Небольшая асимметрия между ними обнаруживается, когда мы обращаем внимание, что Куайн (Quine 1969) называет *центрированным* возможным миром именно контекст выражения, а не обстоятельства оценки. Речь идет об упорядоченной паре, состоящей из мира и *центра*, репрезентирующего точку зрения агента внутри этого мира, использующего интересующий нас термин: центр состоит (по меньшей мере) из «выделенного» индивида и времени. (Эта идея восходит к Льюису (Lewis 1979); Куайн полагает, что центр мог бы быть точкой в пространстве-времени.) Подобный центр нужен для того, чтобы отразить тот факт, что термин, такой как «вода», имеет разный экстенсионал для меня и для моего двойника на Двойнике Земли, несмотря на то что мы живем в одном мире²⁶. Отличие состоит только в нашем положении в мире, и именно это положение производит соответствующее различие в процессе фиксации референции.

Этот феномен особенно заметен в случае индексикальных терминов, таких как «я», референция которых очевидным образом зависит не только от общего состояния мира, но и от того, кто использует данный термин: первичный интенсионал «я» указывает на индивида в центре центрированного мира. (Вторичный интенсионал моего понятия «я» указывает на Дэвида Чалмерса во всех возможных мирах.) Не столь явный индексикальный элемент присутствует, однако, и в таких понятиях, как «вода». Его можно выразить примерно так: «*dthat* (преобладающая прозрачная, пригодная для питья жидкость *в нашем окружении*)»²⁷. Именно из-за этого индексикального элемента первичные интенсионалы должны быть поставлены в зависимость от центрированных миров. Но после фиксации референции в актуальном мире для ее оценки в контрафактическом мире больше не нужен никакой центр. Обстоятельства оценки могут, таким образом, быть представлены простым возможным миром, лишенным центра.

Все это можно формализовать, заметив, что полное описание референции в контрафактических мирах не определяется априори сингулярно индексированной функцией

$$f : W \rightarrow R.$$

На деле референция в контрафактическом мире зависит как от этого мира, так и от того, каким является актуальный мир. Иначе говоря, понятие определяет двойственную индексированную функцию

$$F : W^* \times W \rightarrow R,$$

где W^* есть пространство центрированных возможных миров, а W есть пространство обычных возможных миров. Первый параметр репрезентирует контексты выражения, или возможные варианты актуального мира, второй — обстоятельства оценки, или контрафактические возможные миры. Или, что то же самое, понятие определяет семейство функций

$$F_v : W \rightarrow R$$

для всякого $v \in W^*$, репрезентируя возможные варианты актуального мира, где $F_v(w) = F(v, w)$. В случае «воды» если a — это мир, в котором водянистая материя есть H_2O , то F_a указывает на H_2O в любом возможном мире. Поскольку в нашем мире вода оказалась H_2O , эта F_a конкретизирует условия корректного применения «воды» в контрафактических мирах. Если бы наш мир оказался другим миром b , в котором водянистая материя была бы XYZ , соответствующие условия применения конкретизировались бы посредством F_b , другого интенсионала, выделяющего XYZ в любом возможном мире.

Функция F определяется априори, поскольку все апостериорные факторы оказываются включенными в ее параметры. Из F мы можем получить оба наших сингулярно индексированных интенсионала. Первичный интенсионал есть функция $f : W^* \rightarrow R$, определяемая «диагональным» отображением $f : w \mapsto F(w, w')$, где w' идентичен w , за исключением отсутствия центра. Это функция, посредством которой фиксируется референция в актуальном мире. Вторичный интенсионал есть соотнесение $F_a : w \mapsto F(a, w)$, где a — наш актуальный мир. Этот интенсионал вычисляет референцию в контрафактических мирах. Непосредственным следствием оказывается то, что первичный и вторичный интенсионалы совпадают в их применении к актуальному миру:

$$f(a) = F_a(a') = F(a, a').$$

В обратном направлении двойственную индексированную функцию F , а поэтому и вторичный интенсионал F_a , обычно могут быть извлечены из первичного интенсионала f при помощи «правила»,

говорящего о том, каким образом вторичный интенсионал зависит от первичного интенсионала и актуального мира a . Это правило зависит от типа понятия. Для понятия, являющегося жестким десигнатором, оно состоит в том, что в мире w вторичный интенсионал указывает в w на то, на что указывает первичный интенсионал в a (или, быть может, в случае терминов естественных видов, то, что имеет такую же базовую структуру, как то, на что указывает первичный интенсионал в a). Более формально: пусть $D: R \times W \rightarrow R$ будет «проективным» оператором от класса, выделяемого в каком-то мире, к членам «этого» класса в другом возможном мире. Тогда вторичный интенсионал F_a есть не что иное, как функция $D(f(a), -)$, которую можно мыслить как *dthat* в ее применении к интенсионалу, данному посредством f .

У других понятий получение вторичного интенсионала из первичного будет более простым. В случае «дескриптивных» выражений, таких как «доктор», «площадь» или «водянистая материя», жесткая десигнация не играет особой роли: они применяются к контрафактическим мирам независимо от того, каким именно оказывается актуальный мир. Здесь вторичный интенсионал является простой копией первичного интенсионала (за исключением различий, связанных с центрированием). Очерченная мной модель хорошо работает с обоими видами понятий.

Термины, обозначающие *свойства*, такие как «теплое», могут представляться в интенсиональных рамках одним из двух способов. Интенсионал свойства можно рассматривать в качестве функции от мира к классу индивидов (индивидуов, реализующих данное свойство) или же от мира к самим свойствам. Оба этих способа совместимы с предложенной моделью: в каждом из этих случаев несложно отыскать первичный и вторичный интенсионал; несложно и переходить от одной схемы к другой. Впрочем, обычно я буду действовать первым способом, принимая, что первичный интенсионал «теплого» указывает на такие вещи, которые могут считаться «теплыми» в актуальном мире в зависимости от того, каков он, тогда как вторичный интенсионал указывает на теплые вещи в контрафактическом мире, если актуальный мир является таким, каков он есть.

Как первичные, так и вторичные интенсионалы могут рассматриваться в качестве кандидатов на роль «значения» понятия. Думаю, нет смысла выбирать один из них в качестве *подлинного* значения; термин «значение» здесь по большей части почетная формальность. С тем же успехом мы могли бы представить первичный и вторичный интенсионалы в качестве соответственно априорного и апостериорного аспектов значения.

Если мы идентифицируем их подобным образом, то окажется, что оба этих интенсионала фундируют определенный тип концептуальных истин, или истинность по значению. Первичный интенсионал фундирует априорные истины, такие как «вода есть водянистая материя». Подобное утверждение будет истинным независимо от состояния актуального мира, хотя оно и не обязано быть верным во всех неактуальных возможных мирах. Вторичный интенсионал не является основой априорных истин, но фундирует истины, значимые для всех контрафактических возможных миров, такие как «вода есть H_2O ». И те, и другие могут считаться истинами по значению: но суть в том, что речь идет о разных аспектах значения.

Можно также истолковать обе как варианты *необходимых* истин. Вторая из них соответствует более привычной конструкции необходимой истины. Но и первая может быть представлена как истина возможных миров, если только сконструировать их как контексты выражения, или в качестве возможных вариантов актуального мира. Согласно этой несколько иной конструкции, положение S является необходимо истинным в том случае, если S окажется истинным, каким бы ни оказался актуальный мир. Если актуальный мир оказывается миром, в котором водянистая материя есть XYZ, то мое утверждение «XYZ есть вода» окажется истинным. Таким образом, в соответствии с этой конструкцией, при которой возможные миры *рассматриваются как актуальные*, утверждение «вода есть водянистая материя» является необходимой истиной.

Этот вид необходимости Эванс (Evans 1979) называет «глубокой необходимостью» в противоположность «поверхностным» необходимостям, таким как «вода есть H_2O ». Она детально анализируется Дэвисом и Хамберстоуном (Davies and Humberstone 1980) при помощи модального оператора, который они называют «фиксировано актуально». Глубокая необходимость, в отличие от поверхностной необходимости, не затрагивается апостериорными соображениями. Два этих варианта возможности и необходимости всегда имеют отношение к *утверждениям*. Есть только один надлежащий тип возможности *миров*; и отличие двух этих подходов состоит в способах оценки в мире истинности тех или иных утверждений.

Мы можем выразить это и иным способом, заметив, что с любым утверждением связаны два набора *условий истиности*. Если мы оцениваем термины какого-то утверждения сообразно их первичным интенсионалам, мы приходим к *первичным* условиям истинности утверждения, то есть к множеству центрированных возможных миров, в которых данное утверждение, оцененное в соответствии с первичными интенсионалами его терминов, оказывается истинным. Первичные

условия истинности говорят нам о том, каким должен быть актуальный мир, чтобы высказанное утверждение оказалось истинным в этом мире; то есть они специфицируют те *контексты*, в которых оно оказалось бы истинным. К примеру, первичные условия истинности утверждения «вода есть нечто мокрое» специфицируют что-то вроде того, что подобное высказывание будет истинным во множестве миров, водянистая материя в которых является чем-то мокрым.

Если же мы оцениваем используемые термины в соответствии с их вторичными интенсионалами, мы приходим к более привычным *вторичным условиям истинности*. Эти условия специфицируют истинностное значение того или иного утверждения в контрафактических мирах при условии того, что актуальный мир является таким, каков он есть. К примеру, вторичные условия истинности утверждения «вода есть нечто мокрое» (высказанного в нашем мире) специфицируют те миры, в которых вода является мокрой: поэтому при условии того, что вода есть H_2O , они специфицируют те миры, в которых H_2O оказывается чем-то мокрым. Заметим, что здесь нет опасности двусмысленности относительно истинности в актуальном мире: первичные и вторичные условия истинности всегда будут давать одно истинностное значение при их оценке в актуальном мире.

Если мы рассматриваем пропозицию как функцию от возможных миров к истинностным значениям, то два этих набора истинностных значений дают две *пропозиции*, связанные с любым утверждением. Соединение первичных интенсионалов используемых терминов дает *первичную пропозицию*, верную именно в тех контекстах выражения, в которых данное утверждение выражало бы истину. (Это «диагональная пропозиция» Столнейкера (Stalnaker 1978). Строго говоря, это центрированная пропозиция, или функция от центрированных миров к истинностным значениям.) Вторичные интенсионалы дают *вторичную пропозицию*, верную при тех контрафактических обстоятельствах, при которых данное утверждение в том виде, как оно выражается в актуальном мире, является истинным. Вторичная пропозиция — это «содержание» выражения Каплана, и она чаще рассматривается в качестве пропозиции, выражаемой тем или иным утверждением. Однако первичная пропозиция тоже имеет важное значение.

Два этих вида необходимой истинности утверждения точно соответствуют необходимости двух видов связанной с ним пропозиции. Утверждение необходимо истинно в первом (априорном) смысле, если связанная с ним первичная пропозиция верна во всех центрированных возможных мирах (то есть если данное утверждение было бы истинным в любом контексте выражения). Утверждение необходимо истинно в апостериорном смысле, если связанная с ним вторичная

пропозиция верна во всех возможных мирах (то есть если данное утверждение в том виде, в каком оно высказано в *актуальном* мире, является истинным во всех контрафактических мирах). Первый из этих видов соответствует глубокой необходимости Эванса, второй — более привычной поверхностной необходимости.

В качестве иллюстрации возьмем утверждение «вода есть H_2O ». Первичные интенсионалы «воды» и « H_2O » различны, и поэтому мы не можем априори знать, что вода есть H_2O ; связанная с ними *первичная* пропозиция не является необходимой (она верна в тех центрированных мирах, в которых водянистая материя имеет определенную молекулярную структуру). Тем не менее вторичные интенсионалы совпадают, так что «вода есть H_2O » истинно во всех возможных мирах, если оно оценивается сообразно вторичным интенсионалам — то есть связанная с ним *вторичная* пропозиция является необходимой. Крипкеанская апостериорная необходимость появляется именно тогда, когда вторичные интенсионалы некоего утверждения фундируют необходимую пропозицию, а первичные — нет.

Рассмотрим по контрасту утверждение «вода есть водянистая материя». Здесь с «водой» и «водянистой материей» связаны одни и те же первичные интенсионалы, и поэтому при наличии у нас данных понятий мы можем априори знать об истинности этого утверждения. Связанная с ним первичная пропозиция необходима, так что данное утверждение оказывается необходимо истинным в эвансовском «глубоком» смысле. Вторичные интенсионалы, однако же, различны, поскольку «вода» трактуется в жестком смысле, а «водянистая материя» — нет: в мире, где XYZ — прозрачная, пригодная для питья жидкость, вторичный интенсионал «водянистой материи» указывает на XYZ , а вторичный интенсионал «воды» не делает этого. Связанная с данным утверждением вторичная пропозиция поэтому не необходима, и это утверждение не является необходимой истиной в более привычном смысле; оно будет примером крипкеанского «контингентного априори».

В общем, многие мнимые «проблемы», возникающие из этих крипкеанских выкладок, оказываются следствием попыток втиснуть двойственную картину референции в единое понятие значения или необходимости. От этих проблем обычно можно избавиться путем развернутых указаний на двумерный характер референции и детального определения, какое именно понятие значения или необходимости представляет для нас интерес²⁸.

Данную двумерную модель можно также использовать для характеристики семантики *мышления*, равно как и языка. В другом месте я обсуждаю этот вопрос гораздо более подробно (Chalmers 1994c). Этот аспект данной модели не будет играть большой роли в дальнейшем,

но о нем стоит упомянуть, так как пару раз мы все же мимоходом столкнемся с ним. Главная идея очень сходна с той, о которой шла речь выше: при наличии у мыслящего индивида *понятия* мы можем приписать ему первичный интенсионал, соответствующий тому, что оно будет вычленять в зависимости от того, каким окажется актуальный мир, и вторичный интенсионал, соответствующий тому, что оно вычленяет в контрафактических мирах при условии того, что актуальный мир оказывается таким, каков он есть. При наличии *убеждения* мы можем сходным образом приписывать ему первичную и вторичную пропозиции (названные мной в другом месте «понятийным» и «реляционным» содержанием данного убеждения).

К примеру, такие понятия, как «Геспер» и «Фосфор», будут иметь разные первичные интенсионалы (одно выделяет вечернюю звезду в данном центрированном мире, другое — утреннюю звезду), но одинаковые вторичные интенсионалы (оба они выделяют Венеру во всех мирах). Мысль «Геспер есть Фосфор» будет связана с первичной пропозицией, истинной во всех центрированных возможных мирах, где вечерняя звезда есть утренняя звезда: факт информативности, а не триадичности данной мысли соответствует факту контингентности этой первичной пропозиции, объясняющейся различием первичных интенсионалов двух этих терминов.

Первичная пропозиция в большей степени, чем вторичная, передает то, как выглядят вещи с позиции субъекта: она доставляет набор центрированных миров, признаваемых субъектом, имеющим убеждение, в качестве потенциального окружения, в котором он мог бы жить (будучи убежденным, что Геспер есть Фосфор, я признаю все те центрированные миры, в которых вечерняя звезда оказывается тождественной утренней звезде в окрестностях центра). Не очень сложно также показать, что именно первичная, а не вторичная пропозиция отвечает за когнитивные и рациональные отношения между мыслями. Поэтому естественно представлять первичную пропозицию в качестве *когнитивного* содержания мышления²⁹.

Логическая необходимость, концептуальная истина и представимость

Опираясь на эту модель, мы можем артикулировать отношения между логической необходимостью, концептуальной истиной и представимостью.

Начнем с логической необходимости: это и есть необходимость в объясненном выше смысле. Утверждение логически необходимо, если и только если оно истинно во всех логически возможных мирах. Конечно, у нас есть две разновидности логической необходимости утверждений, зависящие от того, оцениваем ли мы истинность в возможном мире

сообразно их первичным или вторичным интенсионалам. Мы могли бы назвать их соответственно *необходимостью-1* и *необходимостью-2*.

Этот анализ эксплицирует логическую необходимость и возможность того или иного *утверждения* в терминах (а) логической возможности *миров* и (б) интенсионалов, определяемых терминами из этого утверждения. Интенсионалы я уже обсуждал. Что же касается понятия логически возможного мира, то оно относится к числу изначальных понятий: как и прежде, мы можем интуитивно представлять логически возможный мир как мир, который мог бы быть создан Богом (оставляя в стороне вопросы о самом Боге). Я не буду касаться болезненного вопроса об онтологическом статусе этих миров, но просто приму их как отправную точку, в качестве инструмента, так же как можно принять за отправную точку математику³⁰. Если же говорить об *объеме* этого класса, то наиболее важным является то, что всякий представимый мир оказывается логически возможным — в самое ближайшее время мы более подробно обсудим это обстоятельство.

Что же касается концептуальной истины, то если мы отождествим значение с интенсионалом (первичным или вторичным), то будет несложно установить связь между истинностью по значению и логической необходимостью. Если утверждение логически необходимо, его истинность будет автоматическим побочным продуктом интенсионалов его терминов (и композиционной структуры данного утверждения). Нам не нужно будет отводить какую-либо дополнительную роль миру, так как интересующие нас интенсионалы будут удовлетворяться в любом возможном мире. Аналогичным образом, если некое утверждение истинно вследствие его интенсионалов, оно будет истинным в любом возможном мире.

Как и прежде, можно говорить о двух разновидностях концептуальной истины — в зависимости от того, отождествляем ли мы «значения» с первичными или вторичными интенсионалами, — параллельных двум разновидностям необходимой истины. Пока мы принимаем параллельные решения в двух этих случаях, положение концептуально истинно, если и только если оно необходимо истинно. «Вода есть водянистая материя» есть концептуально истинное и необходимо истинное утверждение в первом смысле; а «вода есть H_2O » — во втором. Только первая разновидность концептуальной истины будет, как правило, достижима априори. Вторая разновидность будет включать множество апостериорных истин, так как вторичный интенсионал зависит от того, каким окажется актуальный мир.

(Я не утверждаю, что значения могут корректно мыслиться только через интенсионалы. Значение — многогранное понятие, и некоторые из его граней могут несовершенно отображаться интенсионалами;

поэтому не все согласились бы отождествить их, по крайней мере в некоторых случаях³¹. Скорее отождествление значения и интенсиона должно рассматриваться здесь в качестве своего рода условного соглашения: если мы отождествляем их, мы можем извлечь из этого отождествления самые разные полезные следствия. Не так уж много зависит от использования слова «значение». В любом случае истинность по интенсионалу — единственный вид истинности по значению, который мне понадобится.)

При должной осмотрительности мы можем также установить связь между логической возможностью утверждений и их *представимостью*. Можно сказать, что утверждение представимо (или представимо в качестве истинного), если оно истинно в каком-либо представимом мире. Нельзя смешивать этот смысл с другими смыслами «представимого». К примеру, в одном из смыслов утверждение представимо, если, насколько нам известно, оно истинно или если мы не знаем о его невозможности. Если говорить в этом смысле о гипотезе Гольдбаха, то представимым будет как она, так и ее отрицание. Но ложный элемент этой пары не будет считаться представимым в том смысле, который я использую, так как нет представимого мира, в котором он является истинным (он ложен в любом мире).

Согласно этому взгляду на представимость, представимость утверждения предполагает две вещи: во-первых, представимость соответствующего мира и, во-вторых, истинность данного утверждения в этом мире³². Это значит, что, высказывая суждения о представимости, надо удостовериться в корректном описании представляемого мира, надлежащим образом оценив истинность утверждения в данном мире. На первый взгляд могло бы показаться, что ложность гипотезы Гольдбаха представима через представление мира, где математики обзывают ее таковой; но если эта гипотеза на деле истинна, то мы *неверно описываем* этот мир; на деле это мир, где данная гипотеза верна, но какие-то математики ошибаются.

На практике для вынесения суждения о представимости нужно лишь рассмотреть представимую *ситуацию* — небольшую часть мира, — а затем удостовериться в ее корректном описании. Если имеется представимая ситуация, в которой утверждение оказывается истинным, то, разумеется, будет иметься и представимый мир, в котором оно окажется истинным. Так что этот метод будет давать осмысленные результаты, не напрягая при этом наши когнитивные ресурсы необходимостью представления целого мира!

Иногда утверждается, что такие примеры, как «вода есть H_2O », показывают, что из представимости не следует возможность, но, на мой взгляд, вопрос этот не столь однозначен. В действительности есть две

разновидности представимости, которые мы могли бы назвать *представимостью-1* и *представимостью-2* в зависимости от того, оцениваем ли мы какое-то утверждение в представимом мире сообразно первичным или сообразно вторичным интенсионалам входящих в него терминов. Утверждение «вода есть XYZ» представимо-1, так как представим мир, в котором оно (оцениваемое сообразно первичным интенсионалам) будет истинным, но не представимо-2, поскольку нельзя представить мир, в котором данное утверждение (оцениваемое сообразно вторичным интенсионалам) будет истинным. Два этих вида представимости в точности отражают два упомянутых выше вида логической возможности.

Представимость утверждения зачастую отождествляется с представимостью-1 (в этом смысле представимо, что «вода есть XYZ»), поскольку представимость такого рода достижима априори. Возможность же утверждения чаще всего отождествляется с возможностью-2 (в этом смысле невозможно, что «вода есть XYZ»). При таком рассмотрении из представимости не следует возможность. Но это не меняет того, что представимость-1 влечет возможность-1, а представимость-2 — возможность-2. Надо просто быть осмотрительным и не говорить о представимости-1, когда речь идет о возможности-2. Иными словами, надо быть осмотрительным и не давать описание представляемого мира (скажем, мира XYZ) через первичные интенсионалы, когда лучше было бы делать это через вторичные³³.

Из всего этого следует, что часто упоминаемое различие «логической» и «метафизической» возможности, восходящее к крипкеанским примерам, — в соответствии с которым считается логически, но не метафизически возможным, что вода есть XYZ, — это различие не на уровне *миров*, а в лучшем случае на уровне *утверждений*. В этом смысле утверждение «логически возможно», если оно истинно в каком-то мире при его оценке сообразно первичным интенсионалам; и оно «метафизически возможно», если оно истинно в каком-то мире при его оценке сообразно вторичным интенсионалам. Само пространство миров в обоих случаях одинаково³⁴.

Важнее всего то, что ни один из рассмотренных выше случаев не дает оснований полагать, что какие-либо из представимых *миров* невозможны. Любая обеспокоенность относительно провала между представимостью и возможностью касается лишь утверждений, а не миров: либо мы используем какое-то утверждение для некорректного описания представимого мира (как в крипкеанском случае и во втором гольдбаховском случае), либо мы утверждаем, что оно представимо, вообще не представляя мир (как в первом гольдбаховском случае). Кажется поэтому, что нет основания отрицать, что из представимости мира следует его возможность. В дальнейшем я буду исходить

из этого тезиса о логической возможности; любая вариация возможности, когда из представимости не будет следовать возможность, будет, таким образом, более узким классом. Кто-то мог бы счесть, что существует более узкая разновидность «метафизически возможных миров», но основания верить в наличие подобного класса должны были бы быть совершенно независимыми от стандартных оснований, рассмотренных мной. В любом случае при рассмотрении вопросов, связанных с объяснением, на первый план выходит именно логическая возможность. (Более сильная «метафизическая» модальность могла бы в лучшем случае понадобиться при обсуждении онтологии, материализма и т. п. Я буду говорить о ней в надлежащем месте в главе 4.)

Имплицирование в обратном направлении, от логической возможности к представимости, не столь очевидно, так как ограниченность нашей познавательной способности допускает такие возможные ситуации, которые мы не сможем представить, возможно, из-за их громадной сложности. Однако если мы истолкуем представимость как представимость-в-принципе — быть может, как представимость высшим существом, — то кажется правдоподобным, что из логической возможности мира следует представимость этого мира, и поэтому из логической возможности утверждения следует представимость этого утверждения (в надлежащем смысле). В любом случае меня будет больше интересовать другая импликация.

Если утверждение логически возможно или необходимо в соответствии с его первичным интенсионалом, данная возможность или необходимость познаемы априори, по крайней мере в принципе. Модальность не является эпистемически недосягаемой: возможность какого-либо утверждения есть функция входящих в него интенсионалов и пространства возможных миров. И то, и другое в принципе эпистемически досягаемо, и ни то, ни другое не является в данном случае зависимым от апостериорных фактов. Так что возможность-1 и представимость-1 можно изучать, не вставая с кресла. А вот возможность-2 и представимость-2 во многих случаях будут досягаемы только апостериори, так как в определении вторичных интенсионалов могут играть роль факты, касающиеся внешнего мира.

Класс необходимых-1 истин точно соответствует классу априорных истин. Если какое-то утверждение априори истинно, то оно истинно вне зависимости от того, каким оказывается актуальный мир; то есть оно истинно во всех мирах, рассматриваемых в качестве актуального, и поэтому необходимо-1. И наоборот: если утверждение необходимо-1, то оно будет истинным вне зависимости от того, каким оказывается актуальный мир, и поэтому оно будет истинным априори. В большинстве подобных случаев истинность такого утверждения будет познаваться априори; исключениями могут быть некоторые математические

утверждения, истинность которых мы не можем определить, а также утверждения, слишком сложные для того, чтобы мы могли постичь их. Даже в этих случаях кажется, что мы вправе сказать, что они познаются априори хотя бы в принципе, хотя они и выходят за пределы нашей ограниченной познавательной способности. (Я вернусь к этому вопросу, когда он окажется востребован в дальнейшем.)

В зависимости от того, используем ли мы первичную или вторичную разновидность логической необходимости, мы получаем немного разных понятия логической супервентности. Если бы с «жизней» был связан как первичный, так и вторичный интенсионал, то жизнеобразность могла бы быть логически супервентна на физических свойствах в соответствии либо с первичным, либо с вторичным интенсионалом «жижи». Супервентность по вторичному интенсионалу — то есть супервентность с апостериорной необходимостью как релевантной модальностью — соответствует тому, что некоторые называют «метафизической супервентностью», но мы уже видели, что ее можно трактовать как вариант логической супервентности.

(В действительности существует только один вид логической супервентности *свойств*, так же как существует только один вид логической необходимости *пропозиций*. Но мы видели, что термины или понятия на деле определяют два свойства, одно — через первичный интенсионал («водянистая материя»), другое — через вторичный (H_2O). Поэтому по отношению к данному понятию («вода») можно говорить о двух способах возможной супервентности свойств, связанных с этим понятием. Иногда я буду нестрого говорить о первичных и вторичных интенсионалах, связанных со свойством, и о двух способах возможной супервентности свойства.)

Я рассмотрю как первичный, так и вторичный варианты логической супервентности в конкретных случаях, но первый вариант будет более важным. Первичные интенсионалы оказываются более важными, чем вторичные, особенно когда речь идет об объяснении. Как отмечалось выше, в начале исследования мы можем работать только с первичным интенсионалом, и именно этот интенсионал определяет, удовлетворительным ли оказалось объяснение. К примеру, для объяснения воды мы должны объяснить ее прозрачность, текучесть и т. п. Вторичный интенсионал (H_2O) не фигурирует здесь вплоть до завершения объяснения и поэтому сам по себе не устанавливает критерия успешности объяснения. Именно логическая супервентность сообразно первичному интенсионалу определяет, является ли возможным рекрутывное объяснение. И при отсутствии оговорок, указывающих, что

Логическая необходимость и логическая супервентность

речь идет о другом, я буду, как правило, говорить о логической супервентности сообразно первичному интенсионалу.

Если выбрать какую-либо разновидность интенсионала — скажем, первичный интенсионал — и держаться ее, то мы сможем убедиться в эквивалентности разных вариантов формулы логической супервентности. Согласно дефиниции, данной в начале этой главы, В-свойства логически супервентны на А-свойствах, если в любой логически возможной ситуации Y , А-неотличимой от актуальной ситуации X , все В-факты, истинные относительно X , истинны относительно Y . Или, проще, В-свойства логически супервентны на А-свойствах, если в любой актуальной ситуации X А-факты относительно X влечут В-факты относительно X (где « P влечет Q » понимается как «логически невозможно, что P и не- Q »).

В применении к глобальной супервентности это означает, что В-свойства логически супервентны на А-фактах, если В-факты относительно актуального мира извлекаются из А-фактов. Аналогичным образом В-свойства логически супервентны на А-свойствах, если нельзя представить мир с теми же А-свойствами, что и наш, но с иными В-свойствами. Мы можем также сказать, что логическая супервентность имеет место, если при наличии всей совокупности А-фактов A^* и любого В-факта B о нашем мире W « $A^*(W) \rightarrow B(W)$ » истинно в силу значений А-терминов и В-терминов (где под значениями понимаются интенсионалы).

Наконец, если В-свойства логически супервентны на А-свойствах сообразно первичным интенсионалам, то импликация от А-фактов к В-фактам будет иметь априорный характер. Поэтому если кто-то знает все А-факты о какой-то актуальной ситуации, то он в принципе может удостовериться в В-фактах лишь на основании А-фактов, если он обладает В-понятиями, о которых идет речь. На практике такой вывод может оказаться трудным или невозможным из-за сложности рассматриваемых ситуаций, но он по крайней мере возможен в принципе. Что касается логической супервентности сообразно *вторичным* интенсионалам, то в В-фактах о какой-то ситуации тоже можно в принципе удостовериться на основании А-фактов, но только апостериори. А-факты должны дополняться контингентными фактами об актуальном мире, поскольку такие факты будут участвовать в формировании соответствующих В-интенсионалов.

Таким образом, существует по меньшей мере три пути, следуя которым можно установить логическую супервентность; они связаны с представимостью, эпистемологией и анализом. Чтобы установить, что В-свойства логически супервентны на А-свойствах, мы можем: (1) попытаться доказать, что реализация А-свойств непредставима без

реализации В-свойств; (2) попытаться доказать, что тот, кто располагает А-фактами, мог бы узнать В-факты (по крайней мере в случаях супервентности сообразно первичным интенсионалам) или (3) проанализировать интенсионалы В-свойств с детализацией, достаточной для того, чтобы стало ясно, что В-утверждения вытекают из А-утверждений в силу одних лишь этих интенсионалов. Это работает и в плане установления отсутствия логической супервентности. Доказывая главные тезисы, связанные с логической супервентностью, я воспользуюсь всеми этими тремя методами.

Рассуждения об эквивалентности различных формулировок логической супервентности могли убедить не всех, и поэтому, обосновывая важные выводы, связанные с логической супервентностью, я буду выдвигать несколько вариантов аргументов, используя каждую из этих формулировок. Подобные действия позволят нам убедиться в силе этих аргументов и несущественности тонких различий разных понятий супервентности.

5. Почти все логически супервентно на физическом*

В следующей главе я буду доказывать, что сознательный опыт не является чем-то логически супервентным на физическом и поэтому не может быть редуктивно объяснен. На это нередко отвечают, что сознательный опыт не одинок в этом и что самые разные свойства не являются логически супервентными на физическом. Такие разные свойства, как стольность, жизнь и экономическое процветание, не находятся, как полагают, в логическом отношении к фактам об атомах, электромагнитных полях и т. п. И разве можно сомневаться, что подобные высокоуровневые факты не могли бы логически извлекаться из микрофизических фактов?

Думаю, что при тщательном анализе нетрудно увидеть, что это не так и что эти высокоуровневые факты (глобально) логически супервентны на физическом — в той мере, в какой они вообще факты³⁵. Сознательный опыт почти уникален в плане отсутствия логической супервентности. Отношение между сознанием и физическими фактами — иного рода, чем обычное отношение между высокоуровневыми и низкоуровневыми фактами.

Выявить логическую супервентность большинства свойств на физических свойствах можно по-разному. Здесь я буду говорить только о свойствах, характеризующих *естественные феномены* — то есть о контингентных аспектах мира, которые нуждаются в объяснении. Свойство быть ангелом могло бы и не быть логически супервентным на

физическом, но у нас нет оснований верить в ангелов, так что отсутствие этой супервентности не должно беспокоить нас. Я также не буду обсуждать факты, касающиеся таких абстрактных сущностей, какими являются математические объекты и пропозиции, так как они заслуживают отдельного рассмотрения³⁶.

Следует отметить, что, утверждая супервентность большинства высокоуровневых свойств на физическом, я не хочу сказать, что высокоуровневые факты и законы следуют из микрофизических законов или даже из микрофизических законов в сочетании с микрофизическими ограничительными условиями. Этот тезис был бы слишком сильным, и хотя он мог бы оказаться вполне правдоподобным при надлежащей интерпретации, в пользу него пока еще нет свидетельств. Я высказываю гораздо более слабый тезис о том, что высокоуровневые факты являются следствием совокупности микрофизических *фактов* (возможно, в сочетании с микрофизическими законами). Это громадное по своему охвату множество включает факты о распределении всех частиц и полей во всех уголках пространства-времени: от атомов в шляпе Наполеона до электромагнитных полей в кольце Сатурна. Фиксация этого множества фактов оставляет, как мы увидим, очень мало места для изменений в чем-то другом.

Прежде чем переходить к аргументам, упомяну ряд безвредных причин отсутствия логической супервентности на физическом. Во-первых, некоторые высокоуровневые свойства не являются логически супервентными из-за зависимости от сознательного опыта. Не исключено, что сознательный опыт отчасти конституирует, к примеру, такое свойство, как любовь. От феноменальных качеств могут, как мы увидим, зависеть и первичные (хотя и не вторичные) интенсионалы, связанные с некоторыми внешними свойствами, такими как цвет и тепло. Если это так, то любовь, а возможно и тепло, не являются логически супервентными на физическом. Их нельзя рассматривать в качестве контрпримеров к моему тезису, так как они не выявляют новой лакуны в логической супервентности. И сам тезис, возможно, лучше всего выразить, сказав, что все факты логически супервентны на комбинации физических и феноменальных фактов или что все факты логически супервентны на физических фактах, *если не принимать во внимание вклада, привносимого сознательным опытом*. Сходным образом зависимость от сознательного опыта может препятствовать редуктивной объяснимости некоторых высокоуровневых феноменов, но мы тем не менее можем говорить, что они редуктивно объяснимы, если не принимать во внимание вклада сознательного опыта.

Во-вторых, как мы видели раньше, применение ряда первичных — но не вторичных — интенсионалов не обходится без *индексикального*

элемента. Первичный интенсионал «воды», к примеру, это что-то вроде «прозрачной, пригодной для питья жидкости в нашем окружении», так что если в актуальном мире имеется водянистая H_2O и водянистая XYZ , то ответ на вопрос о том, что из них считается водой, зависит от того, в каком окружении субъект использует этот термин. Поэтому в некоторых случаях мы, в принципе, должны добавлять к супервентностной базе *центр*, репрезентирующий локализацию субъекта. Это дает нам логическую супервентность и редуктивное объяснение, если не принимать во внимание вклада, вносимого сознательным опытом и индексикальностью.

Против логической супервентности, наконец, не свидетельствуют и случаи неопределенности высокоуровневых фактов. Тезис состоит лишь в том, что если высокоуровневые факты определены, то они определяются физическими фактами. Если сам мир не может фиксировать высокоуровневые факты, нельзя ожидать, что физические факты сделают это. Кто-нибудь мог бы подумать, что логическая супервентность не соблюдалась бы при наличии двух в равной степени хороших высокоуровневых теорий мира, различающихся в своем описании высокоуровневых фактов. Одна теория могла бы, к примеру, утверждать, что вирусы — живые существа, другая — что нет, и поэтому факты о жизни не определяются физическими фактами. Это, однако, не контрпример, а случай, в котором факты о жизни оказываются неопределенными. При неопределенности мы вправе употреблять термины по-разному в зависимости от обстоятельств. Но если факты *определенны* — к примеру, если истинно то, что вирусы есть живые существа, — то одно из описаний попросту оказывается неверным. Так или иначе, в той мере, в какой факты относительно какой-либо ситуации вообще являются определенными, они вытекают из физических фактов.

Повсеместность логической супервентности я буду доказывать, используя аргументы, апеллирующие к представимости, к эпистемологическим соображениям, а также к анализу понятий, имеющих отношение к делу.

Представимость. Логическую супервентность большинства высокоуровневых фактов проще всего заметить при использовании представимости в качестве теста на логическую возможность. Какого рода мир мог бы быть идентичен нашему во всех без исключения микрофизических фактах и при этом быть биологически отличным от него? Скажем, у какого-то вомбата в нашем мире есть два детеныша. Физические факты о нашем мире будут включать факты о распределении всех частиц в пространственно-временном участке, соответствующем вомбату, а также о его детенышах, их окружении и их эволюционной истории. Если бы некий мир разделял с нашим миром все эти факты,

но не был бы миром, в котором у этого вомбата было бы два детеныша, в чем могло бы состоять различие между ними? Подобный мир кажется совершенно непредставимым. Если в возможном мире фиксируются все те же физические факты, то автоматически фиксируются и факты о вомбатности и родительстве. Эти биологические факты не являются вещами такого рода, которые могут отрываться от их физических основ, даже в виде концептуальной возможности.

Это же справедливо и относительно архитектурных фактов, астрономических фактов, поведенческих фактов, химических фактов, экономических фактов, метеорологических фактов, социологических фактов и т. д. Нельзя помыслить мир, физически тождественный нашему, но отличающийся от нашего в подобных фактах. Представляя микрофизически тождественный мир, мы представляем мир с той же самой локализацией каждой частицы в пространстве и времени. Из этого следует, что этот мир будет иметь ту же самую макроскопическую структуру, что и наш мир, и ту же самую макроскопическую динамику. После фиксации всего этого попросту не остается места для изменчивости данных фактов (за исключением, возможно, изменения, идущего от изменчивости сознательного опыта).

Более того, эта непредставимость, похоже, не связана с какими-то контингентными ограничениями нашей познавательной способности. Подобный мир непредставим *в принципе*. Даже высшее существо или Бог не могли бы вообразить такой мир. Им просто нечего было бы воображать. После того как они вообразили бы мир со всеми его физическими фактами, они автоматически вообразили бы мир, в котором имеются все те высокоуровневые факты. Физически тождественный мир, в котором эти высокоуровневые факты неверны, таким образом, логически невозможен, и указанные высокоуровневые свойства логически супервентны на физическом.

Эпистемология. Оставляя позади интуиции, связанные с предста- вимостью, мы можем заметить, что *если бы был* возможен мир, физически идентичный нашему, но биологически отличный от него, это породило бы радикальные эпистемологические проблемы. Как мы узнали бы, что мы не находимся в том мире, а находимся в этом? Как мы узнали бы, что биологические факты в нашем мире таковы, какими они являются? Чтобы уяснить проблему, обратим внимание на то, что если бы я находился в этом альтернативном мире, то он наверняка *выглядел бы* в точности как наш. Он реализует такое же распределение частиц, как и то, которое обнаруживается в растениях и животных в этом мире; неотличимые конфигурации фотонов отражаются от этих объектов; даже при самом тщательном исследовании не удалось бы выявить никаких различий. Из этого следует, что всех

имеющихся в нашем распоряжении внешних свидетельств недостаточно для того, чтобы провести различие между этими возможностями. В той мере, в какой биологические факты относительно нашего мира не являются логически супервентными, мы оказываемся не в состоянии узнать об этих фактах, исходя из внешних свидетельств.

В действительности, однако, не существует глубокой эпистемологической проблемы, связанной с биологией. Мы всегда познаем биологические факты о нашем мире на основании внешних свидетельств, и здесь не возникает какой-то особой скептической проблемы. Из этого следует, что биологические факты логически супервентны на физическом. Это же справедливо для архитектурных, экономических и метеорологических фактов. Не существует какой-то особой скептической проблемы, связанной с познанием этих фактов на основании внешних свидетельств, и поэтому они должны быть логически супервентными на физическом.

Мы можем подкрепить этот тезис, заметив, что в тех областях, где *имеются* эпистемологические проблемы, им сопутствует отсутствие логической супервентности, и наоборот, там, где нет логической супервентности, мы видим наличие каких-то эпистемологических проблем.

Наиболее очевидный пример — эпистемологическая проблема, касающаяся сознания, а именно проблема других сознаний. Эта проблема возникает потому, что со всеми внешними свидетельствами кажется логически совместимым как то, что окружающие нас существа обладают сознанием, так и то, что они не обладают им. Мы не можем, к примеру, заглянуть в мозг собаки и увидеть там присутствие или отсутствие сознательного опыта. Статус этой проблемы спорен, но уже одного *prima facie* существования этой проблемы достаточно для опровержения эпистемологического аргумента в пользу логической супервентности сознания, параллельного тем, что были представлены выше. В отличие от этого случая, не существует даже *prima facie* проблемы других биологий или других экономик. Те факты характеризуются прямой публичной доступностью именно потому, что они фиксируются физическими фактами.

(Вопрос: почему сходный аргумент не вынуждает нас заключить, что если сознательный опыт не является логически супервентным, то мы не можем знать даже о нашем *собственном* сознании? Ответ: потому что сознательный опыт находится в самом центре нашего эпистемического универсума. Скептические проблемы относительно несупервентных биологических фактов возникают потому, что мы имеем доступ к биологическим фактам только посредством внешних, физически опосредованных свидетельств; внешние несупервентные факты выходили бы за пределы нашей прямой эпистемической доступности. В случае нашего собственного сознания подобной проблемы не существует.)

Другая знаменитая эпистемологическая проблема связана с фактами, касающимися каузальности. Как доказывал Юм, внешние свидетельства обеспечивают нам доступ лишь к регулярностям в последовательности событий, но не каким-то дополнительным фактам о каузальности. Поэтому если каузальность конструируется как что-то дополнительное к регулярности (как, допускаю, и должно быть), то неясно, можем ли мы знать о ее существовании. Эта скептическая проблема опять-таки сопряжена с отсутствием логической супервентности. В данном случае факты, касающиеся каузальности, лишены логической супервентности на характеристиках конкретного физического факта. При наличии всех фактов о распределении физических сущностей в пространстве и времени логически возможным является то, что все существующие в них регулярности возникают в результате громадной космической случайности без какой бы то ни было реальной каузальности. Если взять более локальный случай, то при наличии конкретных фактов о том, что якобы является примером каузальности, логически возможно, что речь идет просто о последовательности. Мы делаем вывод о каузальности, следуя путем заключения к наилучшему объяснению — полагать иначе значило бы верить в громадные, необъяснимые совпадения, — но вера в каузальность не вынуждается у нас тем прямым способом, каким вынуждается у нас вера в биологию.

Я обошел проблемы, связанные с супервентностью каузальности, допустив, что супервентностная база в нашем случае включает не только конкретные физические факты, но и все физические законы. Мы вправе предположить, что добавление законов фиксирует факты о каузальности. Но, разумеется, по отношению к законам возникает скептическая проблема, параллельная скептической проблеме причинности: вспомним юмовскую проблему индукции и логическую возможность того, что все то, что кажется законом, могло бы оказаться случайной регулярностью.

Насколько я могу судить, две эти проблемы исчерпывают эпистемологические проблемы, возникающие в результате отсутствия логической супервентности на физическом. Есть и другие эпистемологические проблемы, которые в известном смысле предшествуют упомянутым, так как они касаются самого существования физических фактов. Во-первых, это декартовская проблема существования внешнего мира. С нашими опытными данными совместимо несуществование мира, который, как мы думаем, мы видим; возможно, у нас галлюцинация, или мы — мозги в бочке. Можно показать, что эта проблема возникает именно потому, что факты о внешнем мире не являются логически супервентными на фактах о нашем опыте. (Идеалисты, позитивисты и другие пытались доказать, что они супервентны, но их доказательства

спорны. Заметим, что, если принять упомянутые воззрения, скептическая проблема исчезает.) Другая эпистемологическая проблема связана с теоретическими сущностями, постулируемыми наукой, — электронами, кварками и т. п. Их отсутствие было бы логически совместимо с напрямую наблюдаемыми фактами об объектах в нашем окружении, и из-за этого некоторые высказывали скептические сомнения относительно подобных сущностей. Эта проблема, как можно показать, возникает из отсутствия логической супервентности теоретических фактов на фактах наблюдения. В обоих этих случаях скептические сомнения, возможно, лучше всего гасить, действуя путем заключения к наилучшему объяснению, как и в случае каузальности, но принципиальная возможность того, что мы ошибаемся, при этом не устраивается.

В любом случае я обхожу подобные скептические проблемы, по-просту исходя из физического мира и фиксируя все физические факты об этом мире в его супервентностном фундаменте (допуская тем самым существование внешнего мира, электронов и т. п.). При наличии знания о подобных фактах не остается места для скептических сомнений в большинстве высокоуровневых фактов именно в силу их логической супервентности. Представляя это в обратной последовательности: все наши источники внешних свидетельств логически супервентны на микрофизических фактах, и поэтому если некий феномен не супервентен на этих фактах, внешние свидетельства не могут дать нам основание для веры в него. Можно спросить, нельзя ли полагать какие-то дополнительные феномены через заключение к наилучшему объяснению, как это делалось выше, для объяснения микрофизических фактов. Действительно, этот процесс идет от конкретных фактов к простым законам, лежащим в их основании (и тем самым дает каузальность), но затем он, похоже, останавливается. В самой природе фундаментальных законов заложено то, что они являются последним звеном в цепи объяснений (за исключением, возможно, теологических спекуляций). Это оставляет феномены, относительно которых у нас есть *внутренние* свидетельства — а именно сознательный опыт, — и всё. Если отвлечься от вклада, привносимого сознательным опытом, то все феномены логически супервентны на физическом.

Эпистемологический аргумент в пользу логической супервентности можно выстроить и более прямолинейно, доказывая, что тот, кто располагает всеми физическими фактами, в принципе может узнать все высокоуровневые факты при условии, что у него имеются все высокоуровневые понятия, которые с ними связаны. Конечно, на практике такой субъект, возможно, никогда не смог бы удостовериться в этих высокоуровневых фактах, исходя из данного множества микрофизических фактов. Громадность этого множества достаточна для того, чтобы

исключить такую возможность. (Еще меньше я хочу сказать, что он мог бы вывести их формальным способом; формальные системы нерелевантны в данном случае по причинам, которые были указаны выше.) Но если говорить в принципе, то можно самыми разными способами показать, что субъект (высшее существо?), вооруженный только микрофизическими фактами и связанными с ними понятиями, мог бы вывести высокоуровневые факты.

Простейший способ сделать это — заметить, что в принципе можно было бы сконструировать большую ментальную симуляцию мира и, так сказать, умозрительно наблюдать за ней. Допустим, какой-то мужчина идет под зонтиком. Из связанных с этой ситуацией микрофизических фактов мы можем напрямую вывести факты о распределении и химическом составе материи в окружении этого мужчины, давая высокоуровневую структурную характеристику этой области. Можно было бы без особых экивоков удостовериться в существовании телесного двуногого существа мужского пола. К примеру, из структурной информации можно было бы сделать вывод о наличии организма на двух длинных ногах, ответственных за его передвижение, о мужской анатомии этого существа и т. д. Стало бы ясно, что он несет устройство, предотвращающее попадание на него капель воды, в большом количестве присутствующих в этом месте. Сомнения относительно того, что это действительно зонтик, могли бы быть рассеяны наблюдением, что его физическая структура позволяет ему закрываться и раскрываться; что его история такова, что утром оно висело на вешалке и что оно было изначально сделано на фабрике вместе с другими подобными устройствами, и т. п. Сомнения относительно того, что телесное двуногое существо — действительно человек, могли бы быть рассеяны наблюдениями о составе его ДНК, о его эволюционной истории, о его родстве с другими существами и т. п. Нам нужно лишь допустить, что наш субъект достаточно владеет соответствующим понятием, чтобы быть в состоянии корректно применять его к конкретным случаям (то есть что он располагает его интенсионалом). Если это так, то микрофизические факты будут содержать все данные, нужные ему для того, чтобы применить эти понятия и определить, что тут действительно человек идет под зонтиком.

Сказанное справедливо практически для всякого вида высокоуровневых феноменов: столов, жизни, экономического процветания. Зная все низкоуровневые факты, субъект в принципе может вывести все факты, необходимые для определения того, является ли данная ситуация примером реализации того свойства, о котором идет речь. Фактически, при этом конструируется возможный мир, совместимый с данными микрофизическими фактами, и высокоуровневые факты

просто вычитываются из этого мира при использовании надлежащего интенсионала (поскольку релевантные факты в физически идентичных возможных мирах неизменны). Значит, высокоуровневые факты логически супервентны на физическом.

Анализируемость. До сих пор я доказывал, что микрофизические факты фиксируют высокоуровневые факты, особо не детализируя высокоуровневые понятия, которые при этом используются. В любом конкретном случае, однако, это отношение выводения базируется на интенсионале понятия. Если микрофизические факты влекут за собой высокоуровневый факт, то это происходит потому, что микрофизических фактов достаточно для фиксации тех черт мира, благодаря которым возможно применение высокоуровневого интенсионала. Иначе говоря, мы должны быть в состоянии *проанализировать*, при каких условиях некая сущность соответствует интенсионалу высокоуровневого понятия, по крайней мере до такой степени, чтобы мы увидели, почему эти условия соответствия могут быть удовлетворены путем фиксации физических фактов. Полезно поэтому более внимательно взглянуть на интенсионалы высокоуровневых понятий и изучить те черты мира, благодаря которым они оказываются применимыми.

На пути прояснения этих интенсионалов и их словесного выражения существует ряд препятствий. Как мы видели ранее, зачастую условия применения понятия отчасти неопределенны. Является ли чашеобразный предмет, сделанный из ткани, чашей? Является ли компьютерный вирус живым? Будет ли книгой случайно возникший книгообразный объект? Наши обыденные понятия не дают прямых ответов на эти вопросы. В известном смысле это вопрос соглашения. Отсюда следует отсутствие определенных условий применения, которые могли бы быть использованы в процессе выводения. Однако, как мы видели ранее, эта неопределенность в точности отображает неопределенность, касающуюся самих фактов. В той мере, в какой интенсионал «чаша» является предметом соглашения, предметом соглашения оказываются и факты о чашах. Для наших целей важно, что интенсионал вместе с микрофизическими фактами детерминирует высокоуровневые факты — в той мере, в какой они действительно фактичны. Размытость и неопределенность могут затруднить обсуждение, но они не оказывают какого-либо существенного влияния на сами проблемы.

С этим связана и другая проблема, состоящая в том, что любой краткий анализ понятия с неизбежностью будет неточным. Как мы видели, у понятий, как правило, нет четких дефиниций. В первом приближении мы можем сказать, что объект является столом, если у него есть плоская горизонтальная поверхность с поддерживающими ее ножками; но это слишком широко (представим франкенштейнов-

ског монстра на ходулях) и слишком узко (как быть со столом без ножек, выступающим из стены?). Можно уточнять эту дефиницию, добавлять новые условия и оговорки, но мы быстро столкнемся с проблемами неопределенности — и в любом случае результат, который мы получим, никогда не будет совершенным. Нет, однако, необходимости уточнять все детали, требующиеся для того, чтобы справиться с каждым конкретным случаем: в какой-то момент детали начинают попросту повторяться. Если мы знаем, благодаря какому *типу* свойств применяется интенсионал, мы будем в состоянии понять суть дела.

Как мы видели ранее, для выводимости В-фактов из А-фактов не требуется дефиниция В-свойств в терминах А-свойств. Значения существенным образом представляются интенсионалами, а не дефинициями. Роль анализа сводится здесь просто к такой детализации характеристики интенсионалов, которой было бы достаточно для выявления подобной выводимости. Для этого подойдет и грубый анализ. Интенсионалы обычно применяются к индивидам в возможном мире благодаря одним, а не другим их свойствам; задача подобного анализа в том, чтобы понять, благодаря какого рода свойствам применяется интенсионал, и показать, что такие свойства могут быть выводимы из физических свойств.

Третья проблема возникает из различия априорных и апостериорных условий применимости многих понятий. Впрочем, если мы обособляем первичные и вторичные интенсионалы, эта проблема не столь уж серьезна. Вторичный интенсионал, связанный с «водой», есть что-то вроде $\langle H_2O \rangle$, и он очевидным образом логически супервентен на физическом. Но и первичный интенсионал, что-то вроде «прозрачная, пригодная для питья жидкость в нашем окружении», тоже логически супервентен, так как прозрачность, пригодность для питья и жидкая природа воды выводимы из физических фактов³⁷. Мы можем трактовать этот вопрос и так, и так. Как мы видели, в редуктивных объяснениях задействуется первичный интенсионал, так что именно он интересует нас в наибольшей степени. В общем, если первичный интенсионал I логически супервентен на физическом, то таковым будет и прошедший процедуру придания жесткости вторичный интенсионал $dthat(I)$, так как он будет, как правило, представлять собой проекцию некоторой внутренней физической структуры на все другие миры.

Соображения относительно апостериорной необходимости привели некоторых к предположению о невозможности логической выводимости высокоуровневых фактов из низкоуровневых фактов. Обычно говорят примерно так: «Вода с необходимостью H_2O , но это не истина по значению, и поэтому здесь нет никакого концептуального отношения». Но это колоссальное упрощение. Начать с того, что вторичный интенсионал $\langle H_2O \rangle$ в известном смысле может рассматриваться как

часть значения «воды» и что он уж точно логически супервентен. Но еще более важно то, что первичный интенсионал («прозрачная, пригодная для питья жидкость...»), фиксирующий референцию, тоже супервентен, пусть и с поправкой на сознание и индексикальность. Именно из-за соответствия этому интенсионалу мы изначально и сочли, что H_2O — это вода. При наличии первичного интенсионала I высокоуровневые факты без каких-либо проблем выводимы из микрофизических фактов (если отвлечься от вклада, привносимого опытом и индексикальностью). Наблюдение Крипке, что данное понятие может быть лучше представлено в качестве $dthat(I)$, никак не затрагивает этой выводимости. Семантический феномен придания жесткости сам по себе не производит какого-либо онтологического различия.

Устранив эти препятствия, мы можем обратиться к интенсионалам, связанным с различными высокоуровневыми понятиями. В большинстве случаев они характеризуются функциональными или структурными терминами или же сочетанием тех и других. К примеру, для того чтобы нечто было столом, требуется: (1) чтобы объект имел плоскую поверхность и поддерживался ножками и (2) чтобы люди использовали его для размещения на нем каких-то вещей. Первое — структурное условие, то есть условие, касающееся внутренней физической структуры данного объекта. Второе — функциональное условие, связанное с внешней каузальной ролью этой сущности, характеризующее то, как она взаимодействует с другими сущностями. Структурные свойства очевидным образом выводимы из микрофизических фактов. В целом это верно и для функциональных свойств, хотя эта выводимость не столь прямолинейна. Подобные свойства зависят от гораздо более широкой супервентностной базы микрофизических фактов, и поэтому зачастую здесь будут важны факты об окружении объекта; а в той мере, в какой такие свойства характеризуются диспозиционально (нечто растворимо, если бы оно растворилось при погружении в воду), приходится прибегать к контрфактуалам. Но истинностные значения этих контрфактуалов фиксируются включением физических законов в антецеденты наших супервентностных кондиционалов, так что это не проблема.

Если взять другой пример, то условия жизни можно примерно свести к некоей комбинации, среди прочего, способностей размножения, приспособления и осуществления метаболизма (как обычно, нам нет нужды говорить о точных пропорциях или учитывать все другие релевантные факторы). Все эти свойства могут быть охарактеризованы функционально, в терминах отношения некой сущности к другим сущностям, ее способности превращать внешние ресурсы в энергию, а также ее способности надлежащим образом реагировать на свое окружение. Все эти функциональные свойства в принципе могут быть выведены из физических фактов. И опять-таки, даже если

не существует совершенной дефиниции жизни в функциональных терминах, подобного рода характеристика показывает нам, что жизнь есть функциональное свойство, и это значит, что его реализация может рассматриваться как нечто выводимое из физических фактов.

Трудность в том, что функциональные свойства нередко характеризуются в терминах каузальной роли относительно других высокуюровневых сущностей. Из этого следует, что логическая супервентность свойств зависит от логической супервентности других имеющих к ним отношение высокуюровневых понятий, которые сами могут получать функциональную характеристику. В конечном счете это не является проблемой, поскольку каузальные роли в итоге отыгрываются нефункциональными свойствами: обычно либо структурными, либо феноменальными. Во взаимной дефинируемости различных функциональных свойств может обнаружиться определенного рода циркулярность — не исключено, что скрепкосшиватель отчасти определяется тем, что он выдает скрепки, а скрепки — тем, что они выдаются скрепкосшивателем. С этой циркулярностью можно справиться одновременным избавлением от всех каузальных ролей всех свойств³⁸, что можно сделать, если наши анализы содержат какой-то нециркулярный компонент, в конечном счете базирующийся на структурных или феноменальных свойствах. (Может показаться, что апелляция к феноменальным свойствам несовместима с логической супервентностью на физическом, но см. ниже. В любом случае она совместима с логической супервентностью с поправкой на влияние сознательного опыта.)

Многие свойства характеризуются реляционно, в терминах отношения данного объекта к своему окружению. Такие отношения, как правило, каузальны, так что эти свойства оказываются функциональными, хотя это и не всегда так: возьмем, к примеру, свойство быть на том же континенте, что и некая утка. Некоторые свойства подобным образом зависят от истории (хотя обычно их можно сконструировать в каузальном ключе); чтобы быть кенгуру, существо должно иметь надлежащих предков. В любом случае эти свойства не создают проблем для логической супервентности, так как релевантные исторические факты и факты, касающиеся окружающей среды, сами будут фиксироваться глобальными физическими фактами.

Даже такие комплексные общественные факты, как «1950-е годы были временем экономического расцвета»³⁹, могут быть по большей части охарактеризованы в функциональных терминах и поэтому могут рассматриваться как выводимые из физических фактов. Полный анализ был бы очень запутанным и трудноосуществимым из-за неопределенности понятия расцвета, но чтобы представить, как он мог бы проводиться, можно задать себе вопрос: почему мы говорим, что 1950-е годы были временем экономического расцвета? В первом приближении

мы говорим так потому, что эти годы были отмечены высокой занятостью, люди могли позволить себе совершать необычно большие покупки, инфляция была низка, строилось много домов и т. п. И мы можем провести схематичные анализы понятий «дом» (место, где люди спят и едят), «занятость» (организованный труд за вознаграждение), а также монетарных понятий (так, деньги можно схематично проанализировать в терминах возможности осуществления систематического обмена на другие вещи, а их ценность может быть проанализирована в терминах количества того, что мы получаем взамен). Все эти анализы выглядят до смешного упрощенными, но суть достаточно ясна. Всё это, как правило, функциональные свойства, которые могут быть выведены из физических фактов.

Многие высказывали сомнения относительно возможности концептуального анализа. Зачастую их причины — к примеру, из-за неопределенности наших понятий или из-за отсутствия их четких дефиниций — не могли бы оказать влияния на мои аргументы. Иногда этот скепсис мог возникать по более глубоким причинам. Тем не менее если то, что я говорил выше в этой главе, верно и если физические факты о возможном мире фиксируют высокоуровневые факты, то мы должны *ожидать* того, что мы будем в состоянии проанализировать интенционал интересующего нас высокоуровневого понятия, по крайней мере приблизительно, чтобы увидеть, как его применение может быть определено физическими фактами. Именно это я пытался сделать в приведенных здесь примерах. В сходном ключе могут быть рассмотрены и другие примеры⁴⁰.

Я не защищаю программу проведения подобных анализов как таковую. Понятия слишком сложны и непослушны, чтобы она принесла какую-то особую пользу, и любой развернутый анализ — обычно лишь бледная тень реальности. Но важно то, что большинство высокоуровневых понятий не являются чем-то изначальным и недоступным для анализа. Как правило, их можно проанализировать до такой степени, при которой их интенционалы могут быть представлены в качестве спецификаций функциональных или структурных свойств. Благодаря этой открытости для анализа высокоуровневые факты оказываются в принципе выводимыми из микрофизических фактов и редуктивно объяснимыми в их терминах.

Есть несколько видов свойств, которые, как можно было бы подумать, создают особые трудности для логической супервентности, а поэтому и для редуктивного объяснения. Я рассмотрю ряд кандидатов на эту роль, обращая особое внимание на то, аналогичны ли

Некоторые проблемные случаи

проблемы, создаваемые для редуктивного объяснения соответствующими феноменами, тем, которые порождаются для него сознанием. Думаю, что, не считая пары возможных исключений, здесь не возникает новых важных проблем.

Свойства, зависимые от сознания. Как уже обсуждалось, первичные интенсионалы некоторых понятий включают отношение к сознательному опыту. Одним из наглядных примеров является краснота, рассматриваясь в качестве свойства внешних объектов. По крайней мере согласно некоторым толкованиям, первичный интенсионал, связанный с краснотой, требует в качестве критерия красноты вещи то, чтобы она имела тенденцию вызывать переживание красного при надлежащих условиях⁴¹. Поэтому в своем первичном интенсионале краснота не является логически супервентной на физическом, хотя она и супервентна, если не принимать во внимание вклада, привносимого сознательным опытом. Однако если взять ее вторичный интенсионал, то он почти наверняка супервентен. Если оказывается, что в актуальном мире тенденция вызывать переживание красного присуща определенной отражательной способности поверхности, то объекты с такой отражательной способностью будут красными даже в мирах, где нет сознательных существ, которые могли бы их увидеть. Краснота апостериори отождествляется с этой отражательной способностью, логически супервентной на одном лишь физическом.

Ранее мы видели, что отсутствие логической супервентности первичного интенсионала сопровождается невозможностью редуктивного объяснения. Значит ли это, что в случае красноты редуктивное объяснение невозможно? Ответ — да, в слабом смысле. Если краснота конструируется как тенденция вызывать переживание красного, то в той мере, в какой подобный опыт не может быть редуктивно объяснен, не может быть редуктивно объяснена и краснота. Но можно приблизиться к такому объяснению. Можно сказать, что определенное физическое качество вызывает переживание красного; и можно даже объяснить каузальное отношение, существующее между этим качеством и *суждениями о красном*. Необъясненным остается лишь последний шаг к опыту. На практике наши требования к объяснению достаточно лояльны для того, чтобы это имело значение. Для объяснения феномена, референция которого фиксируется опытом, мы не требуем объяснения опыта. Иначе наше ожидание затянулось бы.

Сказанное верно и для таких феноменов, как тепло, свет и звук. Хотя их вторичные интенсионалы определяют структурные свойства (движение молекул, присутствие фотонов, звуковые волны), их первичные интенсионалы включают в себя отношение к сознательному опыту: тепло — это то, что вызывает переживание тепла, свет порождает

визуальный опыт и т. д. Но, как отмечали Нагель (Nagel 1974) и Серл (Searle 1992), при объяснении тепла мы не нуждаемся в объяснении ощущений тепла. Вполне достаточно и объяснения без учета опыта.

Другие свойства еще более непосредственно зависят от сознательного опыта — в том плане, что опыт не только играет определенную роль в фиксации их референции, но и частично конституирует апостериорное понятие. Наглядный пример — свойство нахождения рядом с сознательной личностью. Согласно некоторым толкованиям, такие ментальные свойства, как любовь и убеждение, не будучи сами по себе феноменальными, находятся в концептуальной зависимости от существования сознательного опыта. Если это так, то в мире, где нет сознания, подобные свойства не могли бы быть реализованы. Это означает, что они не являются логически супервентными даже апостериори, и их редуктивное объяснение оказывается невозможным даже в более сильном смысле, чем в предыдущих случаях. Они, однако, логически супервентны и редуктивно объяснимы, если не принимать во внимание вклада, привносимого сознательным опытом, и поэтому здесь не возникает нового провала в редуктивном объяснении.

Интенциональность. Стоит особо рассмотреть статус интенциональности, так как иногда говорят, что она порождает проблемы, аналогичные тем, которые порождает сознание. Вполне вероятно, однако, что любое отсутствие логической супервентности интенциональных свойств производно от несупервентности сознания. Как я отмечал в главе 1, кажется невозможным представить мир, физически и феноменально идентичный нашему, но отличный по своим интенциональным содержаниям⁴². Если, как полагают некоторые философы, феноменология частично конституирует интенциональное содержание, то интенциональные свойства могут и не быть логически супервентными на физическом, оставаясь при этом супервентными, если не принимать во внимание вклада, привносимого сознательным опытом. Тезис о том, что сознание частично конституирует содержание, спорен, но в любом случае нет серьезных оснований считать, что интенциональность несупервентна особым, непроизводным образом.

Если отвлечься от любых феноменальных аспектов, то интенциональные свойства лучше всего представлять в виде некоей конструкции с позиции третьего лица, служащей объяснению человеческого поведения. Это означает, что они должны допускать анализ в терминах каузальных связей с поведением и окружающей средой. Если это так, то интенциональные свойства напрямую логически супервентны на физическом. Льюис (Lewis 1974) предпринимает основательную попытку показать выводимость интенциональных фактов из физических фактов путем надлежащего функционального анализа. Более поздние

концепции интенциональности, в частности те, которые были предложены Деннетом (Dennett 1987), Дретске (Dretske 1981) и Фодором (Fodor 1987), могут быть истолкованы в качестве того, что содействует реализации того же самого проекта. Ни один из предложенных ими анализов не выглядит абсолютно убедительным, но не исключено, что новые вариации на ту же тему устраният их недоделки. Не существует аргумента, аналогичного аргументу против супервентности сознания, который показывал бы, что интенциональность *не может* быть логически супервентной на физических и феноменальных свойствах⁴³. В действительности аргументы от представимости указывают на то, что интенциональные свойства должны быть логически супервентными на них, если подобные свойства вообще реализованы. К сходному выводу ведут нас и эпистемологические аргументы. Так что не существует особой онтологической проблемы интенциональности.

Моральные и эстетические свойства. Нередко утверждается, что не существует концептуального звена, ведущего от физических к моральным и эстетическим свойствам. Согласно Муру (Moore 1922), значение таких понятий, как «доброта», не допускает необходимости выведения фактов о доброте из физических фактов. На деле Мур утверждал, что не существует концептуального звена, позволяющего переходить от естественных фактов к моральным фактам, причем естественное может включать как ментальное, так и физическое (поэтому здесь не может помочь отсылка к супервентности с поправкой на сознательный опыт). Означает ли это, что моральные свойства столь же проблематичны, как и сознательный опыт?

Аналогия между ними, однако, не проходит по двум причинам. Во-первых, похоже, нельзя представить мир, идентичный нашему в естественном смысле, но морально отличный от него, и поэтому маловероятно, что моральные факты являются дополнительными в сильном смысле этого слова. Во-вторых, моральные факты — это не такие феномены, которые не оставляют нам выбора относительно того, принимать их или нет. Если нас прижимают к стенке, мы можем вообще отрицать существование моральных фактов. Именно такова стратегия моральных антиреалистов, таких как Блэкберн (Blackburn 1971) и Хейр (Hare 1984). Эти антиреалисты доказывают, что, поскольку моральные факты невыводимы из естественных фактов и, вероятно, не являются «странными» дополнительными фактами, они не имеют объективного существования — и мораль должна быть релятивизирована и превращена в конструкцию или проекцию нашего когнитивного аппарата. Эту стратегию нельзя использовать по отношению к феноменальным свойствам, существование которых не вызывает сомнений.

Для моральных свойств имеется по меньшей мере две осмысленные альтернативы. Первая — некий антиреализм, состоящий, к примеру, в релятивизации «объективных моральных фактов» в «субъективные моральные факты»⁴⁴ или же занимающий позицию, согласно которой в моральном дискурсе вообще не говорится о фактах. Вторая альтернатива состоит в допущении априорной связи между естественными и моральными фактами, которая (в противовес Муру) могла бы быть установлена средствами анализа и экспликации моральных понятий. Если такое понятие, как «добро», определяет стабильный неиндексикальный первичный интенсионал, то следствием этого оказывается вторая позиция: мы будем располагать априорной функцией от естественно специфицированных миров к моральным фактам. Если это понятие определяет лишь индексикальный первичный интенсионал, или если субъекты вправе связывать с ним разные первичные интенсионалы, или если оно вообще не определяет первичного интенсионала, то будет верна какая-то из разновидностей первой позиции.

Иногда предлагаются и другие решения, но они не кажутся удачными. Мур считал, что существует неконцептуальная априорная связь между естественными и моральными фактами, раскрываемая таинственной способностью «моральной интуиции», но очень многие отвергают такую позицию (трудно понять, что могло бы служить основанием истинности или ложности подобных интуиций). Позиция, согласно которой моральные свойства супервентны благодаря наличию какого-то базового номического звена, кажется неприемлемой, так как нельзя представить мир с теми же естественными фактами, как в нашем, но с иными моральными фактами. Среди современных моральных реалистов популярна позиция (см., напр., Boyd 1988; Brink 1989), в соответствии с которой моральные факты супервентны на естественных фактах с апостериорной необходимостью, то есть супервентны сообразно вторичным, но не первичным интенсионалам моральных понятий. Эту позицию, однако, непросто удержать, так как даже апостериорные отождествления должны базироваться на априорной фиксации референции. Несмотря на то что вода есть H_2O апостериори, факты о воде априори следуют из микрофизических фактов. Сходным образом, если моральные понятия имеют первичный интенсионал и если идентичные в естественном плане центрированные миры морально идентичны, то из этого, похоже, следует существование априорного звена, связующего естественные и моральные факты (Хорган и Тимmons (Horgan and Timmons 1992a, 1992b) критикуют эту позицию в подобном ключе).

Сходным образом можно рассуждать и об эстетических свойствах. Во всяком случае антиреалистская трактовка выглядит здесь еще более убедительной. Так что, хотя с размышлениями о том, как следует

трактовать области морального и эстетического, и связаны интересные концептуальные вопросы, они не создают метафизических и объяснительных проблем, сравнимых с теми, которые ставит перед нами сознательный опыт.

Имена. Многие утверждают (см., напр., Kaplan 1989), что имена, такие как «Рольф Харрис», не увязаны с каким-либо анализом и что они напрямую указывают на свои референты. Означает ли это, что свойство быть Рольфом Харрисом не является логически супервентным на физическом? Относительно супервентности вторичного интенсионала здесь нет проблемы (к примеру, Рольф мог бы быть человеком, возникшим из данного сперматозоида и данной яйцеклетки во всех возможных мирах), но можно было бы подумать, что отсутствие первичного интенсионала создает проблемы для редуктивного объяснения. Правдоподобно тем не менее, что даже при отсутствии общепризнанного первично-го интенсионала с каждым конкретным употреблением этого имени связан первичный интенсионал. Когда я употребляю имя «Рольф Харрис», его референт *неким* систематическим образом зависит от того, каким является мир; для меня этот первичный интенсионал мог бы быть чем-то вроде «человек, именуемый „Рольфом Харрисом“», барабанящий по банкам с краской и находящийся в таком-то каузальном отношении ко мне»⁴⁵. Подобный интенсионал будет логически супервентным. Однако вместо детального обосновывания сказанного проще заметить, что отсутствие логической супервентности не будет здесь со-пряженено с объяснительной тайной. Свойство быть Рольфом Харрисом не составляет феномен, нуждающийся в объяснении, которое следует отличать от экспликации. В объяснении нуждается существование человека по имени «Рольф Харрис», барабанящего по банкам с краской и т. д. Эти свойства наверняка супервентны и в принципе могут быть объяснены способом, который обычно используется в таких случаях.

Индексикалы. Фиксация референции многих понятий, от «воды» до «моей собаки», содержит в себе индексикальный элемент. Референция этих понятий фиксируется на основе как физических фактов, так и соотнесенного с субъектом «индексикального факта», репрезентирующего позицию субъекта, использующего данный термин. Подобный факт является чем-то определенным для данного субъекта, и поэтому определенной оказывается и фиксация референции. Супервентность и объяснение проходят, если сделать поправку на влияние этого индексикального факта.

Создает ли индексикальность трудности для редуктивного объяснения? Для субъектов вообще, возможно, нет, так как от этого «факта» можно избавиться. Но в случае самого себя это не так просто. Индексикальный факт выражает нечто очень важное об известном мне мире:

что Дэвид Чалмерс — это я. Как можно было бы объяснить этот, как кажется, элементарный факт? И есть ли тут на самом деле факт, который надо объяснять, а не простая тавтология? Эта проблема крайне трудна для понимания, но мне представляется, что даже если индексикал не является объективным фактом о мире, это факт о мире, насколько он известен мне, и в объяснении нуждается мир в этом своем аспекте. Но природа простых индексикалов весьма туманна, и совершенно неясно, как можно было бы объяснить их⁴⁶. (Разумеется, мы можем редуктивно объяснить, почему фраза Дэвида Чалмерса «Я — Дэвид Чалмерс» истинна. Но этот неиндексикальный факт кажется совершенно отличным от индексикального факта, что я Дэвид Чалмерс.)

Заманчиво в этой связи было бы обратиться к сознанию. Но хотя объяснение сознания могло бы объяснить «точки зрения» вообще, трудно понять, как оно могло бы объяснить, почему одна из этих точек зрения, которая выглядит как нечто произвольное, — *моя* точка зрения, если, конечно, не признавать истинность солипсизма. Не исключено, что индексикальный факт надо рассматривать как что-то изначальное. Если это так, то редуктивное объяснение окажется здесь невозможным, как оно невозможно и для сознания, причем эти случаи различны. Впрочем, отсутствие редуктивного объяснения будет тут не столь тревожным, как в случае сознания, так как этот не допускающий объяснения факт выглядит слишком уж «худосочным» в сравнении с фактами о сознании во всем его величии. Допущение этого элементарного индексикального факта потребует гораздо менее существенного пересмотра нашего материалистического мировоззрения, чем признание нередуцируемых фактов о сознательном опыте.

Негативные факты. Как мы видели ранее, некоторые факты, включающие негативные экзистенциалы и кванторы всеобщности, не являются логически детерминированными физическими фактами, да и вообще любым множеством локализованных фактов. Рассмотрим следующие факты о нашем мире: ангелов не существует; Дон Брэдман — величайший игрок в крикет; все живое основано на ДНК. Все их можно фальсифицировать без изменения физических фактов о нашем мире, путем простого добавления новой нефизической сущности, к примеру ангелов, состоящих из эктоплазмы и играющих в крикет. И здесь не поможет даже добавление фактов о сознательном опыте или индексикальности⁴⁷.

Означает ли это, что данные факты не являются редуктивно объяснимыми? Похоже, что да, если не существует физического объяснения того, почему в нашем мире отсутствуют добавочные нефизические сущности. Это и в самом деле дополнительный факт. Лучший способ справиться с этой ситуацией — ввести второпорядковый факт,

говорящий о множестве базовых конкретных фактов, любых — микрофизических, феноменальных, индексикальных и т. д.: *это всё*. Этот факт говорит о том, что все конкретные факты о мире включены в это множество фактов или выводимы из него. Из указанного второрядкового факта, в сочетании со всеми базовыми конкретными фактами, будут вытекать все негативные факты.

Невозможность редуктивного объяснения не имеет здесь такого уж большого значения. Вполне возможно, что факт «и это всё» будет верен в любом мире, и подобный факт никогда нельзя будет вывести из конкретных фактов. Он просто выражает сущностную ограниченность нашего мира или любого мира. И так, без особых затрат, мы можем ухватить все негативные экзистенциальные и всеобщие квантифицированные факты.

Физические законы и каузальность. Согласно наиболее убедительным концепциям, объясняющим природу физических законов, последние не являются чем-то логически супервентным на физических фактах, рассматриваемых в качестве собрания конкретных фактов об истории мира в пространстве и времени. В этом можно убедиться, обратив внимание на логическую возможность мира, физически неотличимого от нашего на протяжении всей его пространственно-временной истории, но имеющего иные законы. К примеру, один из законов такого мира мог бы быть таким: при скоплении двухсот тонн чистого золота в вакууме происходит его превращение в свинец. В остальном его законы такие же, с небольшими модификациями там, где это необходимо. Однако в пространственно-временной истории нашего мира не было случаев, когда в вакууме скапливалось двести тонн золота. Из этого следует, что, хотя наш мир и этот другой мир имеют одинаковые истории, их законы тем не менее различны.

Подобные аргументы наводят на мысль, что законы природы не являются чем-то логически супервентным на множестве конкретных физических фактов⁴⁸. Действуя в сходном ключе, можно показать, что каузальная связь между двумя событиями — нечто большее, чем их регулярность. Сторонники разного рода юмистских взглядов оспаривают эти выводы, но мне кажется, что в данном случае им не хватает аргументов⁴⁹. В существовании законов и каузальности имеется что-то нередуцируемое.

В других местах я обходил эти проблемы, включая физические законы в супервентностную базу, но это означает, что мы скорее перешагиваем через метафизическую головоломку, чем решаем ее. Конечно, законы и каузальность выявляют не столь серьезный провал в редуктивном объяснении, какой выявляется в нем благодаря сознанию. Законы и каузальные отношения сами вводятся для объяснения

физических феноменов, а именно множества регулярностей, существующих в природе, тогда как сознание есть исходная данность, нуждающаяся в объяснении. Тем не менее само существование подобных нередуцируемых дополнительных фактов порождает глубокие вопросы об их метафизической сути. Если не считать сознательный опыт и, возможно, индексикальность, то они оказываются единственными дополнительными фактами такого рода, в которые у нас есть хоть какое-то основание верить. И естественным было бы предположить, что два этих вида несупервентного, сознание и каузальность, могут находиться в тесном метафизическом отношении друг к другу.

Итак, мы остались с тем, что почти все факты логически супервентны на физических фактах (включающих физические законы). Возможными исключениями являются сознательный опыт, индексикальность и негативные экзистенциальные факты. Иначе говоря, можно утверждать, что факты о мире исчерпываются (1) конкретными физическими фактами, (2) фактами о сознательном опыте, (3) законами природы, (4) второпорядковым фактом «и это всё» и, возможно, (5) индексикальным фактом о моей локализации. (Два последних незначительны в сравнении с другими, и статус последнего из них сомнителен, но я включаю их для полноты картины.) Если не принимать во внимание вклад, привносимый сознательным опытом и индексикальностью, то дело выглядит таким образом, будто все позитивные факты логически супервентны на физическом. Для окончательности этого вывода потребовалось бы более детальное исследование всех видов феноменов, но и увиденного нами достаточно, чтобы счесть его резонным. Мы можем суммировать онтологическую и эпистемологическую ситуации, рассказав парочку вымышленных историй. Не исключено, что пусть и не в деталях, но в обрисе этих историй можно было бы отыскать зерно истины.

Резюме

Миф о творении. Создавая мир, Бог должен был бы лишь зафиксировать только что упомянутые факты. Для максимальной экономии усилий он начал с фиксации законов природы — законов физики и тех законов, которые соотносят физику с сознательным опытом. Потом он зафиксировал ограничительные условия: к примеру, временной срез физических фактов или значения, выдаваемые генератором случайных чисел. Он объединил их с законами для фиксации остальных физических и феноменальных фактов. И в самом конце он провозгласил: «И это всё».

Эпистемологический миф. Вначале у меня имеются только факты о моем сознательном опыте. Отсюда я вывожу факты о средней величины объектах в мире, а в конечном счете и микрофизические факты.

Из регулярности этих фактов я заключаю к физическим законам, а значит, и к новым физическим фактам. Из регулярных соотношений моего сознательного опыта и физических фактов я заключаю к психофизическим законам, а значит и к фактам, касающимся сознательного опыта других существ. Мне кажется, что я достиг пределов абдукции, и я предполагаю, что это всё. Мир гораздо больше, чем вначале казалось, и поэтому я выделяю изначальный сознательный опыт, обозначая его в качестве *моего опыта*.

Отметим совершенно разную последовательность действий в этих двух перспективах. Эпистемология кажется чуть ли не онтологией наоборот. Отметим также, что Бог, похоже, не в силах зафиксировать мой индексикальный факт. Возможно, это еще один повод усомниться в этом факте.

Тезис о логической супервентности большинства высокоуровневых феноменов не получил такого широкого признания, какое он мог бы получить, даже среди тех, кто дискутирует на тему супервентности. Хотя этот момент часто остается без обсуждения, многие подозрительно относились к тому, чтобы рассматривать концептуальную модальность в качестве надлежащего инструмента для работы с отношениями супервентности. Насколько я могу судить, эта нерешительность была вызвана рядом причин, ни одна из которых в конечном счете не кажется мне весомой.

Во-первых, проблема логически возможных физически идентичных миров с разным добавочным нефизическим содержанием (ангелами, эктоплазмой) подтолкнула некоторых авторов к предположению о том, что отношения супервентности не могут быть логическими (Haugeland 1982; Petrie 1987), но мы уже видели, как можно решить эту проблему. Во-вторых, многие полагали, что соображения относительно апостериорной необходимости показывают невозможность выражения отношений супервентности через значения (Brink 1989; Teller 1984). Мы видели, однако, что отношения супервентности, основанные на апостериорной необходимости, могут рассматриваться в качестве разновидности логической супервентности. В-третьих, существует общий скепсис по поводу понятия концептуальной истины, идущий от Куайна; но мы видели, что здесь он неуместен. В-четвертых, беспокойства в связи с «редуцируемостью» заставляли некоторых предположить, что супервентность в целом не является концептуальным отношением (Hellman and Thompson 1975). Несмотря, однако, существуют ли серьезные аргументы против редуцируемости, которые были бы серьезными аргументами и против логической супервентности. В-пятых, сам феномен сознательного опыта иногда используется для демонстрации

того, что отношения супервентности в целом не могут иметь логический характер (Seager 1988); но мы видели, что сознательный опыт является чуть ли не уникальным примером отсутствия логической супервентности. Наконец, тезис о том, что отношения супервентности не являются в целом логическими, зачастую высказывается без каких бы то ни было доказательств, как то, в чем, как предполагается, должен быть убежден любой разумный человек (Bacon 1986; Heil 1992)⁵⁰.

Вполне вероятно, что любое отношение супервентности высокоуровнего свойства на физическом в конечном счете оказывается либо (1) отношением логической супервентности в его первичной или вторичной разновидности, либо (2) контингентным отношением естественной супервентности. Если некий кандидат на роль отношения супервентности не подходит под эти случаи, у нас есть серьезное основание считать, что здесь вообще отсутствуют объективные высокоуровневые факты (возможно, что именно так обстоит дело с моральными фактами). В главе 4 я приведу дополнительные аргументы в пользу несуществования сколь-либо глубокой разновидности супервентности, занимающей место между логической супервентностью и естественной супервентностью.

В принципе, это дает нам унифицированную объяснительную картину. Почти каждый феномен является редуктивно объяснимым в слабом смысле, очерченном ранее. Исключением является сознательный опыт и, возможно, индексикальность, наряду с предельными микрофизическими фактами и законами, которые надо признать фундаментальными.

Здесь стоит на мгновение задержаться и ответить на вопрос, поставленный Блэкберном (Blackburn 1985) и Хорганом (Horgan 1993): как мы объясним сами отношения супервентности? Если речь идет об отношении логической супервентности, основанном на первичном интенсионale понятия, для этого нужно лишь провести надлежащий анализ понятия, возможно, в функциональных или структурных терминах, и отметить, что его референция инвариантна в физически идентичных мирах. Здесь сам супервентностный кондиционал является априорной концептуальной истиной. Если речь идет об отношении супервентности, основанном на вторичном интенсионale, то супервентность можно объяснить, заметив, что первичный интенсионал интересующего нас понятия выделяет некий референт в актуальном мире, инвариантно проецирующийся (посредством придания жесткости) на физически идентичные миры. Всё, что здесь нужно нам для объяснения, — это априорный концептуальный анализ в сочетании с контингентными фактами, связанными с актуальным миром⁵¹. Что же касается отношения сугубо естественной супервентности, то оно само будет контингентным законом. В лучшем случае оно будет объяснимо в терминах

более фундаментальных законов, в худшем — закон супервентности сам будет иметь фундаментальный характер. Как бы то ни было, мы будем объяснять некие регулярности в мире при помощи фундаментальных законов, так же как мы делаем это в физике, и, как всегда, наше объяснение должно будет остановиться на фундаментальных законах. Мы видели, что сугубо естественная супервентность — дорогое удовольствие в онтологическом плане, так что нам еще повезло, что правилом является логическая супервентность, а естественная есть лишь исключение из него.

Часть II

НЕРЕДУЦИРУЕМОСТЬ СОЗНАНИЯ

Глава 3

Можно ли редуктивно объяснить сознание?

1. Является ли сознание логически супервентным на физическом?

Почти всё в мире можно объяснить в физических терминах, и естественно ожидать, что сознание тоже могло бы объясняться подобным образом. В этой главе я, однако, попытаюсь показать, что сознание ускользает из сети редуктивного объяснения. В одних лишь физических терминах мы никогда не сможем объяснить появление сознательного опыта. Этот вывод может показаться негативным, но он ведет к сильным позитивным следствиям, которые будут раскрыты мной в других главах.

Для демонстрации отсутствия редуктивного объяснения мы должны показать, что сознание не является чем-то логически супервентным на физическом. В принципе, мы должны показать, что оно не является *глобально* супервентным — то есть что факты о сознании не вытекают из всей совокупности микрофизических фактов мира. На практике аргументацию проще проводить на *локальном* уровне, доказывая, что микрофизические факты, касающиеся того или иного индивида, не влекут факты о сознании. Когда речь идет о сознании, локальная и глобальная супервентность, похоже, разделяют одну и ту же судьбу,

так что не имеет большого значения, каким образом будет разворачиваться наш аргумент: если сознание вообще супервентно, оно почти наверняка будет локально супервентным. Но если это оспаривается, то все аргументы могут быть развернуты и на глобальном уровне, для чего не потребуется никаких хитроумных изменений.

Как можно показать, что сознание не является логически супервентным на физическом? Есть разные пути. Мы можем поразмышлять о том, что может быть представлено, чтобы напрямую заключить от этого к логической возможности ситуации, в которой имеются точно такие же физические факты, хотя факты, касающиеся опыта, оказываются иными. Мы можем апеллировать к эпистемологии, доказывая, что между знанием о физических фактах и знанием о сознании отсутствует надлежащее связующее звено. И мы можем напрямую апеллировать к понятию сознания, доказывая, что никакой его анализ не может обосновать логическую зависимость феноменального от физического. В дальнейшем я приведу аргументы, использующие все три упомянутые стратегии. Первые два аргумента — это, по сути, аргументы от представимости, вторые два — от эпистемологии, а пятый аргумент есть аргумент от анализа. В этих пяти аргументах есть некая избыточность, но вместе они составляют сильный довод.

Можно действовать и более прямым способом, доказывая отсутствие редуктивного объяснения без развернутого обращения к логической супервентности. Я избирал этот путь в другом месте, но здесь я предприму более детализированный анализ для придания полноты своему аргументу. Тем не менее довод против редуктивного объяснения и критика существующих редуктивных объяснений (в параграфе 2 ниже) должны сохранять свое значение даже без этого анализа. И читатели при желании могли бы напрямую обратиться к ним, по крайней мере при первом чтении.

(Техническое замечание: задачей этой главы, по сути, является доказательство отсутствия априорной выводимости феноменальных фактов из физических фактов. Необходимость, характеризующая соответствующее отношение супервентности, — это априорная версия логической необходимости, центральное место в которой занимают первичные интенсионалы. Как мы видели во второй главе, именно это отношение релевантно для проблем с объяснением; от моментов, связанных с апостериорной необходимостью, можно отвлечься. В следующей главе, где на первый план выйдут скорее онтологические, чем объяснительные проблемы, я попытаюсь отдельно показать, что между физическими и феноменальными фактами не существует какой-либо апостериорной необходимой связи.)

Наиболее очевидный (хотя и не единственный) способ исследования логической супервентности сознания — рассмотрение логической возможности зомби: кого-то или чего-то, физически идентичного мне (или любому другому сознательному существу), но абсолютно лишенного сознательного опыта¹. На глобальном уровне мы можем рассматривать логическую возможность зомбийного мира — мира, физически идентичного нашему, но при полном отсутствии сознательного опыта. В таком мире все будут зомби.

Посмотрим теперь на моего зомбийного двойника. Эта тварь является моей молекулярной копией и идентична мне во всех своих низкоуровневых свойствах, постулированных совершенной физикой, но она полностью лишена сознательного опыта. (Кому-то могло бы показаться более уместным говорить о зомби «оно», но я использую личностное местоимение — я успел привязаться к своему зомбийному двойнику.) Для определенности мы можем вообразить, что в данный момент я выглядываю из окна, ощущаю приятную зелень деревьев на улице, получаю удовольствие от шоколадки, которую жую, и чувствую ноющую боль в своем правом плече.

А что происходит в моем зомбийном двойнике? Он физически идентичен мне, и мы вполне можем допустить, что он находится в идентичном окружении. Он будет наверняка идентичен мне функционально: он будет обрабатывать аналогичную информацию, аналогичным образом реагировать на стимулы с соответствующими изменениями его внутренних конфигураций и неотличимым от моего итоговым поведением. Он будет идентичен мне и психологически — в смысле, указанном в главе 1. Он будет воспринимать деревья за окном, в функциональном смысле, и чувствовать вкус шоколада, в психологическом смысле. Всё это логически следует из того факта, что он физически идентичен мне, в силу функциональной анализируемости психологических понятий. Он будет даже обладать «сознанием» в очерченных ранее функциональных смыслах — он будет бодрствовать, сможет отчитаться о содержании своих внутренних состояний, сфокусировать внимание на разных местах и т. п. Но ни один из этих процессов не будет сопровождаться реальным сознательным опытом. Там не будет феноменального чувства. Нельзя ощутить, каково это — быть зомби.

Зомби такого рода заметно отличается от зомби из голливудских фильмов, у которых обычно наблюдаются серьезные функциональные нарушения (рис. 3.1). Голливудским зомби явно недостает психологического сознания: у них, как правило, недостаточно развита способность к интроспекции и отсутствует выраженный произвольный контроль за собственным поведением. Что же касается феноменального

Аргумент 1:

Логическая

возможность зомби



Рис. 3.1. Кальвин и Гоббс о зомби. (Calvin and Hobbes © Watterson. Distributed by Universal Press Syndicate. Перепечатка разрешена. Все права сохранены.)

(4) — Конечно, настоящие зомби никогда не хохочут, глядя друг на друга.

сознания, то оно может отсутствовать у них, хотя может и присутствовать; как замечает Блок (Block 1995), логично предположить, что они ощущают вкус поедаемых ими жертв. Мы можем назвать их *психологическими зомби*, но меня интересуют *феноменальные зомби*, физически и функционально идентичные, но лишенные опыта. (Быть может, и неудивительно, что феноменальные зомби не пользовались популярностью в Голливуде, так как по понятным причинам снять о них фильм было бы непростой задачей.)

Идея зомби, как я описал их, выглядит странно. Начать с того, что зомби едва ли являются чем-то естественно возможным. В реальном мире любая моя копия, скорее всего, будет обладать сознанием. Поэтому совершенно естественно представлять бессознательных тварей физически отличными от сознательных — к примеру, с дефективным поведением. Но вопрос состоит не в том, вероятно ли существование зомби в нашем мире, и даже не в том, естественной ли будет идея зомбийной копии, а в том, является ли понятие зомби концептуально когерентным. И для того чтобы убедиться в этом, достаточно уже одной мыслимости данного понятия.

Логическую возможность нельзя доказать напрямую. Как, к примеру, можно было бы доказать логическую возможность моноцикла высотой в милю? Такая возможность кажется попросту очевидной. Хотя ничего подобного в реальном мире не существует, описание не вызывает сомнений в плане его когерентности. Если кто-то возразит, сказав, что такая вещь логически невозможна — и лишь кажется логически возможной, — то едва ли мы сможем что-то сказать в ответ, разве что повторить описание и заявить о его очевидной когерентности. Кажется совершенно несомненным, что в данном описании не таятся какие-то скрытые противоречия.

Признаюсь, что логическая возможность зомби кажется мне столь же очевидной. Зомби — это просто нечто физически идентичное мне, но лишенное сознательного опыта, где всё темно внутри. Хотя подобное, скорее всего, эмпирически невозможно, трудно сомневаться в том, что здесь описывается когерентная ситуация; я не могу заметить какого-либо противоречия в этом описании. В известном смысле утверждение данной логической возможности сводится к грубой интуиции, но подобным образом обстоит дело и в случае с моноциклом. Мне кажется, что почти все могут представить эту возможность. Кто-то мог бы оказаться вынужден отрицать данную возможность для оправдания какой-то теории, но обоснование подобных теорий должно опираться на вопрос о возможности, а не наоборот.

В общем, бремя доказательства лежит на тех, кто утверждает, что та или иная дескрипция является логически *невозможной*. Если кто-то действительно уверен, что моноцикл высотой в милю логически невозможен, он должен указать, в чем тут состоит явное или скрытое противоречие. Если он не сможет указать на те моменты интенсионалов понятий «высота в милю» и «моноцикл», которые могли бы привести к противоречию, то его довод не будет убедительным. Не более убедительным, с другой стороны, был бы и очевидно неверный анализ понятий, о которых идет речь, к примеру утверждение о том, что нечто может считаться моноциклом лишь в том случае, если его параметры меньше, чем у статуи Свободы. Если никакой разумный анализ интересующих нас терминов не указывает на противоречие или хотя бы не делает его вероятным, то естественным будет допустить, что здесь можно говорить о логической возможности.

При всём том имеются и позитивные моменты, которые могут быть использованы в качестве подкрепления своей позиции теми, кто отстаивает логическую возможность. Они могут выдвигать разного рода косвенные аргументы, апеллируя к тому, что нам известно о тех феноменах, о которых идет речь, и к тому, что именно мы думаем о гипотетических случаях, касающихся подобных феноменов, дабы установить, что очевидная логическая возможность действительно есть логическая возможность и действительно очевидна. Так, можно сfantазировать, что обычный человек едет на моноцикле, и вдруг вся система расширяется в тысячу раз. Или можно было бы описать ряд моноциклов, каждый из которых оказывается больше предыдущего. В известном смысле всё это апелляции к интуиции, и оппонент, желающий отрицать данную возможность, может всякий раз говорить, что наша интуиция ввела нас в заблуждение, однако сама очевидность описанного нами говорит в нашу пользу и помогает еще в большей степени сместить бремя доказательства на другую сторону.

Тезис о логической возможности зомби может быть, к примеру, косвенно подкреплен рассмотрением *нестандартных реализаций* моей функциональной организации². Моя функциональная организация — то есть схема каузальной организации, воплощенная в механизмах, ответственных за продуцирование моего поведения, — в принципе может быть реализована множеством весьма странных способов. Возьмем широко известный пример (Block 1978): люди, составляющие большую нацию, вроде китайцев, могли бы организоваться таким образом, что они реализовывали бы каузальную организацию, изоморфную той, что имеется в моем мозге, причем каждый человек симулировал бы поведение одного-единственного нейрона, а синапсы соответствовала бы связь через радиопередатчики. Эта нация могла бы осуществлять контроль над пустой головой робота, снабженного сенсорными датчиками и моторными эффекторами.

Многие считают маловероятным, что такая структура порождала бы сознательный опыт — что в этой системе возник бы «коллективный ум». Меня, однако, сейчас не интересует вопрос, *действительно* ли в таком случае возникнет сознательный опыт; подозреваю, что да, как я доказываю в главе 7. Важным пока является лишь *когерентность* идеи об отсутствии у подобной системы сознательного опыта. Здесь выражена осмысленная возможность, и открытым вопросом остается то, возникает при этом сознание или нет. К аналогичным выводам можно прийти и рассматривая мой кремниевый изоморф, организованный так же, как я, но имеющий кремниевые чипы вместо нейронов. Будет ли подобный изоморф *действительно* обладать сознанием, неочевидно, но большинство считает, что отрицающие это говорят о когерентной возможности. Из подобных случаев можно заключить, что существование моего сознательного опыта логически невыводимо из фактов о моей функциональной организации.

Но если признать концептуальную когерентность того, что у структуры с групповым умом или у моего кремниевого изоморфа могло бы обнаружиться отсутствие сознательного опыта, то из этого следует, что мой зомбийный двойник — в равной степени когерентная возможность. Очевидно ведь, что сознание ничуть не в большей степени *концептуально* выводимо из биохимии, чем из кремния или из какой-то группы гомункулов. Если можно представить кремниевого изоморфа без сознательного опыта, то нам нужно лишь в нашем представлении заменить кремний на нейроны, оставив неизменной функциональную организацию, и мы получим моего зомбийного двойника. Ничто в этой замене не могло бы привнести опыт в данное представление; указанные реалиационные различия попросту не являются тем, что могло бы иметь концептуальное отношение к опыту. Значит, сознание не является логически супервентным на физическом.

Аргумент в пользу зомби можно сформулировать и без обращения к этим нестандартным реализациям, но их эвристическая ценность состоит в устранении источника концептуальной путаницы. Интуиции о логической возможности бессознательной физической копии поначалу могут казаться совсем не очевидными, возможно, потому, что привычное сосуществование биохимии и сознания может наводить на мысль о концептуальной связи. Рассмотрение менее привычных случаев лишает картину этих эмпирических корреляций и поэтому позволяет более непосредственно судить о логической возможности³. Но признав логическую возможность этих бессознательных функциональных копий, мы не сможем избежать аналогичного вывода и относительно той самой физической копии.

Кто-то мог бы подумать, что аргументы от представимости недостоверны. Иногда, к примеру, высказывается возражение, что в действительности мы не в состоянии детально вообразить многие миллиарды нейронов человеческого мозга. Конечно, это так; но для нашего аргумента не требуется воображать каждый из этих нейронов. Одна лишь множественность нейронов не могла бы концептуально влечь за собой сознание; если вся эта нейронная структура должна иметь отношение к сознанию, то она должна иметь указанное отношение *вследствие* неких высокоуровневых свойств, которые она делает возможными. Достаточно поэтому вообразить данную систему на грубом уровне и убедиться в том, что мы представляем ее с надлежащим образом детализированными механизмами восприятия, категоризации, высокого пропускного доступа к информационным данным, способности давать отчет и т. п. Вне зависимости от того, насколько детализированными представляются эти механизмы, сценарий с зомби не утрачивает своей когерентности. Быть может, оппонент мог бы заявить, что ускользающие от воображения нейронные детали неким образом концептуально релевантны независимо от их вклада в реализацию сложных функций, но в таком случае он должен был бы разъяснить, в чем могла бы состоять эта релевантность, — но у него нет путей сделать это. Эти реализационные детали попросту находятся не на том уровне, на котором они могли бы оказаться концептуально релевантными для сознания.

Иногда утверждается также, что представимость не может быть надежным проводником для установления возможности. Представимость и возможность расходятся главным образом в случаях апостериорной необходимости. Так, гипотеза, что вода не есть H_2O , кажется концептуально когерентной, но вода, похоже, есть H_2O во всех возможных мирах. Однако апостериорная необходимость не имеет отношения к проблемам данной главы. Как мы видели в предыдущей главе, объяснительные связи основаны на априорной выводимости

высокоуровневых фактов из физических фактов. Соответствующий тип возможности должен оцениваться на основании использования первичных интенсионалов релевантных терминов, а не вторичных интенсионалов, имеющих отношение к апостериорной необходимости. Поэтому если зомбийный мир представим, хотя бы в том смысле, в каком можно представить, что вода не является H_2O , то этого достаточно для вывода о невозможности редуктивного объяснения сознания.

Если отвлечься от указанных соображений, то аргументы от представимости могут оказаться ошибочными главным образом в случае неявной концептуальной путаницы: при недостаточной отрефлексированности мы можем не заметить некогерентности в том, что якобы возможно, — когда мы представляем некую ситуацию и *неверно описываем* ее. К примеру, кто-то мог бы подумать, что он в состоянии представить ситуацию, в которой последняя теорема Ферма оказывается ложной, вообразив ситуацию, когда ведущие математики заявляют, что нашли контрпример. Но если допустить, что эта теорема в действительности верна, то мы видим, что данная ситуация была неверно описана: на самом деле речь идет о сценарии, в котором последняя теорема Ферма верна, а какие-то математики ошибаются. Важно, однако, что подобного рода ошибки всегда имеют априорный характер, так как они возникают из неверного применения первичных интенсионалов наших понятий к представленной ситуации. Надлежащая рефлексия сможет выявить, что данные понятия были неверно применены и что тезис о логической возможности неоправдан.

Таким образом, единственный путь, открытый для оппонента в нашем случае, состоит в утверждении, что при описании зомбийного мира как зомбийного мира мы неверно применяем понятия и что в этом описании в действительности заложено скрытое противоречие. Быть может, если бы мы подумали о нем с достаточной степенью ясности, мы осознали бы, что, воображая физически идентичный мир, мы тем самым *автоматически* воображаем мир, в котором существует сознательный опыт. Но в таком случае оппонент должен был бы указать нам, где именно в том, что кажется вполне когерентным описанием, могло бы скрываться противоречие. Если же внутренняя некогерентность не может быть раскрыта, то это очень серьезный довод в пользу логической возможности зомбийного мира.

Как и прежде, я не могу заменить никакой внутренней некогерентности; у меня есть ясная картина представляемого мной при представлении зомби. И все же некоторые выказывают недоверие аргументам от представимости, особенно если в них используются столь странные идеи, как здесь. Хорошо поэтому, что всё, что можно извлечь из доводов с зомби, можно получить и другими способами,

к примеру рассматривая вопрос с эпистемологической стороны или со стороны анализа. Подобные аргументы (вроде приведенных ниже аргументов 3–5) покажутся многим менее замысловатыми и поэтому более весомыми в плане обоснования тезиса об отрицании логической супервентности. Однако зомби по крайней мере позволяют живо проиллюстрировать ряд его важных моментов.

Даже при аргументе от представимости против логической супервентности, строго говоря, необязательно допускать логическую возможность зомби или зомбийного мира. Достаточно допустить логическую возможность мира, физически идентичного нашему, в котором факты, касающиеся сознательного опыта, всего лишь *отличны* от соответствующих фактов в нашем мире, хотя сам сознательный опыт может и присутствовать. При отсутствии какого-то позитивного факта об опыте в нашем мире в физически идентичном мире сознание не является логически супервентным.

Достаточно поэтому отметить, что мы можем когерентно вообразить физически идентичный мир с *инвертированным* сознательным опытом или (на локальном уровне) вообразить физически идентичное мне существование с инвертированными сознательными переживаниями. Можно было бы, к примеру, вообразить, что там, где у меня есть опыт красного, у моего инвертированного двойника имеется опыт синего, и наоборот. Конечно, он будет называть свои переживания синего «красным», но это неважно. Важно то, что вещи, которые мы оба называем «красными», — кровь, пожарные машины и т. п. — будут переживаться им так же, как переживаются мной вещи, которые мы оба называем «синими», вроде моря или неба.

Остальной его колористический опыт систематическим образом инвертирован в сравнении с моим для гармонизации с инверсией красного и синего. Если говорить о человеческом колористическом опыте, то, возможно, лучше всего представить это в виде перестановки двух осей нашего трехмерного колористического пространства, при которой красно-зеленая ось оказывается на месте желто-синей, и наоборот⁴. Для получения подобной инверсии в реальном мире, вероятно, потребовалось бы надлежащим образом перестроить нейронные процессы, но если говорить о логической возможности, то кажется совершенно когерентным допущение, что указанные переживания могли бы инвертироваться и при точном воспроизведении физической структуры. Ничто в нейрофизиологии не диктует, что один тип процессов должен сопровождаться скорее переживанием чего-то красного, чем переживанием желтого.

Аргумент 2: Инвертированный спектр

На это иногда возражают (Harrison 1973; Hardin 1987), говоря, что колористическое пространство человека асимметрично таким образом, что оно препятствует подобному инвертированию. Так, с некоторыми цветами ассоциирована теплота или холодность, которые кажутся напрямую связанными с различными функциональными ролями (к примеру, теплота воспринимается как нечто «позитивное», а холодность как «негативное»). Если поменять местами теплый и холодный цвет, то феноменальное чувство «тепла» оказалось бы рассогласованным с «теплой» функциональной ролью — «холодный» опыт зеленого характеризовался бы как скорее позитивный, чем негативный, и т. п. Подобным образом существует, как кажется, большее число различных оттенков красного, чем желтого, и поэтому при прямой подстановке переживаний желтого на место переживаний красного и наоборот могла бы возникать странная ситуация, при которой субъект смог бы функционально различать больше оттенков желтого, чем было бы различимо с феноменальной точки зрения. Не исключено, что количество асимметрий колористического пространства таково, что любое подобное инвертирование приводило бы к странному рассогласованию феноменального чувства и «надлежащей» функциональной роли.

В ответ на эти соображения можно сказать три вещи. Во-первых, с представлением о подобном рассогласовании (к примеру, холодной феноменологии и теплых реакций), как кажется, не связано ничего некогерентного, хотя можно признать, что оно выглядит как-то странновато⁵. Во-вторых, вместо точного переноса красного на синее и наоборот можно вообразить, что они переносятся на несколько иные цвета. К примеру, красное могло бы переноситься на «теплую» разновидность синего (как предлагает Левин (Levin 1991)) или даже на цвет не из нашего колористического пространства. В случае красного и желтого можно было бы представить, что красное переносится на расширенный ряд переживаний желтого, где можно проводить большее количество различий. Нет оснований утверждать, что сценарии с инвертированным спектром должны использовать цвета из обычного колористического пространства. В-третьих, и это, возможно, самый убедительный ответ, можно попытаться доказать (вместе с Шумейкером (Shoemaker 1982)), что даже если наше колористическое пространство и асимметрично, какие-то существа, несомненно, могли бы иметь симметричное колористическое пространство. Так, скорее всего, естественно возможны существа, видящие (и переживающие) всего лишь два цвета, *A* и *B*, соответствующие различным и четко ограниченным друг от друга длинам световых волн. Различие этих двух цветов исчерпывает у них структуру их колористического пространства. Кажется абсолютно когерентным вообразить два таких существа, физически

идентичных, но с инвертированным опытом *A* и *B*. И этого будет достаточно для получения того вывода, который мы хотим получить.

Даже многие редуктивные материалисты (к примеру, Шумейкер (Shoemaker 1982)) признавали когерентность инвертирования колористических переживаний при сохранении функциональной организации. Предполагается, что система с иными базовыми нейрофизиологическими свойствами или, к примеру, с кремнием вместо нейробиологии могла бы иметь другие колористические переживания. Но если признать это, то отсюда будет автоматически следовать, что инвертирование переживаний в физической копии является по крайней мере концептуально когерентным. Дополнительные нейрофизиологические свойства, не реализованные в подобном случае, опять-таки не есть нечто такое, что могло бы логически определять природу опыта. Даже если здесь и можно говорить о некоем апостериорном отождествлении определенных нейрофизиологических структур и определенных переживаний (как считает Шумейкер), мы все равно должны допускать представимость иной схемы связей в том смысле представимости, который релевантен для редуктивного объяснения.

Хотя отсутствие логической супервентности сознания устанавливается как возможностью инвертированных спектров, так и возможностью зомби, первая позволяет делать гораздо более слабый вывод, чем вторая. Вполне представима ситуация, когда мы признаем логическую возможность инвертированных спектров, но отрицаем логическую возможность зомби. Если бы дело действительно обстояло таким образом, то *существование* сознания могло бы получать редуктивное объяснение, а вот специфический *характер* конкретных сознательных переживаний — нет.

Как мы видели ранее, в сознании как одной из черт универсума есть нечто неожиданное. Мы верим в сознание исключительно на основании нашего собственного опыта сознания.

Даже если бы физическая сторона универсума была известна нам во всех деталях, если бы мы знали конфигурацию, причинность и изменение во времени всех полей и частиц пространственно-временного многообразия, эта информация не привела бы нас к допущению существования сознательного опыта. Мое знание о сознании первым делом идет от меня самого, а не от какого-то внешнего наблюдения. Проблема навязана мне моим опытом сознания от первого лица.

Из совокупности низкоуровневых фактов о физических конфигурациях и причинности мы в принципе можем вывести самые разные высокоуровневые факты о макроскопических системах, их организации и каузальных связях между ними. Действуя подобным образом,

Аргумент 3:
От эпистемической асимметрии

можно было бы установить все факты о биологической функции, а также о человеческом поведении и механизмах в мозге, которые его порождают. Но ничто в этой громадной каузальной истории не привело бы того, кто не переживал сознания напрямую, к убеждению, что здесь должно быть какое-то *сознание*. Сама эта идея была бы иррациональной, если не мистической.

Нельзя отрицать, что физические факты о мире могли бы служить неким косвенным свидетельством в пользу существования сознания. К примеру, на основании этих фактов можно было бы убедиться в существовании множества организмов, заявляющих о своей сознательности и говорящих о наличии у них таинственных субъективных переживаний. Однако подобные свидетельства не были бы убедительными, и в данной ситуации самым естественным могло бы оказаться элиминативистское заключение — что у этих существ на деле нет *опыта*, а есть лишь множество слов о нем.

Элиминативизм в вопросе о сознательном опыте оказывается неосновательной позицией *исключительно* потому, что мы сами обладаем подобным опытом. При отсутствии подобного непосредственного знания сознание могло бы разделить судьбу жизненного духа. Иначе говоря, наше знание о сознании характеризуется *эпистемической асимметрией*, отсутствующей в нашем знании о других феноменах⁶. Наше знание о существовании сознательного опыта идет главным образом от нас самих, а внешние свидетельства играют в лучшем случае второстепенную роль.

Суть дела можно также пояснить, указав на наличие проблемы других сознаний. Даже если мы знаем все физические аспекты других существ, мы не можем наверняка знать, что у них есть сознание или каковы их переживания (хотя у нас могут быть весомые основания верить в их существование). Примечательно то, что не существует проблемы «других жизней», «других экономик» или «других высот». Эпистемическая асимметрия отсутствует в указанных случаях, и она отсутствует именно потому, что подобные феномены логически супервентны на физическом.

Эпистемическая асимметрия в знании о сознании делает очевидной невозможность логической супервентности сознания. Если бы оно было логически супервентным, такой асимметрии не существовало бы; логически супервентное свойство может непосредственно обнаруживаться на основании внешних свидетельств, и позиция первого лица не играет здесь никакой особой роли. Конечно, некоторые супервентные свойства, такие, возможно, как память, легче обнаруживаются с позиции первого лица. Но это лишь вопрос о количестве затраченных усилий. Наличие памяти в принципе можно столь же надежно

установить с позиции третьего лица, как и с позиции первого лица. Эпистемическая асимметрия, связанная с сознанием, гораздо более фундаментальна, и она говорит нам о том, что никакой набор фактов о комплексной каузальности в физических системах не может сформировать некий факт о сознании.

Самый яркий аргумент против логической супервентности сознания предложен Джексоном (Jackson 1982) по мотивам сходных аргументов Нагеля (Nagel 1974) и других авторов. Вообразим, что мы живем в эпоху совершенной нейронауки и знаем всё, что можно знать о физических процессах в нашем мозге, ответственных за продуцирование нашего поведения. Мэри выросла в черно-белой комнате и не видела никаких цветов помимо черного, белого и оттенков серого⁷. При этом она — один из ведущих мировых нейроученых, специализирующихся по нейрофизиологии цветного зрения. Она знает всё, что можно знать о нейронных процессах, участвующих в обработке зрительной информации, о физике оптических процессов и о физическом устройстве объектов окружающей среды. Но она не знает, каково это — видеть что-то красное. Никакие рассуждения на основе одних лишь физических фактов не дадут ей этого знания.

Из этого следует, что факты о субъективном опыте цветного зрения невыводимы из физических фактов. В противном случае Мэри в принципе могла бы узнать, каково это — видеть красное, исходя из ее знания о физических фактах. Но она не может узнать этого. Не исключено, что Мэри могла бы узнать, каково это — видеть что-то красное, косвенным методом, к примеру при помощи надлежащих манипуляций со своим мозгом. Суть, однако, в том, что это знание не вытекает из одного лишь физического знания. Знание всей полноты физических фактов в принципе позволяет Мэри вывести все факты о реакциях системы, ее умениях и когнитивных способностях; но она останется в полном неведении о ее опыте красного.

Данное обстоятельство можно продемонстрировать и рассмотрев значительно отличающиеся от нас — возможно, гораздо более простые — системы, к примеру мышей или летучих мышей, отмечая, что физические факты, касающиеся этих систем, не говорят нам о том, каковы их сознательные переживания, если они вообще у них имеются (Нагель делает акцент именно на такого рода проблемах). Даже если в нашем распоряжении есть все физические факты о какой-то мыши, природа ее сознательного опыта остается *открытой*: с физическими фактами о ней совместимо присутствие у нее сознательного опыта, но с этими же фактами совместимо и его отсутствие. На основании физических фактов о летучей мыши мы можем удостовериться во *всех*

Аргумент 4:

Аргумент знания

фактах, касающихся летучей мыши, за исключением фактов о ее сознательном опыте. Зная все эти физические факты, мы по-прежнему не знаем, каково это — быть летучей мышью.

В подобном ключе мы можем поразмышлять и о компьютере, спроектированном как простой когнитивный агент (его интеллект может быть таким же, как у собаки), но сходном с нами в определенных отношениях, таких как способность к перцептивному различению. В частности, он очень похож на нас категоризирует цветные стимулы, группируя вещи, которые мы назвали бы «красными», под одной рубрикой, а вещи, которые мы назвали бы «зелеными», — под другой. Даже если нам известны все детали электрических схем этого компьютера, остаются следующие вопросы: (1) переживает ли компьютер хоть что-то, глядя на розы; (2) если да, то переживает ли он то самое чувственное качество цвета, которое имеется у нас, когда мы смотрим на розу, или же какое-то совершенно другое качество? Это вполне осмысленные вопросы, и полное знание о физических фактах не диктует нам какого-то определенного ответа. Значит, физические факты логически не влекут факты о сознательном опыте.

Джексон выдвигал свой аргумент как аргумент, направленный скорее против материализма, чем против редуктивного объяснения. Этот аргумент вызвал множество откликов; я буду обсуждать их в следующей главе, где в центре нашего рассмотрения будет не редуктивное объяснение, а материализм. Пока же интересно отметить, что в большинстве возражений против этого аргумента признавалось то обстоятельство, которое значимо для аргумента против редуктивного объяснения, а именно то, что знание о качественности красного является фактическим знанием, априорно не выводимым из знания о физических фактах. Единственный способ избежать данного вывода — отрицать, что знание о том, каково это — иметь опыт красного, вообще является собой знание о *факте*. Этой стратегии следовали Льюис (Lewis 1990) и Немиров (Nemirow 1990), доказывавшие, что у Мэри отсутствует лишь некое *умение*, такое как умение распознавать красные вещи. Я обсуждаю эту идею в следующей главе; здесь же я просто отмечу, что в той мере, в какой кажется очевидным, что, когда Мэри впервые видит красное, она *открывает* для себя нечто о мире, кажется очевидным и то, что получаемое ею знание есть знание о факте.

Аргумент 5: От отсутствия анализа

Если сторонники редуктивного объяснения хотят сохранить надежду на опровержение приведенных выше аргументов, они должны будут дать нам хоть какое-то представление о том, каким образом сознание *могло бы* оказываться выводимым из физических фактов. Несправедливым было бы требовать от них полной детализации, но мы

вправе рассчитывать хотя бы на объяснение того, как *могли бы* проходить такие выводы. Однако любая попытка показать это обречена на провал. Для выводимости сознания из набора физических фактов нам потребовался бы определенный анализ понятия сознания — анализ такого рода, реализация которого могла бы быть фундирована физическими фактами, но его не найти.

Единственным видом анализа сознания, который имеет хотя бы отдаленное отношение к достижению указанных целей, является функциональный анализ. В соответствии с таким анализом оказывалось бы, что понятие об осознанности чего-то сводится к представлению, что это что-то должно играть определенную функциональную роль. Так, можно было бы сказать, что для осознанности какого-то состояния требуется лишь возможность вербально отчитаться о нем или то, чтобы оно было результатом определенных перцептивных различий или определенным образом делало доступным информацию для последующих процессов и т. п. Похоже, однако, что эти рассуждения никуда не годятся в качестве анализов. Они попросту упускают то, что имеется в виду под сознательным опытом. Хотя осознанные состояния могут играть различные каузальные роли, они *не определяются* своими каузальными ролями. Осознанными их делает скорее присутствие в них феноменального чувствования, и это чувствование не является чем-то таким, от чего можно функционально отделаться.

Чтобы убедиться в неудовлетворительности этих анализов, заметим, до какой степени они тривиализируют проблему объяснения сознания. Неожиданно оказывается, что для объяснения сознания нам нужно лишь объяснить наше умение давать определенные вербальные отчеты, осуществлять некие различия или проявлять иную способность. Кажется, однако, что вполне можно представить, что мы могли бы объяснить всё это, вообще не объяснив при этом самого сознания, то есть не объяснив тот *опыт*, который сопровождает отчет или различие. Анализ сознания в терминах какого-либо функционального понятия предполагает либо подмену предмета обсуждения, либо уход от проблемы. С тем же успехом можно было бы определить «мир во всем мире» как «бутерброд с ветчиной». Обрести такой мир было бы гораздо проще, но это было бы мнимым достижением.

Функциональные анализы сознания могут быть оспорены и по более специальным соображениям. Так, любое понятие, допускающее функциональный анализ, будет в известной степени семантически неопределенным. Есть ли у мыши убеждения? Обучаются ли бактерии? Является ли компьютерный вирус живым? Лучшим ответом на такие вопросы обычно оказывается: в каком-то смысле да, в каком-то — нет. Все зависит от того, как мы ограничиваем понятия, и в любых

высокоуровневых функциональных понятиях подобные границы будут размытыми. Но сравним: есть ли у мыши сознательный опыт? Есть ли он у вируса? Это не такого рода вопросы, которые можно решить на уровне соглашений. Либо есть нечто, выражающее то, каково это — быть мышью, либо нет, и мы не в силах дефинитивно создавать или уничтожать опыт мыши. Конечно, сознательный опыт, скорее всего, континуален — от крайне тусклого до очень богатого, но если нечто обладает сознательным опытом, каким бы тусклым он ни был, мы не можем устраниТЬ его путем соглашений. Эта определенность не может быть получена из функционального анализа понятий, сопряженных с сознанием, так как все подобные функциональные понятия несколько размыты. И если это так, то отсюда следует, что понятие сознания не может быть раскрыто путем функционального анализа.

Другое возражение состоит в том, что функционалистский анализ приводит к коллапсированию важного различия понятий осведомленности и сознания, о котором шла речь в главе 1. Если сознание можно функционально проанализировать, то этот анализ, судя по всему, не будет существенно отличаться от проведенного там анализа осведомленности в терминах определенного рода доступности информации для последующих процессов и для контроля за поведением. Осведомленность — замечательное понятие, но оно совершенно отлично от понятия сознательного опыта. Функционалистская трактовка сжимает два понятия — сознания и осведомленности — в одно и поэтому искажает нашу концептуальную систему.

Альтернативы функционального анализа выглядят еще хуже. Далеко не ясно, можно ли вообще найти какой-то иной способ анализа, который мог бы оказаться пригодным для редуктивного объяснения. Единственной альтернативой мог бы быть структурный анализ — быть может, сознание можно было бы проанализировать в качестве некоей биохимической структуры, — но неадекватность такого анализа была бы еще более очевидной. Вне зависимости от того, является ли сознание биохимической структурой, под «сознанием» понимается нечто иное. Подобный анализ сознания опять-таки тривиализирует объясняющую проблему путем подмены предмета обсуждения. Понятие сознания, похоже, нередуцируемо и характеризуемо только в терминах понятий, предполагающих сознание.

Заметим, что нередуцируемость, о которой идет речь, совершенно отлична от той, которая, как иногда утверждается, имеет место в высокоуровневых понятиях в целом. Мы видели, что многие высокоуровневые понятия лишены четких definиций и их нельзя реально проанализировать в терминах необходимых и достаточных условий. Тем не менее, как мы видели в предыдущей главе, такие понятия

допускают хотя бы грубый анализ, дающий приблизительно верные результаты, хотя в силу этого неизбежно и не позволяющий нарисовать детальной картины. Важнее всего, однако, то, что мы без особого труда можем понять, что такие свойства, как жизнь, обучение и т. п., могут быть проанализированы в качестве функциональных свойств, даже если артикуляция того, *какими именно* будут эти свойства, оказывается непростым делом. Даже если у этих свойств отсутствуют четкие функциональные дефиниции, они все же являются вполне совместимыми с выводимостью из физических фактов.

С сознанием связаны проблемы совсем другого уровня. Предлагаемые анализы не дают даже приблизительно верных результатов. Все гораздо хуже: они совершенно не в состоянии охарактеризовать то, что нуждается в объяснении. Здесь не возникает искушения хотя бы *попробовать* добавить какие-то эпизоды к предполагаемому функциональному анализу сознания, чтобы сделать его приемлемым, в отличие от аналогичных анализов жизни и обучения. И это связано прежде всего с тем, что сознание не может быть охарактеризовано как функциональное свойство. Это же справедливо и относительно анализа сознания как структурного свойства или в других редуктивных терминах. Таким образом, невозможно даже и подступиться к выводимости сознания из физических фактов.

2. Крах редуктивного объяснения

Отсутствие логической супервентности сознания на физическом означает невозможность редуктивного объяснения сознания. При любой характеристике физических процессов, предположительно лежащих в основе сознания, всегда будет возникать дополнительный вопрос: почему эти процессы сопровождаются сознательным опытом? В случае большинства других феноменов на подобный вопрос несложно ответить: физические факты о таких процессах *влекут* существование указанных феноменов. К примеру, если мы говорим о таком феномене, как жизнь, то физические факты имплицируют осуществление определенных функций, а осуществление этих функций и есть всё, что нам нужно объяснить для объяснения жизни. Но когда речь идет о сознании, такого рода ответ не проходит.

Физическое объяснение хорошо подходит для объяснения *структуры и функции*. Структурные и функциональные свойства могут быть напрямую выводимы из низкоуровневого физического описания и поэтому с очевидностью пригодны для редуктивного объяснения. И почти все высокоуровневые феномены, нуждающиеся в объяснении, в конечном счете сводятся к структуре и функции: подумаем, к примеру, об объяснении водопадов, планет, переваривания пищи, размножения,

языка. Но объяснение сознания не сводится к объяснению структуры и функции. После объяснения всей физической структуры, связанной с мозгом, и объяснения механизмов осуществления различных функций мозга нам еще надо объяснить само сознание. Почему вся эта структура и функция должны порождать опыт? Рассказ о физических процессах не отвечает на этот вопрос.

Мы можем представить это в терминах мысленных экспериментов, о которых мы говорили ранее. Любое рассуждение о физических процессах в равной мере применимо ко мне и к моему зомбийному двойнику. Отсюда следует, что ничто в этом рассуждении не объясняет, почему в моем случае возникает сознание. Аналогичным образом любое рассуждение о физических процессах в равной мере применимо к моему инвертированному двойнику, видящему синее там, где я вижу красное, из чего следует, что ничто в этом рассуждении не объясняет, почему мои переживания являются именно такими, а не какими-то иными. Сам факт логической возможности одинаковости физических фактов при различии фактов о сознании демонстрирует нам то, что Левин (Levine 1983) называет *объяснительным провалом* между физическим уровнем и сознательным опытом.

Если это так, то факт сопровождения сознанием того или иного физического процесса — *дополнительный факт*, не объяснимый одними лишь рассуждениями о физических фактах. В каком-то смысле это сопровождение надо принимать просто как данность. Мы могли бы попытаться систематизировать и объяснить эти грубые факты в терминах какой-то простой и более фундаментальной схемы, но при этом всегда будет оставаться элемент, логически независимый от этой физической истории. Возможно, мы могли бы получить что-то вроде объяснения, сочетаая базовые физические факты с дополнительными *связующими* принципами, сопрягающими эти факты с сознанием, но такое объяснение не будет редуктивным. Сама потребность в развернутых связующих принципах показывает, что сознание объясняется не редуктивно, а в его собственных терминах.

Разумеется, сказанное мной не означает, что физические факты *не имеют отношения* к объяснению сознания. Мы всё же можем ожидать, что физические объяснения играют весомую роль в теории сознания, предоставляя, к примеру, информацию о физической *основе* сознания и, возможно, позволяя в деталях представить соответствие между различными аспектами физических процессов и сознательного опыта. Подобные объяснения могут быть особенно полезными для понимания *структурь* сознания: схем сходства и различия переживаний, геометрической структуры феноменальных полей и т. п. Подробнее мы поговорим об этих и других вещах, которые можно узнать об опы-

те из физических объяснений в нередуктивном контексте, в главе 6. Но одного физического объяснения мало.

На этой стадии естественно может возникнуть ряд возражений.

Можно было бы сказать, что объяснение любых высокоуровневых феноменов будет сопровождаться допущением «связующих законов», дополняющих низкоуровневое описание, и что только с помощью этих связующих законов можно конкретизировать детали высокоуровневых феноменов. Однако, как можно заключить из обсуждения в предыдущей главе (и как тщательно доказывает Хорган (Horgan 1978)), в подобных случаях связующие законы не являются дополнительными фактами о мире. Скорее, единительные принципы сами оказываются логически супервентными на низкоуровневых фактах. Предельным случаем такого связующего принципа является супервентностный кондиционал, который, как мы видели, обычно может быть представлен в качестве концептуальной истины. Другие, более «локализованные» связующие принципы, такие как звено, сопрягающее молекулярное движение и тепло, могут по крайней мере выводиться из физических фактов. В случае же сознания подобные связующие принципы должны приниматься как нечто изначальное.

**Возражение 1:
Не слишком ли высока
наша планка?**

Интересно посмотреть, каким образом типичное высокоуровневое свойство, такое, к примеру, как жизнь, ускользает от аргументов, обсуждавшихся нами, когда речь шла о сознании. Во-первых, непосредственно не представима физическая копия живого существа, которая сама не была бы живой. Здесь могла бы возникать проблема, связанная со свойствами, зависимыми от контекста (был бы живым или был бы человеком двойник, случайно возникший в каком-то болоте?), но фиксация фактов об окружении элиминирует даже эту возможность. Во-вторых, здесь невозможна ситуация с «инвертированной жизнью» по аналогии с инвертированным спектром. В-третьих, если мы знаем все физические факты об организме (и, возможно, о его окружении), в нашем распоряжении имеется достаточно материала для знания обо всех биологических фактах. В-четвертых, в отношении жизни не существует эпистемической асимметрии; факты, касающиеся жизни других существ, в принципе не менее доступны, чем факты, касающиеся нашей собственной жизни. В-пятых, понятие жизни вполне можно проанализировать в функциональных терминах: быть живым — значит, грубо говоря, обладать определенными способностями к адаптации, размножению и метаболизму. В общем, большинство высокоуровневых феноменов сводится к физической структуре и функции, и у нас есть серьезное основание полагать, что структурные и функциональные свойства логически супервентны на физическом.

Возражение 2:

Не может ли виталист сказать то же самое о жизни?

Несмотря на все вышесказанное, типичной реакцией на аргументацию, подобную той, которая была приведена мной, является замечание, что виталист мог бы сказать то же самое относительно жизни⁸. К примеру,

виталист мог бы заявить, что логически возможным является то, что моя физическая копия могла бы и не быть *живой*, чтобы провести тезис о невозможности редуктивного объяснения жизни. И виталист мог бы доказывать, что жизнь — это дополнительный факт, необъяснимый каким-либо описанием физических фактов. Но виталист был бы *неправ*. Аналогично, не может ли ошибаться и противник редуктивного объяснения сознания?

Думаю, что эта реакция неверно диагностирует истоки виталистских возражений. Стимулом к витализму были по большей части сомнения относительно возможности физических механизмов реализовать все те сложные функции, которые связаны с жизнью: адаптивное поведение, размножение и т. п. В то время знания об исключительной хитроумности биохимических механизмов были крайне скучны, так что такого рода сомнения были вполне естественными. Но в самих этих сомнениях неявно присутствовало концептуальное допущение того, что для объяснения жизни нужно объяснить реализацию различных функций. И примечательно, что в ходе постепенного появления физических объяснений соответствующих функций виталистские сомнения по большей части исчезли. Однако в случае сознания проблема остается даже при объяснении функций.

Если представить вменяемому виталисту полное физическое объяснение, показывающее, как физические процессы реализуют релевантные функции, то он согласится с тем, что жизнь получила свое объяснение. Не существует даже *концептуальной* возможности реализации этих функций при отсутствии жизни. Не исключено, что крайний виталист будет отрицать даже это, заявляя, что что-то все же остается за бортом функционального объяснения жизни — к примеру, жизненный дух. Но очевидное и убедительное возражение на это состоит в том, что, в отличие от опыта, жизненный дух не является тем, в существовании чего мы можем быть убеждены по независимым основаниям. В той мере, в какой вообще имелось основание быть убежденным в его существовании, оно было связано с его трактовкой в качестве объяснительного конструкта — «Мы должны признавать нечто подобное, чтобы делать такие-то удивительные вещи». Но в качестве объяснительного конструкта жизненный дух может быть элиминирован, если отыскать лучшее объяснение того, как реализуются интересующие нас функции. Сознательный же опыт навязывается нам в качестве того, что нуждается в объяснении, и от него так просто не избавиться.

Одна из причин, по которой виталист мог бы счесть, что при функциональном объяснении жизни что-то все же остается за бортом, состоит именно в том, что физические объяснения не объясняют, почему в соответствующих явлениях есть что-то, выражающее то, каково это — быть живым. Быть может, вера в жизненный дух была отчасти сопряжена с феноменами нашей внутренней жизни. Многие чувствовали связь между понятиями жизни и опыта, и даже сегодня мы вправе сказать, что при объяснении жизни надо объяснить в том числе и то, что многие живые существа обладают сознанием. Но существование *таких* виталистских сомнений может вызвать лишь дискомфорт у защитника идеи редуктивного объяснения сознания, так как подобные сомнения остаются неразрешенными.

Философы нередко с подозрением относятся к аргументам, отводящим ключевую роль представимости, замечая, что представимости недостаточно для возможности. Это тонкий вопрос, который я обсуждал ранее и к которому еще вернусь. Но в данном случае эти тонкости не имеют такого уж большого значения. Когда речь идет об *объяснении*, представимость, очевидно, играет ключевую роль. Если в итоге размышлений мы увидим, что можем представить, что все эти физические процессы могли бы протекать при отсутствии сознания, то нельзя будет найти удовлетворительного редуктивного объяснения сознания: все время будет возникать дополнительный вопрос, почему существуем *мы*, а не зомби. Даже если представимость ограничена пределами человеческих способностей, аналогичным образом ограничены ими и даваемые на-ми объяснения.

Эту мысль можно выразить и по-другому, отметив то обстоятельство, что редуктивное объяснение феномена в физических терминах требует априорного имплицирования физическими фактами соответствующих высокоразвитых факторов (логической супервентности в соответствии с первичным интенсионалом, как я обозначал его раньше). Если подобная связь отсутствует, мы всегда сможем задать дополнительный вопрос о том, почему физические процессы порождают сознание. Мы видели, что такая связь существует почти во всех областях, и это делает возможным в них редуктивное объяснение; но она, похоже, отсутствует в случае сознательного опыта. Можно высказывать сомнение относительно того, предполагаются ли подобные априорные связи *онтологическими* взглядами, такими как материализм, — я рассмотрю этот вопрос в следующей главе, — но когда речь идет о редуктивном объяснении, они играют ключевую роль.

Возражение 3: Вытекает ли из представимости возможность?

Возражение 4:
А не набор ли это
замкнутых в круг
интуиций?

Можно также возразить на все это, сказав, что выдвинутые мной аргументы — это, по сути, не более чем набор каких-то интуиций. Разумеется, в определенном смысле все эти аргументы основаны на интуиции, но я пытался показать, насколько естественны и очевидны эти интуиции и насколько натужным может быть их отрицание. Главная из них — *что тут есть то, что нуждается в объяснении*, — некий феномен, связанный с опытом от первого лица и составляющий проблему, не возникающую при наблюдении познания с позиции третьего лица. Если признать, что опыт от первого лица навязывает нам нечто нуждающееся в объяснении, не навязываемое наблюдением с позиции третьего лица, то это сразу же развернет перед нами большинство вышеприведенных аргументов. Из этого, к примеру, непосредственно следует, что то, что нуждается в объяснении, не может быть проанализировано в качестве исполнения какой-то функциональной роли, так как последнее известно нам из наблюдения с позиции третьего лица и представляет собой гораздо более простой феномен.

Задействованная здесь «интуиция» — самый что ни на есть *raison d'être* проблемы сознания. Единственный осмыслиенный способ обойти эту интуицию — полностью отрицать данную проблему и сам феномен. Всегда — или по крайней мере всегда в «философских» рассуждениях — можно полностью отрицать данные интуиции и отрицать наличие того, что нуждается в объяснении (за исключением реализации разного рода функций). Но если мы всерьез относимся к сознанию, из сказанного должны вытекать те выводы, о которых идет речь и которые я пытаюсь обосновать.

Возражение 5:
Не должны ли все
объяснения где-то
заканчиваться?

Последнее возражение состоит в том, что никакие объяснения не берутся из пустоты: они должны где-то заканчиваться. Так, при объяснении движения планет мы принимаем за аксиому законы гравитации и существование массы. Быть может, и в нашем случае мы должны просто принять что-то за аксиому? Мне близок такой подход; я считаю, что мы должны принимать что-то за аксиому при объяснении сознания. Но действуя подобным образом, мы неизбежно выходим за пределы *редуктивного* объяснения. В самом деле, подобная аналогия подкрепляет нередуктивную позицию, которую я отстаиваю. Мы принимаем за аксиому законы физики вследствие их *фундаментальности*. Если мы принимаем за аксиому звено, связующее физические процессы и сознательный опыт, то это находит на мысль, что это связующее звено тоже должно считаться чем-то фундаментальным. В следующей главе я вернусь к этому вопросу.

3. Когнитивное моделирование

В этом и последующих параграфах я проиллюстрирую крах редуктивного объяснения критическим рассмотрением ряда концепций сознания, предложенных исследователями, представляющими самые разные дисциплины. Не все эти концепции предлагались в качестве редуктивных объяснений сознательного опыта, хотя зачастую они и трактовались подобным образом; в любом случае, однако, полезно посмотреть, что можно и чего нельзя достичь с помощью этих концепций. По ходу дела небезынтересным будет отмечать различное отношение этих исследователей к трудным вопросам, возникающим в связи с сознательным опытом.

Вначале я рассмотрю концепции, основанные на *когнитивном моделировании*. Когнитивное моделирование хорошо подходит для решения большинства проблем когнитивной науки. Создавая модель каузальной динамики когнитивных процессов, можно объяснить производование поведения когнитивным агентом. Это позволяет хорошо объяснять такие психологические феномены, как обучение, память, восприятие, контроль за действием, внимание, категоризация, лингвистическое поведение и т. п. Если в нашем распоряжении находится модель, схватывающая каузальную динамику того, кто, к примеру, находится в процессе обучения, то это означает, что все, в чем будет реализована такая динамика в надлежащем окружении, будет находиться в процессе обучения. На основании этой модели мы можем понять то, как осуществляются определенные функции, и это все, что нам нужно объяснить для объяснения обучения. Но этого недостаточно для объяснения сознания. В связи с любой показанной нами моделью можно задать дополнительный вопрос о том, почему реализация этой модели должна сопровождаться сознанием. И на этот вопрос нельзя ответить с помощью одного лишь описания или анализа подобной модели.

Иногда высказывается выражение, что предполагаемые модели сознания не могут быть протестированы, так как невозможно верифицировать, будут ли сознательными те вещи, которые реализуют эти модели. Это действительно проблема, но тут имеется и более глубокая проблема. Даже если бы у нас (*per impossible*) был «опытометр», с помощью которого можно было бы заглядывать внутрь подобных объектов и говорить, являются ли они сознательными, это позволяло бы лишь устанавливать корреляцию. Мы знали бы, что при реализации этой модели всегда обнаруживается сознание. Но она не объясняла бы сознание в том смысле, в каком подобные модели объясняют другие ментальные феномены.

Такие модели, разумеется, могут объяснять «сознание» в психологических смыслах этого термина, если конструировать его в качестве некоей когнитивной или функциональной способности. Многие

из существующих «моделей сознания» при самом благожелательном отношении могут быть истолкованы в этом свете. Мы можем рассматривать их в качестве объяснений способности давать отчеты, внимания, интроспективных способностей и т. д. Но ни одна из них даже не приближается к объяснению того, почему эти процессы должны сопровождаться сознательным опытом. Примеры, о которых сейчас пойдет речь, проиллюстрируют это.

Первый пример связан с когнитивной моделью, представленной Бернардом Баарсом (Baars 1988) в качестве одной из частей развернутой на целую книгу трактовки сознания с позиции когнитивной психологии. Баарс привлекает самые разные экспериментальные данные для обоснования своего главного тезиса о том, что сознание есть нечто вроде *глобального рабочего пространства* в рассредоточенной системе интеллектуальных информационных процессоров. Когда эти процессоры получают доступ к глобальному рабочему пространству, они передают сообщение всей системе, как если бы они написали его на классной доске. То, что наполняет глобальное рабочее пространство, и составляет содержание сознания.

Баарс использует эту модель для объяснения впечатляющего множества свойств протекающих у нас процессов. Эта модель создает очень перспективный фон для объяснения доступа субъекта к информации и его роли во внимании, способности дать отчет, произвольном контроле и даже в формировании представления о самом себе. Модель глобального рабочего пространства, таким образом, хорошо подходит для объяснения сознания во всей совокупности его психологических смыслов. Теперь у нас имеется хотя бы общая теория *осведомленности*.

Здесь, однако, мы не найдем редуктивного объяснения *опыта*. Вопрос о том, почему эти процессы должны порождать опыт, попросту не рассматривается. Можно было бы предположить, что, согласно этой теории, содержание опыта точно совпадает с тем, что наполняет глобальное рабочее пространство. Но даже если это так, ничто в самой этой теории не объясняет, почему информация внутри глобального рабочего пространства оказывается тем, что переживается в опыте. В лучшем случае эта теория может сказать, что данная информация переживается потому, что она является *глобально доступной*. Но тогда тот же вопрос воспроизводится в ином виде: почему глобальная доступность должна порождать сознательный опыт? Этот сопрягающий вопрос не рассматривается в работе Баарса.

Баарс мимоходом касается этой проблемы: «Сkeptически настроенный читатель может... задачиться тем, действительно ли мы описываем сознательный опыт или же мы можем иметь дело лишь с побочными феноменами, связанными с ним» (с. 27). Его ответ состоит в том, что научные теории обычно хотя бы *находят подходы* к «самой вещи».

К примеру, биология объясняет *саму* наследственность, а не всего лишь связанные с ней феномены. Но это, как мы видели, означает, что здесь попросту игнорируется родовое отличие сознания от таких феноменов. Когда речь идет о наследственности, нам нужно объяснить лишь функции. В случае же сознания имеется и дополнительное нечто, нуждающееся в объяснении, — сам опыт. Таким образом, теорию Баарса можно рассматривать в качестве интересного подхода к когнитивным процессам, лежащим в основе сознания, который косвенным образом улучшает наше понимание сознания, но при этом оставляет незатронутыми ключевые вопросы — почему существует сознание, и как оно возникает из когнитивных процессов?

Дэниел Деннет тоже предлагает когнитивную модель сознания. В действительности он создал по меньшей мере две таких модели. Первая из них (см. Dennett 1978c), «рамочно-стрелочная» модель, изображает поток

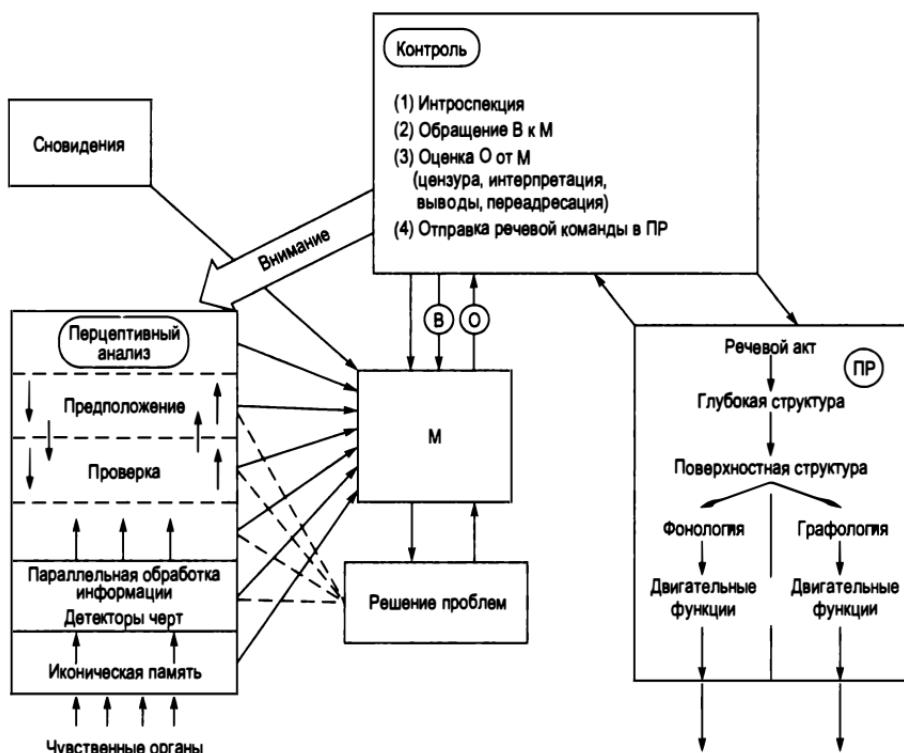


Рис. 3.2. Деннетовская когнитивная модель сознания. (Источник: Figure 9.1, p. 155 в Daniel C. Dennett, *Brainstorms: Philosophical Essays on Mind and Psychology*, The MIT Press. Copyright © 1987 by Bradford Books, Publishers. С разрешения The MIT Press)

информации между различными модулями (рис. 3.2). Ключевыми в этой модели являются: (1) перцептивный модуль, (2) хранилище кратковременной памяти M , получающее информацию от перцептивного модуля, (3) система контроля, взаимодействующая с хранилищем памяти посредством вопросов и ответов и могущая направлять внимание на данные перцептивного модуля, и (4) инстанция, осуществляющая «связи с общественностью», получающая указания осуществить речевые акты от системы контроля и конвертирующая их в высказывания публичного языка.

Что могла бы объяснить эта модель? Хотя она представлена в очень упрощенном виде (и Деннет, возможно, не стал бы спорить с этим), при ее конкретизации она могла бы объяснить *способность давать отчет*, то есть нашу способность отчитываться о содержании наших внутренних состояний. Она также содержит каркас для объяснения нашей способности использовать перцептивную информацию при контроле над поведением, интроспективно постигать наши внутренние состояния и т. д. Но она не объясняет, почему в системе, где протекают эти процессы, должно быть что-то, выражющее, каково это — быть системой с подобными процессами.

В «Объясненном сознании» Деннет (Dennett 1991) выдвигает более детализированную концепцию, опирающуюся на большой массив недавних исследований в области когнитивной науки. Предлагаемая здесь модель, по сути, является моделью «пандемония», в которой мы видим множество мелких агентов, борющихся за внимание к себе, причем главную роль в дирижировании позднейшими процессами играет тот, кто кричит громче всех. Согласно этой модели, существует не «штаб-квартира», из которой осуществляется контроль, а множество каналов одновременного влияния. Деннет дополняет эту концепцию, обращаясь к данным нейронауки, эволюционной биологии, к конекционистским моделям и порождающим системам, обсуждаемым в работах по искусственному интеллекту.

Несмотря на комплексность этой концепции, она направлена на рассмотрение главным образом тех же феноменов, что и более ранняя. В лучшем случае она могла бы объяснить способность давать отчет, а в более общем плане — влияние различных видов информации на контроль за поведением. Она также могла бы объяснить фокусировку внимания. Она провокативно трактует некоторые из наших когнитивных способностей, но, как и предшествующая ей модель, ничего не говорит о том, почему эти способности должны сопровождаться сознательным опытом.

В отличие от большинства авторов, предлагающих когнитивные модели, Деннет недвусмысленно заявляет, что его модели могли бы объяснить все, что нуждается в объяснении относительно опыта. В част-

ности, он полагает, что для объяснения сознания нужно лишь объяснить такие функциональные феномены, как способность давать отчет и контроль; любой феномен, который, похоже, остается за пределами таких объяснений, есть не более чем химера. Иногда кажется, будто он попросту допускает, что при объяснении различных функций объясняется вообще всё (см., напр., Dennett 1993а, с. 210), однако в других случаях он приводит аргументы. Некоторые из этих аргументов я рассмотрю впоследствии⁹.

Подобная критика может быть высказана и по отношению к когнитивно-моделирующим подходам к сознанию Черчленда (Churchland 1995), Джонсон-Лэрда (Johnson-Laird 1988), Шеллиса (Shallice 1972, 1988а, 1988б) и многих других. Все они предлагают интересные трактовки когнитивных функций, не касаясь при этом действительно трудных вопросов.

4. Нейробиологическое объяснение

В последнее время приобрели популярность нейробиологические подходы к сознанию. Подобно когнитивным моделям, они могут немало способствовать объяснению психологических феноменов, в частности различных видов осведомленности. Они могут также сообщать нам что-то о процессах в мозге, *скоррелированных* с сознанием. Но ни одна из таких концепций не объясняет данную корреляцию: мы не узнаем из них, почему процессы в мозге вообще должны порождать сознание. С точки зрения нейронауки подобная корреляция есть не более чем некий исходный факт.

В методологическом плане не очевидно, как вообще можно приступить к созданию нейронаучной теории. Как проводить эксперименты, устанавливающие корреляцию между некоторыми нейронными процессами и сознанием? Обычно теоретики неявно опираются на какой-то психологический критерий сознания — фокусировку внимания, контроль за поведением, но чаще всего — на способность вербально отчитаться о внутреннем состоянии. Затем отмечается, что при наличии этих критерий реализуется некое нейрофизиологическое свойство, и вот перед нами теория сознания.

Однако сам факт использования здесь таких косвенных критериев ясно показывает, что тут не дается редуктивного объяснения сознания. В лучшем случае нейрофизиологическая концепция могла бы объяснить механизмы реализации релевантного психологического свойства. Вопрос о том, почему данное психологическое свойство должно сопровождаться сознательным опытом, остается без ответа. Поскольку успех этих теорий зависит от *допущения* связи между психологическими

свойствами и сознательным опытом, ясно, что они не объясняют эту связь. В этом можно убедиться, рассмотрев ряд недавних нейронаучных концепций сознания.

В последнее время внимание в нейронауке было приковано к колебаниям частотой в 40 Гц в зрительной коре и в других участках мозга. Фрэнсис Крик и Кристофф Кох (Crick and Koch 1990) выдвинули гипотезу о том, что такого рода колебания могут оказаться фундаментальной нейронной характеристикой, ответственной за сознательный опыт, и отстаивали идею развития нейробиологической теории в подобном ключе¹⁰.

Но почему колебания частотой в 40 Гц? Главным образом потому, что, как можно заключить из имеющихся данных, эти колебания играют важную роль в *связывании* различных видов информации в единое целое. Два разных типа информации о какой-то сцене — к примеру, относительно формы и локализации объекта — могут представляться по отдельности, однако эта теория указывает, что обособленные нейронные представления могут совпадать по частоте и фазе их колебаний, что допускает объединение этой информации в позднейших процессах и ее сохранение в рабочей памяти. Подобным образом самая разная фрагментированная информация могла бы оказываться интегрированной в «содержания сознания».

Такая теория и в самом деле могла бы нейробиологически прояснить процесс связывания и рабочую память, а в конечном счете и то, каким образом информация интегрально используется при контроле за поведением. При этом, однако, остается без ответа ключевой вопрос: почему эти колебания должны сопровождаться сознательным опытом? Данная теория дает частичный ответ: потому что эти колебания ответственны за связывание. Но она не рассматривает вопрос, почему само это связывание должно сопровождаться опытом. Успех этой теории зависит от допущения сопряженности связывания и сознания, и поэтому она не объясняет ее.

Крик и Кох, похоже, сочувственно относятся к «большой» проблеме сознания, называя ее «главной загадкой, с которой приходится иметь дело при создании нейрональной концепции психического». Они доказывают, что чисто когнитивистские подходы обречены на провал и что нужны теории нейронального уровня. Но они не дают нам основания считать, что их теория превосходит когнитивные теории в плане ее способности отвечать на действительно трудные вопросы. На самом деле они и не утверждают, что их проект нацелен на решение проблемы опыта. В одном из интервью Кох достаточно ясно высказывается об ограничениях их подхода:

Для начала давайте забудем о действительно трудных аспектах, таких как субъективные чувства, так как у них может и не быть научного решения. Субъективные состояния игривости, боли, удовольствия, видения синевы или обоняния розы — разрыв между материалистическим уровнем объяснения молекул и нейронов и субъективным уровнем кажется здесь очень большим. Сосредоточимся на том, что легче изучать, к примеру на визуальной осознанности. Вы сейчас говорите со мной, но Вы не смотрите на меня, Вы смотрите на капучино и, значит, осознаете его. Вы можете сказать: «Это чашка, и в ней находится какая-то жидкость». Если я передам ее Вам, Вы протянете руку и возьмете ее — Вы осмысленно среагируете на мое движение. Именно это я и называю осознанностью¹¹.

Другая нейрофизиологическая теория сознания была очерчена Джералдом Эдельманом в «Воспомянутом настоящем» (Edelman 1989), а также в других его книгах и статьях. Ядром его теории является идея

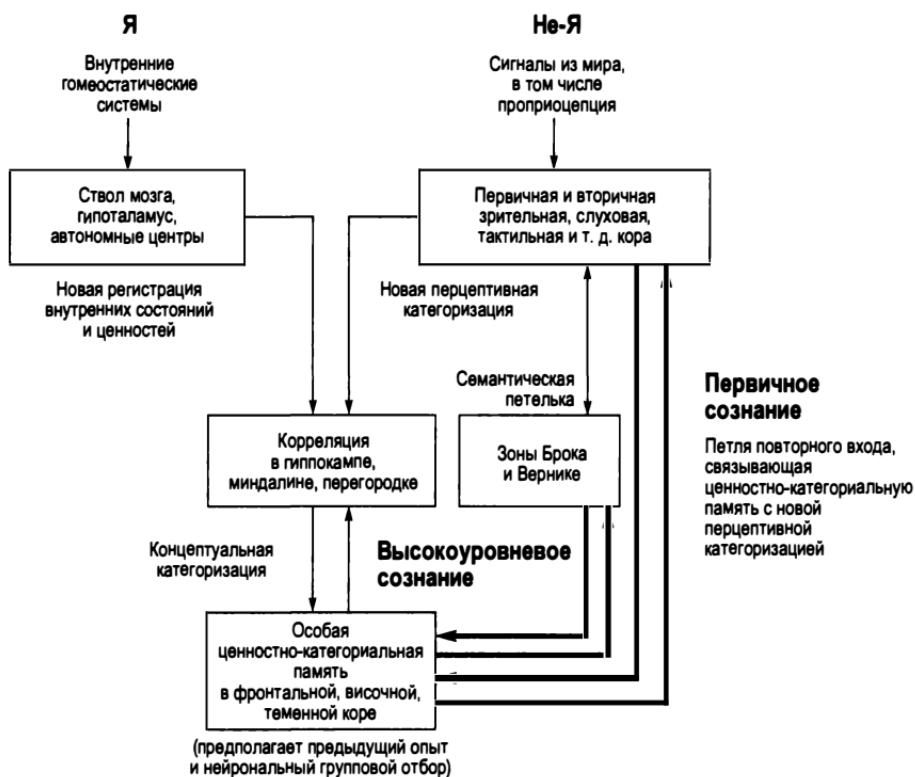


Рис. 3.3. Эдельмановская схема высокоуровневого сознания. (Источник: Figure 12-4, p. 132 в *Bright Air, Brilliant Fire* by Gerald M. Edelman. Copyright © 1992 by BasicBooks, Inc. С разрешения BasicBooks, подразделения HarperCollins, Inc.)

нейронных цепей повторного входа, благодаря которым перцептивные сигналы могут концептуально категоризоваться перед запоминанием. Перцептивная информация сложным образом взаимодействует с внутренним состоянием (как показано на рис. 3.3), порождая «первичное сознание». Его модель «высокоуровневого сознания» привносит новый элемент памяти, связанный с «семантической самонастройкой», что приводит к появлению понятий самости, прошлого и будущего. Все это увязывается с продуцированием языка, происходящим в зонах Брука и Вернике.

Значительная часть работ Эдельмана посвящена объяснению скорее восприятия, памяти и языка, нежели сознания. В тех их частях, где речь идет о сознании, обсуждения зачастую туманны, но тем не менее возникает ощущение, что подобная модель в конечном счете может объяснить перцептивную осведомленность — то есть влияние процессов восприятия на последующие процессы и на контроль за поведением, — а также некоторые аспекты самосознания, в частности происхождение понятия самости.

Эдельман не объясняет, как все эти процессы должны порождать сознательный опыт. Он попросту принимает факт корреляции между ними. Он не скрывает этого обстоятельства, отмечая, что феноменальный опыт составляет труднейшую проблему для теории сознания, а также то, что никакая физическая теория не может в полной мере объяснить нам квалиа:

Это указывает на то, как надо относиться к проблеме квалиа. В качестве основы теории сознания разумно допустить, что подобно тому, как квалиа существуют у нас самих, они существуют и у других сознающих людей, неважно, в качестве научных наблюдателей или субъектов... Затем мы можем признать людей лучшим каноническим объектом для исследований сознания. Это оправдывается тем, что субъективные отчеты людей (в том числе отчеты о квалиа), действия, а также структуры и функции мозга могут быть скоррелированы друг с другом. После построения теории, основанной на допущении существования квалиа у людей, мы можем заново взглянуть на некоторые из свойств квалиа, отталкиваясь от этих корреляций. Таким образом, наша способность давать отчет и устанавливать корреляции при индивидуальном переживании квалиа открывает возможность для научного исследования сознания (Edelman 1992, с. 115).

Опять-таки, поскольку эта теория основана на *допущении* корреляции, ясно, что она не может предложить редуктивного объяснения опыта. Эдельман обычно говорит лишь, что он объясняет процессы, лежащие в основе сознательного опыта, и не утверждает, что он объясняет сам опыт¹².

5. Апелляция к новой физике

Иногда утверждается, что ключ к объяснению сознания может быть найден в какой-то новой разновидности физической теории. Быть может, доказывая, что сознание не вытекает из физики нашего мира, мы неявно допускали, что эта физика подобна сегодняшней физике, в которой говорится о расположении частиц и полей в пространственно-временном многообразии, где протекают сложные каузальные и эволюционные процессы. Оппонент мог бы согласиться с тем, что существование сознания невыводимо из *такого* рода физики, но при этом доказывать, что сознание могло бы оказываться следствием новой разновидности физической теории.

При отсутствии конкретных идей оценить этот тезис непросто. Неплохо было бы увидеть хотя бы пример того, какой *могла бы быть* эта новая физика. Пример, о котором идет речь, не обязан выглядеть правдоподобным в свете современных теорий, но он должен был бы быть таким, чтобы можно было опознать, что речь в нем идет о физике. Ключевой вопрос: как из очевидно физической теории могло бы выводиться существование сознания? Если подобная теория состоит в описании структуры и динамики полей, волн, частиц и т. п., то в связи с ней будут возникать все обычные проблемы. И неочевидно, что *какая-либо* из физических теорий могла бы отличаться от нее до такой степени, чтобы избежать этих проблем.

Проблема в том, что основные элементы физических теорий, как кажется, всегда сводятся к двум вещам: структуре и динамике физических процессов. Разные теории говорят о различных структурах. Ньютонианская физика говорит о евклидовом пространстве-времени; теория относительности — о неевклидовом дифференцируемом многообразии; квантовая теория — о гильбертовом пространстве для волновой функции. И разные теории говорят о различных видах динамики этих структур: ньютоновских законах, принципах относительности, волновых уравнениях квантовой механики. Но из структуры и динамики мы можем получать только структуру и динамику. Это позволяет удовлетворительно объяснить различные высокоуровневые структурные и функциональные свойства, но сознательный опыт остается незатронутым. Никакой набор фактов о физической структуре и динамике не может составлять феноменологический факт.

Конечно, в определенном смысле физика универсума *должна* влечь за собой существование сознания, если *определить* физику как фундаментальную науку, из фактов и законов которой вытекает все остальное. Такая трактовка физики, однако, тривиализирует интересующий нас вопрос. Если включить в физику теории, специально разработанные для рассмотрения феномена сознания, при отсутствии более осно-

вательной мотивации, то мы можем получить «объяснение» сознания, но оно точно не будет редуктивным. Для нас лучше всего трактовать физику как фундаментальную науку, объясняющую наблюдения за внешним миром. Если бы из такой физики вытекали факты относительно сознания без придания центральной роли самому сознанию, то сознание действительно оказывалось бы редуктивно объяснимым. Однако по причинам, о которых я говорил, есть серьезное основание отрицать возможность таких редуктивных объяснений.

Почти все идеи относительно использования физики для объяснения сознания сфокусированы на наиболее загадочной части физики, а именно на квантовой механике. Это понятно: объяснение физикой сознания было бы чем-то необычным, а квантовая механика по общему мнению является самой необычной частью современной физики. Впрочем, в итоге создается впечатление, что она недостаточно необычна для решения этих задач.

Так, Пенроуз (Penrose 1994) считает, что ключом к пониманию сознания может быть теория, примиряющая квантовую теорию с общей теорией относительности. Он полагает, что еще не понятые гравитационные эффекты могут быть ответственны за коллапс квантовой волновой функции, привнося неалгоритмизируемый элемент в законы природы. Опираясь на идеи Хамероффа (Hameroff 1994), он допускает, что человеческое познание может зависеть от квантовых коллапсов в микротрубочках — белковых структурах, обнаруживаемых в цитоскелете нейронов. Более того, Пенроуз и Хамерофф не исключают, что квантовый коллапс в микротрубочках может быть физической основой сознательного опыта.

Эти идеи крайне спекулятивны, но по крайней мере можно представить, что они будут способствовать объяснению ряда элементов функционирования человеческого познания. Пенроуз считает, что неалгоритмизируемый элемент при коллапсе мог бы объяснить некоторые аспекты наших математических интуиций, которые, как он полагает, превосходят возможности любых алгоритмических систем. Хамерофф считает, что коллапс суперпозиционной волновой функции мог бы способствовать объяснению ряда аспектов принятия решений, характерных для людей. Но ничто из этого, похоже, не помогает объяснить сознательный опыт. Почему квантовые процессы в микротрубочках должны порождать сознание? Этот вопрос не менее труден, чем аналогичный вопрос о классических процессах в классическом мозге. Когда речь идет о проблеме опыта, неалгоритмические и алгоритмические процессы оказываются в одной лодке.

Некоторые полагали, что ключом к теории сознания могла бы стать *нелокальность* квантовой механики, о наличии которой, возможно, говорят недавние эксперименты, опирающиеся на парадокс

Эйнштейна—Подольского—Розена и на теорему Белла (см. подобные рассуждения в (Lahav and Shanks 1992)). Но даже если физика нелокальна, трудно понять, как это должно способствовать объяснению сознания. Даже при нелокальности физического процесса остается логически возможным, что этот процесс мог бы протекать при отсутствии сознания. Провал в объяснении не становится меньше.

Связь между сознанием и квантовой механикой чаще всего упоминается в контексте того факта, что согласно некоторым интерпретациям последней для коллапса волновой функции нужен производящий измерение сознательный наблюдатель. В соответствии с подобными интерпретациями, сознание играет центральную роль в динамике физической теории. Эти интерпретации вызывают множество вопросов, но в любом случае стоит заметить, что они никак не объясняют сознание. Скорее, они просто допускают существование сознания и применяют его в качестве вспомогательного средства для объяснения определенных физических феноменов. Можно встретить теории сознания, использующие это отношение (напр., Hodson 1988; Stapp 1993), но они точно не являются редуктивными¹³.

Нельзя исключить возможность того, что фундаментальные физические теории, такие как квантовая механика, будут играть ключевую роль в теории сознания. Не исключено, к примеру, что сознание окажется связанным с какими-то фундаментальными физическими свойствами или с определенными конфигурациями этих свойств. Или же такого рода связь окажется более тонкой. Тем не менее едва ли приходится надеяться на то, что подобная теория будет содержать сугубо физическое объяснение сознания. Когда речь идет о редуктивном объяснении, теории, базирующиеся на физике, оказываются ничем не лучше нейробиологических и когнитивных теорий.

6. Эволюционное объяснение

Даже те, кто принимает сознание всерьез, нередко симпатизируют идею эволюционного объяснения сознания. В конце концов, сознание — настолько широко распространенная и важная черта, что кажется, будто она должна была возникнуть в ходе эволюционного процесса по какой-то причине. В частности, естественно предположить, что сознание возникло для реализации какой-то функции, которая не могла бы быть реализована без него. Если бы мы смогли в достаточной степени прояснить эту функцию, то у нас возникло бы некое представление о том, почему оно существует.

К сожалению, эта идея преувеличивает возможности эволюционного объяснения. Процесс естественного отбора не может провести

границу между мной и моим зомбийным двойником. Эволюция отбирает свойства в соответствии с их функциональной ролью, а мой зомбийный двойник осуществляет все те функции, которые осуществляю и я сам, и ничуть не хуже меня; в частности, он оставляет столько же копий своих генов. Из этого следует, что одна лишь эволюция не может объяснить, почему в ходе эволюционного процесса возникли сознательные существа, а не зомби.

У кого-то могло бы возникнуть искушение сказать: «Но зомби *не могли бы* делать все то, что могу делать я». Однако мой зомбийный двойник по определению физически идентичен мне на протяжении всего своего существования, так что он уж точно ведет себя так же, как я. Тот, кто хочет усомниться в способностях зомби, должен поэтому найти ошибку в аргументах в начале этой главы, а не выдвигать возражение на данном этапе.

Чтобы взглянуть на сказанное в несколько ином свете, отметим, что реальная проблема, связанная с сознанием, состоит в том, чтобы объяснить принципы, благодаря которым сознание возникает из физических систем. Эти принципы — будь они концептуальными истинами, метафизическими необходимостями или естественными законами — предположительно одинаковы в пространстве и времени: если бы миллион лет назад где-то возникла моя физическая копия, она была бы столь же сознательной, как и я. Сами связующие принципы, таким образом, независимы от эволюционного процесса. Хотя эволюция может быть очень полезной при объяснении того, почему появились те или иные специфические физические системы, она не имеет отношения к объяснению связующих принципов, благодаря которым некоторые из этих систем обладают сознанием.

7. Куда ведет редуктивное объяснение?

Достаточно типичной является картина, когда люди соглашаются с критикой конкретных редуктивных концепций, но оговариваются: «Конечно, это не объясняет сознания, но давайте подождем, и объяснение будет найдено». Надеюсь, что наше обсуждение сделало очевидным то, что проблемы, связанные с подобными объяснениями сознания, имеют более фундаментальный характер. Проблемы с моделями и теориями, представленными в этой главе, касаются вовсе не деталей; по крайней мере нам не надо было рассматривать детали, чтобы увидеть их дефекты. Проблема — в общей объяснительной стратегии. Эти модели и теории попросту не есть теории такого рода, которые могли бы объяснить сознание.

Выдвижение все более утонченных редуктивных «объяснений» сознания неизбежно, но они будут содержать лишь все более и более утонченные объяснения когнитивных функций. Даже такие «революционные» прорывы, как использование коннекционистских сетей, нелинейной динамики, искусственной жизни и квантовой механики, приведут лишь ко все более сильным функциональным объяснениям. Это может сделать когнитивную науку очень интересной, но не устроит тайну сознания.

Любое объяснение, сформулированное исключительно в физических терминах, столкнется с той же самой проблемой. В конечном счете оно будет даваться в терминах структурных и динамических свойств физических процессов, и неважно, насколько утонченным оно будет, — оно будет выдавать только еще больше структуры и динамики. Хотя этого достаточно для рассмотрения большинства естественных феноменов, проблема сознания выходит за рамки проблемы объяснения структуры и функции, и поэтому здесь требуется объяснение какого-то нового типа.

Можно было бы предположить, что в итоге удалось бы отыскать такую объяснительную технику, которая объясняла бы не только структуру и функцию, но очень трудно понять, как это могло бы быть возможным, учитывая, что законы физики в конечном счете выражаются в терминах структуры и динамики. Существование сознания всегда будет дополнительным фактом относительно структурных и динамических фактов и поэтому всегда будет оставаться не объясненным физическими концепциями.

Таким образом, мы должны искать объяснения сознания где-то в другом месте. Мы точно не должны отказываться от объяснения; нам надо лишь отказаться от *редуктивного объяснения*. Возможность нередуктивного объяснения сознания остается открытой. Оно будет совершенно иным объяснением, требующим радикальных изменений в том, как мы представляем себе структуру мироздания. Но если мы произведем эти изменения, на горизонте могут замаячить контуры теории сознания.

Глава 4

Натуралистический дуализм

1. Аргумент против материализма

В предыдущей главе меня интересовал скорее объяснительный вопрос: «Можно ли объяснить сознание физическими теориями?», нежели онтологический вопрос: «Является ли само сознание физическим?» Но два этих вопроса тесно связаны, и в этой главе я извлеку онтологические следствия из аргументов предыдущей главы. В частности, отсутствие логической супервентности напрямую имплицирует ложность материализма: в мире имеются такие черты, которые не являются физическими. Главный аргумент в пользу этого тезиса выглядит следующим образом:

1. В нашем мире существует сознательный опыт.
2. Имеется логически возможный мир, физически идентичный нашему, в котором отсутствуют позитивные факты о сознании из нашего мира.
3. Поэтому факты о сознании — дополнительные факты нашего мира, обособленные от физических фактов.
4. Значит, материализм ложен.

Если физически идентичный зомбийный мир логически возможен, то из этого следует, что присутствие сознания — дополнительный факт о нашем мире, не гарантируемый одними лишь физическими фактами. Характер нашего мира не исчерпывается тем его характером, который основан на физических фактах; у него есть и добавочный характер, связанный с наличием сознания. Говоря словами Льюиса (Lewis 1990), сознание несет феноменальную информацию. Физические факты не вполне определяют способ существования нашего мира; факты сознания довершают дело.

Сходный вывод можно извлечь из логической возможности мира с инвертированным сознательным опытом. Такой мир физически идентичен нашему, но некоторые факты о сознательном опыте нашего мира отсутствуют в нем. Отсюда вытекает, что факты о сознательном опыте в нашем мире выходят за пределы физических фактов и что материализм ложен.

В любом случае, если сознание не является чем-то логически супервентным на физическом, материализм оказывается ложным. Отсутствие логической супервентности означает, что какой-то позитивный факт о нашем мире отсутствует в физически идентичном мире, так что он является фактом, выходящим за пределы физических фактов. Как и в главе 2, я трактую материализм как учение о том, что физические факты о нашем мире исчерпывают все факты в том смысле, что любой позитивный факт вытекает из физических фактов. Если зомбийные или инвертированные миры возможны, то физические факты не включают за собой все позитивные факты о нашем мире, и материализм оказывается ложным.

Здесь можно воспользоваться образом, предложенным Крипке. При творении мира, обеспечив наличие всех физических фактов, Бог должен был сделать и еще что-то. Он должен был обеспечить наличие фактов о сознании. Возможность зомбийных миров и инвертированных миров показывает, что у него был выбор. Мир мог бы оказаться лишенным сознательного опыта или мог бы содержать другой опыт даже при идентичности всех физических фактов. Для того чтобы факты о сознании были такими, какие они есть, мир должен был получить и иные черты.

Крах материализма ведет к какой-то из разновидностей дуализма: у мира есть как физические, так и нефизические черты. Отсутствие логической супервентности означает, что опыт фундаментально отличается от каких бы то ни было физических характеристик. Есть, однако, много видов дуализма, и важно понимать, куда именно ведет нас данный аргумент.

Дуализм какого рода?

Аргументы из предыдущей главы устанавливают, что сознание не является чем-то логически супервентным на физическом, но это не значит, что оно вообще не супервентно. Сознательный опыт, похоже, систематически зависит от физических структур в известных нам случаях, и ничто из упомянутых выше аргументов не говорит об обратном. К примеру, они не уменьшают правдоподобность того, что при воспроизведении моей физической структуры каким-то существом в нашем мире воспроизводился бы и мой сознательный опыт. Поэтому можно по-прежнему говорить о правдоподобности того, что мое сознание является естественно супервентным на физическом. И я буду развивать именно эту позицию — естественной супервентности при отсутствии логической супервентности.

Указанные аргументы не ведут нас к дуализму декартовского типа, допускающему особую ментальную субстанцию, оказывающую влияние на физические процессы. Наиболее достоверные данные современной науки говорят нам о том, что физический мир более или

менее каузально замкнут: любое физическое событие имеет физическую достаточную причину. Если так, то не остается места для того, чтобы ментальное «привидение в машине» могло выполнять какую-то дополнительную каузальную работу. Небольшая брешь может открываться наличием квантовой неопределенности, но в дальнейшем я покажу, что она, похоже, не может быть использована для наделения нефизического духа какой-то каузальной ролью. В любом случае все аргументы предыдущей главы оставляют вероятной возможность объяснения физических событий в физических терминах, и поэтому шаг к картезианскому дуализму был бы неоправданным.

Дуализм же, о котором пойдет здесь речь, — это разновидность дуализма *свойств*: сознательный опыт предполагает такие свойства индивида, которые не могут быть выведены из физических свойств данного индивида, хотя и могут закономерно зависеть от этих свойств. Сознание — это такая *черт* мира, которая выходит за пределы его физических черт. Это не означает, что оно есть особая «субстанция»; вопрос о том, при каких условиях можно было бы говорить о субстанциальном дуализме, кажется мне очень туманным. Всё, что мы знаем, — так это то, что в нашем мире существуют свойства индивидов — феноменальные свойства, — которые являются онтологически независимыми от физических свойств.

Существует и более слабая разновидность дуализма свойств, с которой не следует смешивать ту позицию, о которой мы сейчас говорим. Иногда утверждается, что дуализм свойств имеется везде, где свойства сами по себе не являются физическими или не могут быть напрямую редуцированы к таким свойствам. В этом смысле даже биологическая приспособленность не является физическим свойством. Но такой «дуализм» очень слаб. В свойствах, подобных приспособленности, нет ничего *фундаментально* нового в онтологическом плане, поскольку они все же логически супервентны на микрофизических свойствах. Эта разновидность дуализма свойств полностью совместима с материализмом. Отстаиваемый же мной дуализм свойств предполагает наличие фундаментально новых черт мира. Поскольку эти свойства логически не супервентны даже на микрофизических свойствах, они нефизичны в гораздо более сильном смысле. Говоря о дуализме свойств и о нефизических свойствах, я имею в виду это более сильное возврение и этот более сильный смысл нефизичности.

Остается, впрочем, вероятным, что, хотя сознание не может быть выведено из некоей физической основы, оно *порождается* ею. И позиция, к которой мы пришли, состоит в том, что сознание возникает из физического субстрата благодаря определенным контингентным законам природы, которые сами по себе не являются следствием физических законов. Этого взгляда неявно придерживаются многие из тех, кто

считает себя материалистами. Нередко можно услышать такую фразу: «Конечно, я материалист; ментальное уж точно порождается мозгом». Здесь должно было бы настороживать само присутствие слова «порождается». К примеру, обычно мы не говорим «обучение порождается мозгом» — а если бы мы сказали так, то мы употребили бы слово «порождается» в темпоральном смысле. Более естественным было бы, скорее, сказать, что обучение *есть* процесс в мозге. Сам факт, что ментальное должно *порождаться* мозгом, указывает на то, что здесь случается что-то, выходящее за пределы физических фактов¹.

Кто-то подумает, что это воззрение должно быть сочтено скорее материалистическим, чем дуалистическим, так как оно говорит о сильной закономерной зависимости феноменальных фактов от физических фактов и оставляет физическое автономной областью. Разумеется, нет большого смысла спорить о названии, хотя мне кажется, что существование дополнительных контингентных фактов, выходящих за пределы физических фактов, достаточно существенно модифицирует общепринятые материалистические взгляды, чтобы заслуживать иного обозначения. И если для материализма требуется лишь закономерная связь всех фактов с физическими фактами, то он и впрямь окажется весьма размытым учением.

Но хотя это воззрение и является разновидностью дуализма, в нем нет ничего антинаучного или сверхъестественного. Лучше всего его можно мыслить следующим образом. Физика постулирует в мире множество фундаментальных параметров: пространство-время, масса-энергия, заряд, спин и т. д. Она также признает существование множества фундаментальных законов, определяющих отношения между этими фундаментальными параметрами. Фундаментальные параметры не могут быть объяснены в терминах еще более фундаментальных параметров, а фундаментальные законы не могут быть объяснены в терминах еще более фундаментальных законов; они должны трактоваться просто как изначальные. После же установления фундаментальных законов и распределения фундаментальных параметров из них вытекает почти всё в мире. Именно поэтому о фундаментальной теории физического иногда говорят как о «теории всего». Однако тот факт, что сознание не супервентно на физических параметрах, показывает нам, что теория физического *не в полной мере* является теорией всего. Чтобы сознание оказалось в области фундаментальной теории, мы должны ввести *новые* фундаментальные свойства и законы.

В книге «Мечты об окончательной теории» физик Стивен Вайнберг (Weinberg 1992) отмечает, что характерной чертой фундаментальной теории физического является то, что она протягивает объяснительную цепь до самых вершин, в конечном счете объясняя всё. Однако

он вынужден признать, что подобная теория может и не объяснить сознание. Он говорит, что в лучшем случае мы можем объяснить «объективные корреляты» сознания. «Это может и не быть объяснением сознания, хотя и быть чем-то близким к нему» (с. 45). Но, разумеется, недостаточно близким. Такая теория не объясняет всё, что происходит в мире. Чтобы не впасть в непоследовательность, мы должны признать, что подлинно окончательная теория нуждается в некоем дополнительном компоненте.

И здесь есть два пути. Не исключено, что мы могли бы признать сам опыт одним из фундаментальных параметров мира наряду с пространством-временем, спином, зарядом и т. п. Иными словами, какие-то феноменальные свойства будут признаны нами в качестве базовых свойств. Альтернативный вариант состоит в допущении какого-то другого класса новых фундаментальных свойств, из которых извлекаются феноменальные свойства. Приведенные ранее аргументы показали, что физические свойства не могут быть этими свойствами, но, быть может, речь идет о новой разновидности нефизических свойств, на которых логически супервентны феноменальные свойства. Подобные свойства относились бы к опыту так же, как базовые физические свойства относятся к таким нефундаментальным свойствам, как температура. Мы могли бы назвать эти свойства *протофеноменальными* свойствами, так как, хотя они сами по себе и не феноменальны, феноменальное может быть их совокупным результатом. Конечно, очень трудно вообразить, на что могло бы походить протофеноменальное свойство, но мы не можем исключать возможности их существования. Тем не менее по большей части я буду говорить так, будто эти фундаментальные свойства феноменальны уже сами по себе.

Там, где имеются новые фундаментальные свойства, имеются и новые фундаментальные законы. В данном случае фундаментальные законы будут *психофизическими* законами, специфицирующими то, каким образом феноменальные (или протофеноменальные) свойства зависят от физических свойств. Эти законы не будут пересекаться с физическими законами; физические законы уже формируют замкнутую систему. Они будут *супервентностными* законами, говорящими нам о том, как опыт порождается физическими процессами. Мы видели, что зависимость опыта от физического не может быть извлечена из физических законов, так что любая окончательная теория должна включать в себя законы подобного рода.

Разумеется, на этой стадии у нас может быть лишь самое отдаленное представление о том, на что будет похожа соответствующая фундаментальная теория или какими будут эти фундаментальные психофизические законы. Но не лишено резонов мнение, что подобная

теория существует. Имеется серьезное основание полагать, что между физическими процессами и сознательным опытом есть некое закономерное отношение, а любое закономерное отношение должно базироваться на фундаментальных законах. Пример физики показывает нам, что фундаментальные законы, как правило, просты и элегантны; и мы должны ожидать того же от фундаментальных законов в теории сознания. После того как мы сопроводим фундаментальную теорию физического фундаментальной теорией сознания, мы сможем действительно получить теорию всего. При наличии базовых физических и психофизических законов, а также распределения фундаментальных свойств, мы можем ожидать того, что из них будут вытекать все факты о мире. Разработка подобной теории не будет простым делом, но ее осуществление должно быть в принципе возможным.

Происходящее здесь с сознанием в известном смысле аналогично тому, что случилось с электромагнетизмом в девятнадцатом столетии. Была предпринята попытка объяснить электромагнитные феномены в терминах уже известных физических законов, с использованием механических принципов и т. п., но она не привела к успеху. Оказалось, что для объяснения феноменов электромагнетизма надо признать фундаментальность таких параметров, как электромагнитный заряд и электромагнитные силы, и Максвелл ввел новые фундаментальные законы электромагнетизма. Только так могли быть объяснены эти феномены. Так же и для объяснения сознания недостаточно параметров и законов теории физического. Для теории сознания нужны новые фундаментальные параметры и законы.

Эта концепция полностью совместима с современным научным мировоззрением, и она является в полной мере натуралистической. Согласно данной концепции, мир по-прежнему является собой сеть фундаментальных свойств, соотнесенных базовыми законами, и всё в конечном счете должно объясняться в этих терминах. Происходит лишь расширение набора свойств и законов, подобно тому как это было и в случае Максвелла. Более того, ничто в этой концепции не противоречит каким-либо элементам теории физического; скорее можно говорить о дополнении последней. Теория физического — это теория физических процессов, а психофизическая теория говорит нам о том, как эти процессы порождают сознание.

Чтобы выразить дух той позиции, которую я отстаиваю, я имею ее *натуралистическим дуализмом*. Натуралистический он потому, что здесь предполагается, что всё обусловлено сетью базовых свойств и законов, и потому, что он совместим со всеми результатами современной науки. И подобно натуралистическим теориям в других сферах, это воззрение допускает *объяснение* сознания в терминах базовых

естественных законов. Сознание не обязано быть каким-то уж осо-
бо трансцендентальным; это просто один из естественных феноменов.
Происходит лишь расширение нашей картины природы. Иногда «на-
турализм» считают синонимом «материализма», но мне кажется, что
приверженность натуралистическому пониманию мира может остаться
даже в случае краха материализма. (Если читатель сомневается в этом,
я могу указать на остальные части книги как на подтверждение ска-
занного.) Кто-то мог бы усмотреть в названии этой позиции некую
иронию, но важнее всего то, что оно передает главный посыл: одоб-
рение дуализма может и не сопровождаться разговорами о тайнах.

Сторонники подобного дуализма могут быть по своим настрое-
ниям в известном отношении ближе к материалистам, чем к дуали-
стам иного рода. Отчасти это связано с тем, что они не допускают
каких-либо элементов трансцендентального, а также с их привержен-
ностью к естественным объяснениям и идеи физического причинения
поведения. Наоборот, отрицая привидение в машине, эта позиция
избегает самых неправдоподобных элементов традиционных дуали-
стических воззрений. Часто можно услышать, что успехи когнитивной
науки и нейронауки дискредитируют дуализм, но не все виды дуализма
в равной степени затрагиваются этим обстоятельством. Все эти успехи
основаны на физических объяснениях поведения и других физических
феноменов и поэтому не позволяют отделить материализм от натура-
листического дуализма.

Два финальных замечания. Кто-то может все еще удивляться, по-
чему, если опыт фундаментален, его нельзя считать *физическими* свой-
ством. В конце концов, разве физика как раз и не является наукой
о подлинно фундаментальном? Ответ: конечно, если мы определим
физику подобным образом, опыт и в самом деле можно будет счесть
физическими свойством, и супервентностные законы окажутся законы
физики. Но при более естественном толковании «физики» и «физи-
ческого» опыт не будет таковым. Опыт не является фундаментальным
свойством такого рода, в котором нуждается физика при построении
теории внешнего мира; физика составляет замкнутую и непротиворе-
чивую теорию даже и без опыта. Если признать возможность зомбий-
ного мира, то становится очевидной избыточность опыта для физики
в ее обычном понимании. Более естественно поэтому рассматривать
опыт в качестве фундаментального нефизического свойства, а психо-
физические законы — как фундаментальные законы природы, не яв-
ляющиеся законами физики. Впрочем, если суть этого воззрения ясна,
то терминологические споры не будут иметь большого значения.

Я должен также заметить, что, хотя я называю это воззрение ду-
ализмом, возможна ситуация, когда оно может предстать в качестве

некоей разновидности монизма. Физическое и феноменальное могут оказаться двумя разными аспектами одного охватывающего их рода, подобно тому как двумя аспектами одного-единственного рода являются материя и энергия. Ничто из сказанного мной не исключает такой возможности, и в действительности я питаю определенную симпатию к этой идеи. Но даже если какой-то монизм и окажется истинным, он не может быть *материалистическим* монизмом. Он должен быть чем-то более широким.

В связи с аргументом против материализма, приведенным в начале этой главы, можно высказать целый ряд возражений. Некоторые из них атакуют вторую посылку — отрицание логической супервентности. Я рассматривал такие возражения в предыдущей главе. Здесь я буду иметь дело с возражениями относительно перехода от отсутствия логической супервентности к ложности материализма. Самые серьезные возражения такого рода связаны с отсылкой к апостериорной необходимости. Я обсуджу их в следующем параграфе. Пока же поговорим о менее серьезных возражениях.

Иногда доказывают, что сознание могло бы быть *эмерджентным* свойством в том смысле, который все же совместим с материализмом. В недавних работах о сложных системах и искусственной жизни нередко утверждается, что, хотя эмерджентные свойства нельзя предсказать из низкоуровневых свойств, они тем не менее имеют физическую природу. В качестве примеров можно привести эмерджентное появление самоорганизации в биологических системах или эмерджентное возникновение некоего подобия стай при симуляции птиц на основе набора простых правил (Langton 1990; Reynolds 1987). Однако такие эмерджентные свойства не могут рассматриваться в качестве аналогии сознания. Интерес подобным случаям придает то, что соответствующие свойства не являются очевидными следствиями низкоуровневых законов; но они все же логически супервентны на низкоуровневых *фактах*. При наличии *всех* физических фактов об изменении во времени такого рода систем из них можно напрямую вывести факт о том, что в них происходит процесс самоорганизации. И именно этого можно было бы ожидать, учитывая то обстоятельство, что такие свойства, как самоорганизация или образование стай, являются свойствами, имеющими непосредственное отношение к функциям и структуре.

Если сознание — эмерджентное свойство, оно эмерджентно в гораздо более сильном смысле. Согласно более сильному понятию эмерджентности, использовавшемуся британскими эмерджентистами (к примеру, Броудом (Broad 1925)), эмерджентные свойства не могут быть предсказаны даже при наличии всей совокупности низкоуровневых физических фактов. И не лишено оснований утверждение (которое

Возражения

и высказывали британские эмерджентисты), что сознательный опыт эмерджентен в этом смысле. Однако подобную эмерджентность лучше всего истолковать в качестве одной из разновидностей дуализма свойств. В отличие от тех более «невинных» примеров эмерджентности, эта ее сильная разновидность требует новых фундаментальных законов для появления эмерджентных свойств.

Другое возражение состоит в том, что сознание и физическое могли бы оказаться двумя аспектами одной и той же вещи, подобно тому как Утренняя звезда и Вечерняя звезда являются двумя аспектами Венеры. Если это так, то в известном смысле сознание было бы физическим. Но опять-таки мы должны спросить: вытекает ли феноменальный аспект из физического аспекта? Если да, то это разновидность материализма, но тогда мы возвращаемся к аргументам из главы 3. Если нет, то феноменальный аспект привносит дополнительную контингентность в мир, выходящую за рамки физического аспекта, и дуальность этих аспектов дает нам некий дуализм свойств. Не исключено, что может оказаться так, что дуальность физического и феноменального может быть подведена под более широкий монизм, но он не будет монизмом одного лишь физического.

Намек на третье возражение можно найти у Серла (Searle 1992). Как и я, Серл считает, что сознание всего лишь естественно супервентивно на физическом. Он допускает логическую возможность зомбийных копий, полагая, что сознание лишь *каузально* порождается состояниями мозга. Он, однако, отрицает, что эта позиция является собой разновидность дуализма, даже если говорить только о дуализме свойств. Могло бы показаться, что речь идет лишь о терминологическом разногласии, но Серл настаивает, что онтологический статус сознания не отличается от онтологического статуса таких физических свойств, как текучесть, так что это разногласие имеет *не только* терминологический характер. Аргумент Серла в пользу недуалистичности этого взгляда состоит в том, что аналогичная картина наблюдается и в других случаях: к примеру, H_2O каузально порождает текучесть, но никто не является дуалистом в вопросе о текучести.

Кажется, однако, очевидным, что это ложная аналогия. При наличии всех микрофизических фактов о конкретной совокупности H_2O логически невозможным оказывается то, что эти факты могли бы иметься в наличии при отсутствии текучести. Понятие нетекучей копии массы текучей H_2O попросту некогерентно. Из этого следует, что отношение между микрофизическими фактами и текучестью является гораздо более тесным, чем простое каузальное отношение. Микрофизические характеристики *не каузально порождают* текучесть, а *конституируют* ее. Это полностью отличается от того, что происходит

в случае сознания, так что аналогия не проходит. Сознание обладает онтологической новизной в гораздо более существенном смысле, чем тот, в котором ей обладает текучесть².

Наконец, кто-то мог бы подумать, что выдвигаемый мной аргумент в пользу дуализма напоминает аргумент, выдвигавшийся Декартом. Декарт доказывал, что он мог бы вообразить свой ум существующим отдельно от своего тела, и поэтому его ум не мог быть тождествен телу. Такого рода аргументация, по общему мнению, ошибочна: на основании одного того, что кто-то может вообразить нетождественность *A* и *B*, не следует, что они нетождественны (вспомним об Утренней и Вечерней звезде). Не могла бы сходная ошибка допускаться и в моем аргументе? Зомбийный мир показывает лишь *представимость* того, что некто мог бы находиться в каком-то физическом состоянии при отсутствии сознания, но он не показывает нетождественность такого состояния и сознания.

Подобное возражение, однако, основано на неверном понимании данного аргумента. Существенно важным является то, что мой аргумент касается не вопроса о *тождестве*, а вопроса о *супервентности*. Его вид не таков: «Можно вообразить физическое состояние *P* без сознания, поэтому сознание не является физическим состоянием *P*». Скорее он таков: «Можно вообразить наличие *всех* физических фактов при отсутствии фактов о сознании, и поэтому физические факты не исчерпывают всех фактов». Это совершенно другой тип аргумента. В общем, модальные аргументы в пользу дуализма, выраженные в терминах тождества, менее убедительны, чем модальные аргументы, выраженные в терминах супервентности; и это одна из причин того, что я всюду формулировал их в терминах супервентности и полностью избегал рассуждений о тождестве. Мне кажется, что проблемы, связанные с супервентностью, являются здесь наиболее фундаментальными.

И тем не менее можно было бы попробовать ответить на этот аргумент, прибегая к стратегии, аналогичной той, которая применяется при ответе Декарту. К примеру, можно было бы заметить, что моя стратегия все же опирается на заключения от представимости к возможности, которые можно попытаться оспорить. В следующем параграфе я рассмотрю стратегии, которые можно было бы использовать для этой цели.

2. Возражения от апостериорной необходимости*

Распространенным ответом на подобный аргумент является возражение, что он устанавливает лишь логическую возможность зомбийного мира, совершенно отличную от метафизической возможности. Для ло-

гической возможности достаточно концептуальной когерентности, но метафизическая возможность является более ограниченной. Этот тезис также нередко высказывается в виде указания на отличие *представимости* от подлинной возможности. Хотя зомбийный мир и может быть представимым, для демонстрации его возможности в метафизическом смысле, пригодном для опровержения материализма, требуется нечто большее.

Это возражение чаще всего сопровождается ссылкой на работу Крипке «Именование и необходимость» (Кріпке 1980), где показано существование необходимых истин, таких как «вода есть H_2O », необходимость которых познается только апостериори. Говоря словами тех, кто высказывает подобное возражение: хотя логически возможно то, что вода не есть H_2O , это не является метафизически возможным. И вполне естественно предположить, что аналогичным образом и зомби могли бы быть логически возможными, но метафизически невозможными. И если это так, то не исключено, что этого было бы достаточно для спасения материализма.

Такова самая популярная стратегия материалистов, убежденных в отсутствии выводимости феноменальных понятий из физических понятий. Согласно этой позиции, концептуальный провал может существовать без метафизического провала. Она заманчива тем, что обещает всерьез относиться к сознанию, сохраняя вместе с тем материалистический характер. К сожалению, при более тщательном изучении в ней обнаруживаются явные дефекты. Понятие апостериорной необходимости не может нести ту ношу, в которой нуждается этот аргумент, и в этом контексте, по сути, просто отвлекает внимание³.

Лучше всего убедиться в этом, использовав двумерную модель апостериорной необходимости из четвертого параграфа главы 2. Вспомним, что согласно этой модели с любым понятием связано два интенсионала (функции от возможных миров к референтам): первичный интенсионал (определенный априори), фиксирующий референт в актуальном мире, и вторичный интенсионал (определенный апостериори), осуществляющий референцию в контрфактических мирах. Первичный интенсионал, связанный с «водой», — это что-то вроде «водянистой материи». Вторичный же интенсионал — « H_2O », получаемый из первичного интенсионала путем применения каплановского оператора *dthat*: «*dthat* (водянистая материя)» указывает на H_2O во всех возможных мирах, так как водянистая материя является H_2O в актуальном мире.

«Логическая возможность» сводится к возможной истинности какого-либо утверждения, оцениваемого в соответствии с входящими в него первичными интенсионалами (это то, что я назвал возможностью-1 в главе 2). Первичные интенсионалы «воды» и « H_2O » различны,

а значит, в этом смысле логически возможно, что вода не есть H_2O . «Метафизическая возможность» сводится к возможной истинности какого-либо утверждения, оцениваемого в соответствии с входящими в него вторичными интенсионалами (возможность-2). Вторичные интенсионалы «воды» и « H_2O » одинаковы, и поэтому то, что вода есть H_2O , метафизически необходимо.

Данное возражение сводится, стало быть, к тому тезису, что при использовании аргументов от представимости и т. п. мы продемонстрировали возможность зомбийного мира на основе *первичных* интенсионалов соответствующих понятий, но не на основе более уместных *вторичных* интенсионалов. Хотя первичные интенсионалы феноменальных понятий могут и не соответствовать первичным интенсионалам какого-либо физического понятия, их вторичные интенсионалы могут быть одинаковыми. И если это так, то феноменальные и физические/функциональные понятия могут апостериори указывать на одни и те же свойства, несмотря на их априорное различие. Подобное возражение могло бы исходить от защитника «психофункционализма» (см. Block 1980), апостериори отождествляющего феноменальные и функциональные свойства, или от сторонника той позиции, при которой феноменальные свойства апостериори отождествляются с нейрофизиологическими свойствами определенного рода.

Убедиться в том, что ничто из сказанного не влияет на мой аргумент в пользу дуализма, проще всего, отметив, что этот аргумент проходит, если мы концентрируемся только на первичном интенсионале и игнорируем вторичный интенсионал. Во второй главе мы видели, что именно первичный интенсионал имеет наибольшее значение при объяснении, но он может сослужить нам хорошую службу и при аргументации в пользу дуализма. Ведь независимо от того, совпадают первичные и вторичные интенсионалы или нет, первичный интенсионал определяет такое свойство объектов в возможных мирах, к которому невозможно придраться. Свойство быть водянистой материи — абсолютно законное свойство, даже если оно и не совпадает со свойством быть H_2O . Если мы сможем показать, что в некоторых возможных мирах, физически идентичных нашему, отсутствует свойство, вводимое первичным интенсионалом, мы придем к дуализму.

Именно это и было сделано с сознанием. Мы видели физически идентичные нашему миры, в которых отсутствует сознание, трактуемое согласно его первичному интенсионалу. Этого различия миров достаточно для демонстрации существования таких свойств нашего мира, которые выходят за рамки физических свойств. Аналогично, если бы мы смогли показать, что в некоторых физически тождественных нашему мирах отсутствует водянистая материя, мы пришли бы к водяному

дуализму — так же как и в том случае, когда мы установили бы, что в некоторых мирах, физически идентичных нашему, отсутствует H_2O . Важно то, что различие относительно первичного интенсионала может быть установлено независимо от апостериорных факторов, так что соображения, касающиеся апостериорной необходимости, не имеют отношения к данному вопросу.

(Здесь надо сделать два технических замечания. Строго говоря, первичный интенсионал определяет *центрированное* свойство объекта в возможном мире (или отношение между объектами и центрами), поскольку он применяется к *центризованным* возможным мирам. Но эта относительность не может сыграть на руку нашему критику. При конкретизации местонахождения центра первичный интенсионал определяет безупречное неиндексикальное свойство; и все аргументы из главы 3 проходят, даже если местонахождение центра включается в супервентностную базу. К примеру, даже если факты о мире, известные Мэри, включают факты о ее местонахождении, это не позволит ей узнать, каково это — видеть красное.)

Кого-то мог бы беспокоить также тот факт, что понятие сознания, судя по всему, отсутствует в центре зомбийского мира, хотя применение первичного интенсионала могло бы требовать присутствия соответствующего понятия в центре данного мира. (Можно было бы даже начать беспокоиться относительно применения понятия «*зомби*»!) Думаю, что ситуация не столь однозначна — первичные интенсионалы не обязательно требуют наличия оригинального понятия, — но в любом случае мы можем полностью рассеять эти тревоги, просто рассматривая *частично зомбийский* мир: мир, в котором я нахожусь в центре, обладаю сознанием и всеми надлежащими понятиями, притом что некоторые другие люди — зомби.)

Вывод о нерелевантности апостериорной необходимости можно подкрепить наблюдением о совпадении первичных и вторичных интенсионалов сознания. В актуальном мире состояние оказывается сознательным опытом при наличии связанного с ним феноменального чувства, и то же самое можно сказать относительно сознательности чего бы то ни было в контрафактическом мире. Различие между первичным и вторичным интенсионалами понятия воды отражает то обстоятельство, что в каком-то контрафактическом мире может иметься нечто, выглядящее и ощущаемое в точности как вода, но в действительности являющееся не водой, а всего лишь водянистой материей. Но если нечто чувствуется как сознательный опыт, даже в контрафактическом мире, то оно *есть* сознательный опыт. Для сознательного опыта в любом возможном мире нужно лишь наличие определенного чувствования. (Крипке высказывает аналогичное соображение, правда не в терминах значения, а в терминах сущностных свойств.)

Но даже если настаивать на различии первичного и вторичного интенсионалов в данном случае, аргумент все равно будет работать. Мы просто фокусируемся на первичном интенсионале, используемом для фиксации референции, как и было сделано выше. К примеру, если «сознание» трактуется как «*dthat* (имеет феноменальное чувство)», то мы просто фокусируемся на интенсионале «имеет феноменальное чувство». Аргументы главы 3 устанавливают наличие возможного мира, в котором у моей копии отсутствует феноменальное чувство, и поэтому свойство, состоящее в обладании феноменальным чувством, является фактом, выходящим за рамки физических фактов, и аргумент в пользу дуализма проходит⁴.

В самом общем виде суть дела можно выразить, сказав, что крипкеанская апостериорная необходимость никоим образом не делает какие-либо логически возможные миры невозможными. Она просто говорит нам о том, что некоторым из этих миров дается неверное описание, когда мы применяем термины в соответствии с их первичными интенсионалами, а не в соответствии с более уместными вторичными интенсионалами. Кто-нибудь мог бы априори счесть возможным, что вода есть не H_2O , а XYZ . Представляя это, он воображает что-то вроде мира, в котором XYZ является жидкостью, заполняющей океаны и озера. Анализ Крипке, однако, показывает, что в силу особенностей актуального мира мы неверно описываем тот мир, говоря, что XYZ является в нем водой, поскольку мы описываем его с использованием первичного интенсионала, а не более уместного вторичного интенсионала. Строго говоря, это мир, в котором XYZ является водянистой матерью. Данные соображения не могут показать невозможность этого кажущегося возможным мира; они просто показывают нам то, каким образом надо правильно описывать этот мир.

Как мы видели в главе 2, крипкеанские соображения демонстрируют нам, что вторичный интенсионал $F_a : W \rightarrow R$ иногда отличается от первичного интенсионала $f : W^* \rightarrow R$. Это накладывает некоторые апостериорные ограничения на условия применения понятий, однако релевантное пространство миров совершенно не меняется; единственное различие аргументов двух функций касается местоположения центра. Поэтому, хотя возможность *утверждений* и может допускать разделение на два вида, существует только один релевантный вид возможности *миров*.

Из этого следует, что если можно представить мир, физически идентичный нашему, но такой, в котором отсутствуют какие-то позитивные характеристики нашего мира, то никакие рассуждения о дизайне терминов, таких как «сознание», не смогут исключить метафизической возможности этого мира. Мы можем просто забыть о семантике этих терминов и отметить, что в соответствующем возможном

мире явным образом *что-то* отсутствует — неважно, называем ли мы это «сознанием». Крипкеанские соображения в лучшем случае могли бы говорить нам о том, как уместнее всего было бы описывать этот мир и его соответствующие черты, но они не могут повлиять на его возможность; и одной лишь возможности такого мира, вне зависимости от его описания, достаточно для того, чтобы аргумент в пользу дуализма достиг своей цели.

Альтернативная стратегия

Для уклонения от дуализма можно было бы использовать апостериорную необходимость и иначе. Можно было бы попытаться доказать, что при утверждении о *физической идентичности* зомбийного мира нашему мы неверно описываем его. Так же как мир XYZ кажется содержащим воду, но в действительности не содержит ее, зомбийный мир *кажется* физически идентичным нашему, но не является таковым. Это может звучать странно, но данный тезис можно пояснить. Оппонент мог бы заявить, что некоторые свойства, существенные для физического устройства мира, недоступны для физического исследования. Представляя «физически идентичный» мир, на деле мы представляем мир, идентичный лишь с позиции физических изысканий, но отличающийся в плане недоступных сущностных свойств, которые вместе с тем гарантируют наличие сознания.

Дело, к примеру, могло бы обстоять так, что для того чтобы нечто могло считаться электроном в контрафактическом мире, было бы недостаточно того, что это нечто каузально соотносится с другими физическими сущностями так же, как и электрон. Для этого могла бы также требоваться и скрытая сущность электронности. Согласно этому представлению, понятие электрона есть нечто вроде «*dthat* (сущность, играющая роль электрона)». Референция к электронам фиксируется внешней характеристикой, но затем обретает жесткость таким образом, что в контрафактических мирах отбираются сущности с той же самой внутренней природой независимо от того, играют ли они там надлежащую роль, и поэтому сущности, которые играют такую роль в подобных мирах, считаются электронами лишь в том случае, если они наделены соответствующей внутренней природой. То же самое могло бы оказаться верным для таких свойств, как масса, которую можно было бы истолковать как «*dthat* (сущность, играющая роль массы)». Сущностная природа электронов или массы была бы в таком случае скрытой от физической теории, лишь внешним образом характеризующей электроны и массу. И тогда релевантные сущностные свойства могли бы оказаться феноменальными или протофеноменальными, а их реализация могла бы гарантировать существование сознания в нашем мире.

Если бы дело действительно обстояло таким образом, представляемый нами зомбийный мир был бы лишен этих скрытых сущностных свойств и поэтому был бы физически не идентичен нашему миру. Зомбийный мир не отличался бы от нашего при его оценке сообразно первичным интенсионалам физических предикатов, применяемым на основе внешних отношений, но отличался бы от него при его оценке сообразно вторичным интенсионалам, предполагающим отсылку к скрытой сущности. В такой ситуации сознательный опыт все же мог бы оказываться «метафизически» супервентным на физических свойствах. (Очень близкий аргумент выдвигает Максвелл (Maxwell 1978), и он также вытекает из подхода Локвуда (Lockwood 1989). Согласно Максвеллу, главная идея состоит в том, что, хотя мы и не можем нейтрально проанализировать феноменальные понятия так, чтобы они указывали на лежащие в их основе физические свойства, мы можем нейтрально проанализировать физические понятия так, что они могли бы указывать на лежащие в их основе феноменальные свойства⁵.)

Это возражение во многих отношениях более интересно, чем предыдущее. Конечно, оно опирается на спекулятивную метафизику, но это не мешает ему быть когерентным. Более предметным ответом было бы замечание, что оно опирается на неверное представление о семантике физических терминов. Можно было бы попробовать доказать, что применение физических предикатов даже апостериори основывается на внешних отношениях между физическими сущностями независимо от каких-либо скрытых свойств. Это чисто концептуальный вопрос: если электроны в нашем мире наделены скрытымиprotoфеноменальными свойствами, стали бы мы называть идентичную в иных отношениях контрафактическую сущность, лишенную подобных свойств, электроном? Думаю, да. Референция к электронам не только фиксируется ролью, играемой электронами в теории; само понятие электрона определяется этой ролью, детерминирующей применение этого понятия во всех мирах. Понятие электрона, имеющего все внешние свойства, актуально присущие протонам, не кажется когерентным, как и представление о том, что в каком-то мире масса играет роль, которую в нашем мире реально играет заряд. Сформулированная выше семантическая концепция предсказывает когерентность подобных представлений и поэтому неверно характеризует данные понятия.

Семантические интуиции могут разниться, но, как обычно, можно дать и такой ответ, который окажется глубже семантических интуиций. Даже если мы допустим, что какие-то скрытые свойства могли бы конституировать физические свойства, различие между этим воззрением и отстаивавшейся мной позицией дуализма свойств будет незначительным. По-прежнему будет верно, что мир наделен феноменальными

свойствами, не фиксируемыми свойствами, которые открывает физика. Удостоверившись в том, что с позиции наших физических теорий некий мир идентичен нашему, Бог должен был затратить дополнительные усилия, чтобы сделать такой мир полностью идентичным нашему. Дуализм «физических» и «нефизических» свойств заменяется в этой концепции дуализмом «доступных» и «скрытых» физических свойств, но суть при этом не меняется.

Я еще вернусь к концепции, согласно которой физические сущности имеют внутреннюю протофеноменальную природу, но ее метафизические аспекты мало зависят от трактовки семантики физических предикатов. Как и раньше, вторичные интенсионалы и апостериорная необходимость производят лишь семантическое, но не метафизическое различие. Независимо от того, как артикулируется эта концепция, она признает фундаментальность феноменальных или протофеноменальных свойств и поэтому остается ближе к дуализму (или, возможно, к идеализму или нейтральному монизму — позже я еще обсуджу это), чем к материализму.

Сильная метафизическая необходимость

Обсуждавшийся выше двумерный анализ показывает, что привлечение крипкеанской апостериорной необходимости не может нейтрализовать аргумент от супервентности. Подобная необходимость не накладывает апостериорные ограничения на пространство возможных миров, а лишь ограничивает способ использования определенных терминов для их описания; поэтому если имеется логически возможный мир, идентичный нашему во всех физических, но не во всех позитивных аспектах, то такие соображения не могут служить отрицанию его метафизической возможности.

Несмотря на это, кто-нибудь может заявить, что соответствующие миры *тем не менее* могли бы быть метафизически невозможными, основываясь на представлении о том, что существует такая модальность метафизической возможности, которая отличается от логической возможности и является более ограниченной по сравнению с последней, причем ее наличие связано с обстоятельствами, независимыми от крипкеанских соображений. Согласно этому взрыванию, метафизически возможные миры не столь многочисленны, как логически возможные миры, и апостериорная необходимость определенных утверждений может определяться факторами, совершенно независимыми от семантики используемых в них терминов. Мы можем назвать эту гипотетическую модальность *сильной метафизической необходимостью* в противоположность *слабой метафизической необходимости*, которая вводится в крипкеанской модели.

Эта позиция допускает наличие полностью представимых — даже в соответствии с самыми строгими критериями представимости — миров, которые тем не менее оказываются невозможными. Провал между представимостью и возможностью более глубок, чем все остальные. В определенном смысле истинность таких утверждений, как «вода есть XYZ», представима, но невозможна; однако подобные примеры никогда не исключают возможности какого-либо представимого мира. Они просто иллюстрируют неверное описание подобного мира. Сильная метафизическая необходимость идет дальше. Здесь «зомбийный мир» может быть корректным описанием представляемого нами мира, даже с позиции вторичных интенсионалов. Тем не менее он не является метафизически возможным⁶.

Краткий ответ на это возражение может состоять в том, что у нас нет основания считать, что такая модальность существует. Такие «метафизические необходимости» будут накладывать фундаментальные и необъяснимые ограничения на пространство возможных миров. Признание фундаментальных и необъяснимых фактов *нашего* мира может быть оправданным, но наличие подобных фактов в пространстве возможных миров было бы чем-то весьма необычным. В области возможного (в отличие от области естественного) нет места для таких произвольных ограничений.

Данная позиция не может быть поддержана какой-либо аналогией, так как нужные для этого аналогии отсутствуют⁷. Мы уже видели, что аналогии с необходимостью таких положений, как «вода есть H_2O », «Геспер есть Фосфор» и т. д., проваливаются, так как эти примеры требуют одного-единственного пространства миров. В самом деле, если какие-то миры логически возможны, но метафизически невозможны, то кажется, что мы никогда не смогли бы узнать об этом. По предположению эта информация недоступна априори, а апостериорная информация является информацией только о *нашем* мире. Она может содействовать локализации нашего мира в пространстве возможных миров, но трудно понять, как она могла бы давать информацию о протяженности данного пространства. Любые тезисы о дополнительных ограничениях метафизической возможности казались бы произвольными допущениями; с тем же основанием можно было бы допустить метафизическую невозможность того, что выпущенный из рук камень мог бы двигаться не вниз, а вверх.

Данная позиция, кроме того, ведет к *ad hoc* умножению модальностей. Если принять ее, то мы должны были бы признавать *четыре* вида возможности и необходимости утверждений, даже если оставить в стороне естественную модальность: возможность и необходимость в соответствии с первичными и вторичными интенсионалами, в про-

странстве логически возможных или метафизически возможных миров. А если говорить о возможности не утверждений, а миров, то у нас оказалось бы *три* объективных класса возможных миров: логически возможные миры, метафизически возможные миры и естественно возможные миры. У нас есть серьезные основания верить в существование первого и последнего из этих классов, но очень мало оснований допускать третий, особый метафизический класс.

Тот, кто считает зомбийный мир логически возможным, но метафизически невозможным, должен ответить на ключевой вопрос: *почему Бог не мог бы создать зомбийный мир?* Естественно предположить, что при создании мира Бог в силах сделать всё, что является логически возможным. Но защитник метафизической необходимости должен говорить либо то, что, хотя такая возможность существует, Бог не мог бы создать его, либо то, что, хотя Бог мог бы создать его, он тем не менее метафизически невозможен. Первое положение совершенно необоснованно, а второе абсолютно произвольно. И если говорить о втором положении, то аргумент против материализма в любом случае сохранит свою силу; после фиксации физических фактов о мире Богу надо будет сделать и еще кое-что.

Но даже если принять это воззрение, во многих существенных отношениях оно будет очень похоже на тот дуализм свойств, который я отстаиваю. Согласно этому воззрению, дело по-прежнему будет обстоять так, что существование сознания нельзя будет вывести из физического знания, так что сознание не будет допускать редуктивного объяснения. И, как и прежде, мы будем нуждаться в изначальных связующих принципах для объяснения супервентности феноменального на физическом. Единственное различие между этими воззрениями состоит в том, что соответствующие психофизические принципы окажутся фундаментальными «законами необходимости», а не законами природы. В плане *любых* объяснительных моментов при построении теории мы остаемся там же, где и дуализм свойств; различие, по сути, сводится к онтологическим предпосылкам.

Представляется, что единственной реальной мотивацией в пользу данного взгляда может быть спасение материализма любой ценой — возможно, исходя из представлений о проблематичности дуализма. Но такой материализм кажется гораздо более загадочным, чем его дуалистическая альтернатива. Привлечение базовых «метафизически необходимых» принципов, ограничивающих пространство возможных миров, вводит гораздо более проблематичный, а на деле гораздо менее натуралистичный элемент, чем простое допущение добавочных естественных законов, постулируемых дуализмом свойств. В конечном счете задействование новой степени необходимости — это такое

ad hoc допущение, которое приводит не только к решению проблем, но и к появлению неменьшего числа новых проблем. Эта позиция спасает материализм ценой объявления совершенной тайной вопроса о том, как сознание могло бы быть физическим⁸.

Это последняя возможность для материалиста, допускающего представимость зомбийного мира, но желающего спасти материализм. Обсуждавшаяся выше позиция предполагает, что материалист признает полную когерентность понятия зомби, даже для существа с максимально развитыми рациональными способностями, но тем не менее отрицает его метафизическую возможность, приходя, таким образом, к «двуслойной картине» логически и метафизически возможных миров. Однако материалист может также попытаться показать, что эта кажущаяся представимость обязана своим существованием некой ущербности нашего разума, так что если бы мы были более продвинутыми в интеллектуальном плане, мы увидели бы, что такое описание мира в конечном счете не является когерентным. Согласно этой позиции, подобный мир не оказывается даже *логически* возможным; и мы ошибочно верим в его логическую возможность просто из-за ограниченности человеческих когнитивных способностей. (Так можно интерпретировать взгляды Макгина (McGinn 1989).)

Когнитивные ограничения

Можно было бы попробовать подкрепить эту позицию аналогией с необходимостью ряда сложных математических истин, выходящих за пределы возможностей нашего математического постижения. Если наши математические способности алгоритмизируемы, подобные истины должны существовать (по теореме Геделя), но даже если нет, то они тоже вполне могут существовать. (Примером могла бы быть гипотеза Гольдбаха или, скажем, континuum-гипотеза или ее отрижение.) Эти истины необходимы, даже если не могут быть познаны нами априори, и они не основаны на сочетании априори познаваемых и эмпирических факторов, подобно крипкеанским необходимостям. Не может ли импликация от физических к феноменальным фактам оказаться подобной необходимостью, неким образом выходящей за пределы наших способностей модального постижения⁹?

Эта аналогия, однако, несовершенна. Когда речь идет о математике, наши модальные рассуждения оставляют вопрос *открытым*; наши интуиции представимости ничего не говорят нам об обоих вариантах. Хотя в слабом смысле и можно допускать «представимость» ложных положений — имея в виду, что они, насколько это нам известно, ложны, — но это не тот смысл, который позволяет представлять мир, где они неверны. В случае же с зомби вопрос не остается открытym: кажется, что мы ясно можем представить мир, в котором

данная импликация не проходит. Для спасения материализма возможность этого мира должна быть исключена — несмотря на то, что о ней надежно свидетельствуют наши модальные способности; но в случае с математикой отсутствует даже намек на возможность примера подобного исключения кажущегося возможным мира. И любой провал между представимостью и возможностью, на который мог бы ссылаться здесь материалист, опять-таки должен был бы быть провалом *sui generis*, не подкрепленным какой-либо аналогией в других областях¹⁰.

Конечно, материалист мог бы решиться на этот шаг и заявить о *sui generis* когнитивной ущербности. Для этого он должен был бы признать, что все аргументы в главе 3 ложны по непостижимым для нас причинам. Не говоря уже о том, что такая позиция рисует картину колоссальных заблуждений наших интуиций представимости, ведомых нашим несовершенным разумом, она также предполагает, что более умная Мэри на основе физической информации действительно могла бы узнать, каково это — видеть красное, и что можно осуществить такой анализ (к примеру, структурный или функциональный) феноменальных понятий, который подкреплял бы импликацию от физических к феноменальным фактам, хотя оценка правильности этого анализа выходила бы за пределы наших возможностей.

И хотя нельзя отрицать, что любой философский аргумент *мог бы* оказаться неверным вследствие когнитивной ущербности, при отсутствии серьезных оснований считать так подобное возражение выглядит совершенно искусственным. Как и раньше, главным его мотивом кажется желание любой ценой сохранить материализм. Такой путь всегда должен рассматриваться как *последняя* возможность, к которой можно прибегать лишь после признания неспособности ответить на серьезные аргументы, указывающие на наши ошибки, а также бесперспективности разработки альтернативы материализму. Если нам удастся найти серьезную и удовлетворительную альтернативу, мотивы занимать такую позицию исчезнут¹¹.

3. Другие аргументы в пользу дуализма*

Я не первый, кто использовал аргумент от логической возможности против материализма¹². Более того, я считаю, что в том или ином виде это базовый антиматериалистический аргумент в философии сознания. Тем не менее он не был объектом того пристального внимания, которого он заслуживает. Больше внимания уделялось двум антиматериалистическим аргументам Джексона (Jackson 1982) и Кripке (Kripke 1972). Они кажутся мне вполне уместными, хотя, быть может, и не столь фундаментальными. Аргумент Джексона важен

как прелюдия к аргументу от логической супервентности, а наиболее убедительная часть аргумента Кripке, как мы увидим, зависит от аргумента от логической супервентности.

Я уже обсуждал аргумент Джексона — аргумент знания — в контексте демонстрации отсутствия логической супервентности, где он играет вспомогательную роль. Вспомним, что речь там идет о Мэри, которая занимается нейронаукой, выросла в черно-белой комнате и знает все физические факты о процессах в мозге при восприятии цвета. Позже, когда она впервые видит красный объект, она узнает какие-то новые факты. В частности, она узнает, каково это — видеть красное. Аргумент завершается выводом, что физические факты не исчерпывают все факты и что материализм ложен.

Аргумент Джексона

Этот аргумент тесно связан с аргументами от зомби или инвертированных спектров, поскольку все они врачаются вокруг невыводимости феноменальных фактов из физических фактов. В известном смысле это разные грани одного и того же аргумента. Однако аргумент Джексона нередко считается уязвимым в качестве прямого аргумента против материализма вследствие использования в нем интенсионального понятия знания. Многие критические выпады против этого аргумента фокусировались на этой интенсиональности; доказывалось, к примеру, что один и тот же факт может быть познан двумя различными способами. Эти выпады, на мой взгляд, не приносят результата, но проще всего убедиться в этом можно, непосредственно обратившись к отсутствию супервентности, выражаемому не в эпистемологических, а в метафизических терминах. Разработанная мной модель помогает понять, по какой именно причине разные возражения не приводят к цели. Ниже я рассмотрю некоторые из таких возражений.

Прежде всего, различные критики доказывали, что, хотя Мэри и получает новое знание, когда видит что-то красное, это знание не соответствует какому бы то ни было новому *факту*. Она просто по-новому, сообразно новому «модусу презентации», узнает старый факт благодаря интенсиональности знания (Churchland 1985; Horgan 1984b; Lycan 1995; McMullen 1985; Papineau 1993; Teller 1992; Tye 1986). Тай и Лайкан, к примеру, апеллируют к интенсиональному различию между утверждениями «этая жидкость — вода» и «этая жидкость — H_2O »: в известном смысле они выражают один и тот же факт, но одно может быть известно без другого. В сходном ключе и Черчленд указывает на провал между знанием температуры и знанием средней кинетической энергии. Хорган обсуждает различие между знанием о Кларке Кенте и знанием о Супермене, а Макмаллен указывает в этой связи на Марка Твена и Сэмюэла Клеменса.

Эти провалы возникают именно вследствие различия первичных и вторичных интенсионалов. Можно знать что-то о воде, не зная каких-то вещей об H_2O , так как их первичные интенсионалы различны — не существует априорной связи между мыслями о воде и мыслями об H_2O . Тем не менее в известном смысле имеется только один набор фактов о них: благодаря апостериорному тождеству воды и H_2O их вторичные интенсионалы совпадают. (Не очевидно, что мы должны индивидуализировать факты так, чтобы факты о воде и факты об H_2O оказывались одними и теми же фактами, но в целях аргументации я принимаю этот тезис¹³.) Если использовать принятую ранее терминологию, утверждение «если это вода, то это H_2O » является логически контингентным, но метафизически необходимым. Данное возражение поэтому ничем не отличается от обсуждавшегося выше возражения, связанного с различием логической и (крипкеанской) метафизической необходимости, и для убедительного ответа на него достаточно сослаться на обсуждение там первичных и вторичных интенсионалов.

Мы также можем выразить суть дела более прямым способом. Всякий раз, когда кто-то знает некий факт сообразно одному модусу презентации, но не сообразно другому, здесь всегда будет иметься какой-то *иной* факт, относительно которого этот человек будет лишен знания, — факт, связующий два упомянутых модуса презентации¹⁴. Если кто-то знает, что Геспер доступен для наблюдения, но не знает, что Фосфор доступен для наблюдения (поскольку он не знает, что Геспер и есть Фосфор), то такой человек не знает, что один и тот же объект является одновременно ярчайшей звездой на утреннем небе и ярчайшей звездой на вечернем небе. Это особый факт, который абсолютно неизвестен данному человеку. Аналогичным образом если кто-то знает, что Супермен может летать, но не знает, что Кларк Кент может летать, то он не знает о существовании индивида, одновременно являющегося ведущим репортером *Daily Planet* и носящего накидку. Если кто-то знает, что вода мокрая, но не знает, что H_2O мокрая, он не знает, что вещество в озерах состоит из молекул H_2O , и т. д.

Более формально: допустим, что «*a* есть *G*» и «*b* есть *G*» — один и тот же факт в этом смысле, но некто не может априори соединить эти факты. Это должно объясняться тем, что $a = b$, и одинаковостью вторичных интенсионалов при различии первичных интенсионалов: возможно, *a* эквивалентно *dthat(P)*, а *b* эквивалентно *dthat(Q)*. Если кто-то знает, что *a* есть *G*, но не знает, что *b* есть *G*, то он лишен фактуального знания о том, что нечто является одновременно *P* и *Q*. В более общем виде: он лишен фактуального знания, что нечто является как *P'*, так и *Q'*, под которыми понимаются любые идентифицирующие дескрипции такого рода, когда он знает, что *a* есть *P'* и что *b* есть *Q'*.

Этот факт совершенно отличен от фактов, изначально находившихся в его распоряжении. Даже при трактовке с позиции вторичных интенсионалов можно говорить о возможном мире, в котором *a* есть *F*, но ничто не является как *P*, так и *Q* (или как *P'*, так и *Q'*).

(Как и в параграфе 2, здесь возникает осложнение, связанное с тем, что *P* и *Q* могут быть индексикальными свойствами, однако это обстоятельство не приводит ни к каким существенным изменениям. Чтобы сделать неизвестный новый факт строго неиндексикальным, нужно лишь перейти к факту «существует точка [со свойством *X*], из которой *P* и *Q* указывают на одно и то же». *X* — подстраховка на тот случай, если некто знает какую-то *другую* позицию, в которой *P* и *Q* указывают на одно и то же; в подобной ситуации мы просто специфицируем *X* до такой степени, чтобы ту точку можно было отличить от других местонахождений. Крайний случай, в котором человек не обладает никаким различающим самопознанием, сводится к чисто индексикальной ситуации, обсуждаемой ниже.)

Из этого следует, что, если Мэри получает какое-либо фактуальное знание, отсутствовавшее у нее прежде, — даже если речь идет исключительно о знании старого факта сообразно новому модусу презентации, — должен существовать подлинно новый узнаваемый ею факт. В частности, она должна узнавать новый факт, *касающийся* того модуса презентации. Если она уже знала все физические факты, то материализм ложен. Физические факты никоим образом не исчерпывают все существующие факты.

Этот ответ может показаться менее четким, чем соответствующий ответ на аргумент от логической возможности. Супервентностная модель элиминирует не столь ясный вопрос о том, как индивидуализировать компоненты знания, и поэтому делает обсуждение менее запутанным. Тем не менее тщательный анализ показывает, что аналогии с водой— H_2O и связанные с ними возражения в любом случае проваливаются по одним и тем же причинам. Хотя это, похоже, самое популярное возражение на аргумент знания, представляется вместе с тем, что оно самое слабое из главных возражений. Оно попросту не выдерживает проверки на прочность.

Второе, более утонченное возражение, восходящее к Лору (Loar 1990), также предполагает, что Мэри получает новое знание о старых фактах благодаря интенсиональности, но не останавливается на привычных аналогиях с водой— H_2O . Лор признает, что аналогии с обычными примерами не могут выполнить задачу, поставленную материалистом, поскольку (в нашей терминологии) такие аналогии допускают, что физические и феноменальные понятия имеют разные первичные интенсионалы, и антиматериалист может просто применить данный аргумент

к свойству, соответствующему первичному интенсионалу. Как выражается Лор, даже если «тепло» и какие-то статистико-механические предикаты *обозначают* одно и то же свойство (вторичный интенсионал), они тем не менее *вводят* разные свойства (первичный интенсионал). Поэтому он продолжает аргумент и доказывает, что два предиката могут вводить одно и то же свойство — то есть иметь один первичный интенсионал — даже тогда, когда эта одинаковость не может быть познана априори. Если это так, то знание Мэри о феноменальных свойствах может быть знанием о физических/функциональных свойствах, даже если она заранее не могла бы связать их.

Но как могут совпадать первичные интенсионалы, если мы не в состоянии узнать об этом априори? Только в том случае, если пространство возможных миров меньше, чем можно было бы априори подумать. Мы считаем, что эти интенсионалы различны, так как мы представляем мир, в котором у них имеются разные референты, как в зомбийном мире. Поэтому Лор должен исходить из того, что подобный мир не является реально возможным, хотя мы не можем исключить его возможность по концептуальным соображениям и несмотря на то, что мы не можем ссылаться здесь на крипкеанскую апостериорную необходимость. Таким образом, эта позиция ничем не отличается от рассмотренного выше возражения, связанного с «сильной метафизической необходимостью». Как и в том возражении, позиция Лора подразумевает метафизическую необходимость условного заключения от физических фактов к феноменальным фактам, невзирая на его логическую контингентность, причем этот провал не может быть объяснен различием первичных интенсионалов. Подобно тому возражению, позиция Лора требует изначального и произвольного ограничения возможных миров. Лор не выдвигает никаких аргументов в пользу этого ограничения, и его можно подвергнуть точно такой же критике¹⁵.

Можно было бы ожидать, что здесь можно будет выдвинуть и третье возражение, аналогичное промежуточной «альтернативной стратегии» из второго параграфа. Оно выливалось бы в тезис, что Мэри в действительности не знает всех физических фактов. Она знает все факты, выраженные в терминах физики, но лишена знания о скрытых (феноменальных илиprotoфеноменальных) сущностях физических объектов. Если бы она обладала этим знанием, она в силу этого знала бы и феноменальные факты. Однако это воззрение опять-таки лишь с очень большой натяжкой может претендовать на то, чтобы именоваться «материализмом». Подобно моей собственной концепции, данная позиция должна признавать феноменальные или protoфеноменальные свойства фундаментальными свойствами.

Четвертое возражение увязывает состояние Мэри с отсутствием индексикального знания (Bigelow and Pargetter 1990; McMullen 1985;

Papineau 1993). Здесь доказывается, что, хотя Мэри получает новое знание, ее случай не более загадочен, чем другие случаи, когда кто-то, знающий все релевантные объективные факты, открывает что-то новое, как, к примеру, всезнающий человек с амнезией, узnaющий, что «я — Рудольф Лингенс», или прекрасно информированный человек, страдающий бессонницей, не знающий, что *сейчас* три часа сорок девять минут ночи (см. (Reppu 1979) и (Lewis 1979)). В этих случаях имеется провал между физическим знанием и индексикальным знанием, точно так же как в случае Мэри имеется провал между физическим знанием и феноменальным знанием.

Эта связь может быть прописана двумя способами. Во-первых, критик мог бы попытаться *редуцировать* феноменальное знание к индексикальному знанию, доказывая, что Мэри лишена лишь индексикального знания. Во-вторых, он мог бы попробовать провести *аналогию* между двумя этими случаями, доказывая, что в индексикальном случае эпистемический провал не приводит к онтологическому провалу (индексикальность не фальсифицирует материализм), так что и тогда, когда речь идет о феноменальном, мы не обязаны допускать наличие онтологического провала.

Редуктивная стратегия терпит очевидный крах. Как мы видели в главе 2, подобно фактам о сознательном опыте, индексикалы не являются логически супервентными на физических фактах, но эта проблема полностью устраняется добавлением небольшого «индексикального факта» о местонахождении агента, о котором идет речь. Но даже если мы наделим Мэри полным знанием о ее индексикальном отношении ко всем вещам физического мира, ее знание о переживаниях красного ничуть не возрастет. Испытывая нехватку феноменального знания, она оказывается лишенной чего-то гораздо большего, чем тот, кому недостает индексикального знания.

Стратегия, связанная с применением указанной выше аналогии, представляется более интересной. Можно было бы попытаться ответить на нее, доказывая, что онтологический провал имеется даже и в индексикальном случае (см., напр., Nagel 1983), но можно дать и более прямой ответ. Чтобы убедиться в этом, отметим, что в индексикальном случае нельзя даже приступить к развертыванию аргументации, аналогичной той, которая была использована в первом параграфе, так как нет такого представимого нецентрированного мира, в котором физические факты не отличаются от наших, но индексикальные факты разнятся с фактами нашего мира. В нецентрированных мирах вообще нет места для индексикальных фактов. Разумеется, можно представить релевантный *центрированный* мир, но в интересующем нас онтологическом плане значимы именно нецентрированные миры. (Если нет,

то в индексикальном случае тоже имеется онтологический провал, и критик не сможет даже приступить к своей аргументации¹⁶.) Так что лишь в этом случае мы можем избавиться от эпистемического провала, отмечая, что эпистемические связи определяются *центрированными* первичными интенсионалами, тогда как онтологические связи определяются свойствами, соответствующими нецентрированным интенсионалам. И это отражается в единственной дыре, которая была найдена в аргументе второго параграфа и в аналогичном аргументе этого параграфа, — в том факте, что первичные интенсионалы определяют только соотнесенные с центром свойства. Эта дыра пропускает единственный фрагмент нередуцируемого индексикального знания (локализация центрированного мира) без онтологической мэды, но ничего больше. После конкретизации местонахождения центра, эта дыра закрывается. Феноменальные факты остаются неустроенными даже после конкретизации местонахождения центра, так что сознательный опыт остается снаружи¹⁷.

Если материалист хочет упорствовать в своем материализме, ему придется отрицать, что Мэри вообще *открывает* что-то новое в мире. Материализм нуждается в логической супервентности, требующей, чтобы Мэри не могла получать какого-либо нового фактического знания при ее первом переживании опыта красного. Поэтому, следуя пятой стратегии, Льюис (Lewis 1990) и Немиров (Nemirov 1990) доказывают, что в лучшем случае Мэри обретает новое *умение*. К примеру, она обретает умение воображать то, как выглядят красные вещи, и узнавать их, когда она видит их. Но это лишь знание *как*, а не знание *что*. Впервые переживая опыт красного, Мэри не узнает каких-либо фактов о мире¹⁸.

В отличие от предыдущих вариантов, эта стратегия лишена каких-либо внутренних проблем. Ее главная проблема состоит в ее крайней неправдоподобности. Нельзя сомневаться, что Мэри действительно обретает какие-то умения, когда она впервые переживает опыт красного, так же как она обретает какие-то умения, когда она учится кататься на велосипеде. Но трудно отрицать, что она узнает и что-то еще: некие *факты* о природе опыта. Исходя из ее предыдущего знания, переживание красного могло бы быть таким, или иным, или вообще никаким. Но теперь она знает, что оно именно *такое*. Она сузила пространство эпистемических возможностей. Подобного знания не возникает, когда всезнающий механик учится кататься на велосипеде (возможно, за исключением знания феноменологии прогулок на велосипеде). Так что этот ответ не ухватывает того, что происходит, когда Мэри узнает, каково это — видеть красное.

В том, что открытие, сделанное Мэри, предполагает фактуальное знание, можно убедиться и более косвенными методами. Так, Лор

(Loar 1990) указывает, что подобное знание может быть встроено в кондиционалы: «Если красные вещи видятся так, синие — так, то пурпурные вещи, скорее всего, будут видеться вот так»; «если собаки видят красное так-то, то из этого следует то-то и то-то» и т. д. Другой пример: как указывает Лайкан (Lycan 1995)¹⁹, воображаемое нами может оказаться правильным или неправильным; так, после того как Мэри увидела какие-то цвета, она могла бы вообразить, каково это — видеть какой-то другой цвет, и ее фантазия могла бы оказаться правильной или неправильной. Если это так, то знание о том, каково что-то, есть знание какой-то истины о мире, и анализ, связанный с отсылкой к умениям, терпит крах.

Деннетт (Dennett 1991) занимает близкую, но еще более жесткую позицию, доказывая, что Мэри вообще ничего не узнает. Он замечает, что Мэри могла бы использовать свое нейрофизиологическое знание для распознания того, что красный объект, который она видит, красен, обращая внимание на то, как он воздействует на ее реакции, так как это воздействие может отличаться от того, которое оказал бы на нее синий объект. (Если группа экспериментаторов пытается обмануть ее, показывая ей синее яблоко, она могла бы и не поддаться на обман.) Не исключено, что это так, но из этого следует лишь то, что, в противоположность тому, что говорят Льюис с Немировым, Мэри обладала бы определенными умениями распознавания еще до того, как она впервые пережила опыт красного. Это никоим образом не показывает, что она обладала бы ключевым знанием — знанием о том, каково это — видеть красное. Такой вывод можно было бы сделать лишь при условии предварительного признания корректности анализа «знания о том, каково это» с помощью понятия умения; но, если бы мы приняли этот анализ, аргумент против материализма был бы уже опровергнут. Поэтому аргумент Деннетта лишь отвлекает здесь внимание.

В конечном счете стратегия, которой *должен* следовать материалист, состоит в том, чтобы отрицать обретение Мэри знания о мире. И кажется, что сделать это можно лишь через анализ «знания о том, каково это» через понятие умения. Это единственная внутренне когерентная позиция, которая не может быть отвергнута по техническим соображениям, так же как единственным когерентным способом, посредством которого материалист может сопротивляться аргументу от логической супервентности, в конечном счете является обращение к аналитическому функционализму. Однако сама неправдоподобность отрицания того, что Мэри обретает знание о мире, свидетельствует об обреченности материализма²⁰.

Мы видели, что модальный аргумент (аргумент от логической возможности) и аргумент знания — две стороны одной и той же медали. Думаю, что каждый из них в принципе достигает успеха и сам

по себе. На практике, однако, они лучше всего работают в связке²¹. Если взять один лишь аргумент знания, то, хотя большинство материалистов не видит оснований отрицать, что Мэри обретает знание о мире, они зачастую отрицают, что из этого следует признание ложности материализма. Если взять один лишь модальный аргумент, то, хотя большинство материалистов не видят оснований отрицать правильность аргумента от представимости зомби или инвертированных спектров к признанию ложности материализма, они зачастую отрицают посылку этого аргумента. Но если соединить эти аргументы, то оказывается, что модальный аргумент поддерживает аргумент знания именно там, где он нуждается в поддержке, и наоборот. Быть может, наиболее эффективной оказывается такая комбинация этих двух аргументов, при которой мы сможем использовать аргумент знания для убедительной демонстрации отсутствия логической супервентности, а модальный аргумент — для убедительного перехода от констатации этого отсутствия к выводу о ложности материализма.

Аргумент Крипке

Аргумент Крипке был нацелен против тезиса тождества, выдвинутого Плейсом (Place 1957) и Смартом (Smart 1959), но его можно рассматривать и в более широком плане, как довод против всех форм материализма. Я довольно подробно обсуджу достоинства и недостатки этого аргумента — для подкрепления вывода, что удачными являются именно те его компоненты, которые соответствуют аргументу от логической супервентности.

Аргумент выглядит примерно так: согласно тезису тождества, определенные ментальные состояния (такие как болевые ощущения) тождественны состояниям мозга (к примеру, активации С-волокон), несмотря на то что слова «боль» и «активация С-волокон» имеют разное значение. Изначально предполагалось, что это тождество скорее контингентно, чем необходимо, так же как контингентно тождество воды и H_2O . Крипке, однако, доказывает, что все тождества являются необходимыми: если X есть Y , то X с необходимостью есть Y , если термины X и Y — жесткие десигнаторы, указывающие на одного и того же индивида или вид в разных мирах. Вода, доказывает он, с необходимостью есть H_2O , то есть вода есть H_2O в любом возможном мире. Это тождество может казаться контингентным — то есть могло бы показаться, что в каком-то из возможных миров вода есть не H_2O , а XYZ , — но это иллюзия. В действительности тот возможный мир, который мы воображаем, вообще не содержит воды. В нем есть всего лишь водянистая материя — материя, выглядящая как вода и проявляющая себя как вода, — состоящая из XYZ . Утверждая, что эта водянистая материя является водой, мы даем ей неверное описание.

И Крипке утверждает, что если болевые ощущения тождественны активации С-волокон, то это тождество должно быть необходимым. Однако оно *не кажется* необходимым. На первый взгляд, можно вообразить возможный мир, в котором имеется боль, но вообще отсутствуют состояния мозга (развоплощенная боль), и можно вообразить мир, активация С-волокон в котором не сопровождается болью (к примеру, у зомби). Далее он доказывает, что эти возможности нельзя трактовать как всего лишь кажущиеся возможности, подобно возможности воды без H_2O . Ведь в таком случае мы *неверно описывали бы* мир с «развоплощенной болью» в качестве мира, в котором имеется боль, тогда как в действительности в нем была бы лишь «болевая материя» (нечто ощущаемое как боль). Аналогичным образом мы неверно описывали бы лишенность боли зомби, поскольку в действительности он был бы лишен лишь болевой материи. Согласно этому объяснению, зомбийный мир, похоже, содержал бы реальную боль, то есть активацию С-волокон; она лишь не ощущалась бы таковой.

Согласно Крипке, однако, дело не может обстоять таким образом: чтобы нечто было болью, нужно *лишь*, чтобы оно ощущалось как боль. Не существует различия между болью и болевой материей, подобного различию между водой и водянистой материей. Может быть так, что нечто ощущается как вода, но не является водой, но не может быть так, что нечто ощущается как боль, но не является болью. Ощущение боли *существенно для* нее. Поэтому и нельзя по аналогии с прежними случаями отбрасывать возможность боли без состояний мозга (и наоборот). Те возможные миры действительно возможны, и ментальные состояния не являются с необходимостью тождественными состояниям мозга. А из этого следует, что они вообще не могут быть тождественны состояниям мозга.

Крипке разворачивает свой аргумент двумя различными способами, направляя его как против теорий конкретного тождества, так и против теорий типового тождества. Теории конкретного тождества утверждают, что *конкретные боли* (такие как моя боль в данный момент) тождественны конкретным состояниям мозга (таким как активация С-волокон в моей голове в данный момент). В указанной выше манере Крипке пытается показать, что конкретная боль могла бы иметь место без конкретного связанного с ней состояния мозга, и наоборот, и что поэтому они не могут быть тождественными. Теории типового тождества утверждают, что ментальные состояния и состояния мозга тождественны как *типы*. Так, боль как тип могла бы быть тождественна активации С-волокон. Крипке полагает, что это непосредственно опровергается фактом возможной реализации типа «ментальное состояние» без типа «состояние мозга» и наоборот. В общем, здесь можно

насчитать четыре отдельных аргумента, разделенных сообразно мишени, на которую они направлены (конкретные или типовые теории тождества), и сообразно методу аргументации (от возможности развоплощения и от возможности зомби).

Между аргументом Кripке и моим аргументом имеется ряд очевидных различий. Начать с того, что аргумент Кripке сформулирован исключительно в терминах тождества, тогда как я опирался на понятие супервентности. Во-вторых, аргумент Кripке тесно увязан с его теоретическим аппаратом, включающим жесткие десигнаторы и апостериорную необходимость, тогда как в моем аргументе этот аппарат играет лишь второстепенную роль, применяясь при ответе на некоторые возражения. В-третьих, обычно считается, что аргумент Кripке опирается на определенного рода эссенциализм относительно различных состояний, тогда как мой аргумент не содержит каких-либо ссылок к этому учению. В-четвертых, мой аргумент нигде не апеллирует к возможности развоплощения, как у Кripке. Тем не менее между ними есть и явные сходства. Оба аргумента являются модальными аргументами, отводящими ключевую роль необходимости и возможности. И оба они апеллируют к логической возможности отрыва физических состояний от связанных с ними феноменальных состояний.

Посмотрим теперь, что правильно и что неправильно в аргументах Кripке, начав с его аргументов против конкретного тождества. Большинство считает их неубедительными. Связано это прежде всего с тем, что они опираются на интуиции о том, что считать *той же самой вещью* в разных возможных мирах, а такие интуиции, как известно, недежны. Тезис Кripке о том, что можно было бы иметь *то же самое* состояние боли без *того же самого* состояния мозга, опирается на тезис о существенности чувства боли, и только этого чувства, для того состояния боли. Но подобные утверждения о существенных свойствах индивидов трудно обосновать. Теоретик конкретного тождества может возразить, что не менее правдоподобной является и существенность для того состояния свойства активации С-волокон. Конечно, активация С-волокон не кажется существенной для боли как *типа*, но кто сказал, что она не является существенной для этой конкретной боли, особенно если она тождественна какому-то состоянию мозга? И если это так, то мы просто не могли бы испытывать эту конкретную боль без определенного состояния мозга. (Подобную линию аргументации проводит Feldman 1974), доказывающий, что болезненность не обязательно должна быть чем-то существенным для какой-то конкретной боли, а также Макгинн (McGinn 1977), в конечном счете доказывающий, что для конкретной боли могли бы оказаться существенными *как* болезненность, *так* и активация С-волокон.) Если это

так, то, воображая развоплощенную вариацию моей боли, мы воображаем не *ту самую* боль, а другую, нумерически отличную боль. Это же верно и для случая, когда мы воображаем активацию моих С-волокон без боли. Поэтому эти аргументы против конкретного тождества неубедительны, хотя аргументы против типового тождества и могут сохранять свое значение.

Кроме того, аргумент от развоплощения не может рассматриваться как убедительный довод против материализма. Он мог бы опровергать тезис о типовом тождестве вроде того, что был выдвинут Плейсом и Смартом, но материализм не нуждается в таком тезисе²². Как отмечает Байд (Boyd 1980), материализм может и не предполагать, что ментальные состояния являются физическими состояниями во всех возможных мирах, — с ним совместимо утверждение, что в каких-то мирах ментальные состояния сформированы из чего-то нефизического; главное, чтобы в *нашем* мире они имели физический состав. Возможность развоплощения показывает лишь возможность дуализма, но не его действительность²³. Для иллюстрации этой мысли заметим, что мало кто взялся бы утверждать, что возможность нефизической жизни влечет за собой биологический дуализм. Возможно, что Кripке интересовал лишь аргумент против тезиса тождества, но в любом случае материализм выживает в своем более общем виде.

Это оставляет нас с аргументом от возможности реализации физических состояний без соответствующих им ментальных состояний, то есть, по сути, с аргументом от возможности зомби. Любопытно, что эта часть аргумента Кripке привлекла наименьшее внимание критиков, и большинство комментаторов сосредоточило свое внимание на возможности развоплощения. Аргумент против тезиса о сильном типовом тождестве, который поставляют нам зомби, опять-таки может оказаться нерелевантным, поскольку материализм не нуждается в этом тезисе, но здесь неявно присутствует и более общий аргумент. Возможность реализации соответствующих физических состояний без боли, как доказывает Кripке (с. 153–154), демонстрирует то, что даже после того как Бог создал все физические процессы, протекающие, когда кто-то испытывает боль, — скажем, мозг с активированными С-волокнами, — *ему надо было сделать и кое-что еще*, чтобы эта активация чувствовалась как боль. И этого достаточно для того, чтобы можно было говорить о ложности материализма²⁴.

Данный аргумент от физических состояний без феноменальных состояний прямо соответствует выдвинутому мной аргументу против материализма. Соответствие простирается и на последующие маневры. На возражение, что эта ситуация всего лишь представима, но не реально возможна, Кripке ответит, что мы не можем отделаться от такой

представимой ситуации, как от ситуации, в которой некто лишен ощущения боли, но не самой боли, поскольку в любом возможном мире испытывать боль — значит иметь болевые ощущения (то есть вторичный интенсионал и первичный интенсионал «боли» совпадают). К этому мы могли бы добавить (вместе с Джексоном (Jackson 1980)), что, даже если оспорить эту эквивалентность, аргумент против материализма будет проходить в применении пусть и не к самой боли, но к *чувству* боли (то есть даже если эти интенсионалы не совпадают, аргумент все равно проходит при использовании первичного интенсионала). Эти ответы изоморфны ответам, которые были даны мной на подобные возражения, рассмотренные ранее в этой главе.

(Заметим, что, выдвигая тезис о том, что кажущаяся представимой, но невозможная ситуация должна быть охарактеризована как неверно описанная эпистемически возможная ситуация, Кripке, по сути, одобряет «слабую» трактовку апостериорной необходимости: пространства представимых и возможных миров совпадают, но апостериорные факторы накладывают ограничения на их корректное описание²⁵. Чтобы убедиться в этом, обратим внимание на то, что сторонник «сильной» метафизической необходимости, согласно которой пространство возможных миров есть собственное подмножество пространства представимых миров, не будет отстаивать подобный тезис. Согласно этому представлению, мы могли бы *корректно* описывать эпистемически возможную ситуацию, но при этом она все же могла бы оказываться (изначально) метафизически невозможной. Использование Кripке стратегии неверного описания, напротив, указывает на неявное одобрение им двумерной модели: все его примеры неверного описания и впрямь можно рассматривать как случаи, в которых мир описывается сообразно первичным, а не вторичным интенсионалам.)

Этот аргумент от физических состояний без феноменальных состояний кажется мне самой убедительной частью того анализа, который мы находим у Кripке. Она часто игнорируется в ходе дискуссий о тезисе тождества, развоплощении и т. п.; и сам Кripке выносит эту часть обсуждения на периферию. Я, однако, полагаю, что именно эта часть его анализа в конечном счете несет на себе основное бремя его аргументации.

В итоге мне кажется, что в той мере, в какой аргумент Кripке против материализма достигает цели, (1) возможность развоплощения неубедительна как аргумент против материализма, но это несущественно в данном случае; (2) аналогичным образом неубедительны, но несущественны аргументы, сформулированные в терминах тождества; (3) несущественна и эсценциалистская метафизика, за исключением того, что чувство боли является существенным для боли как

типа, — но это всего лишь факт, говорящий о том, что *означает* «боль»; и (4) крипкеанский аппарат жесткой десигнации и т. п. не играет главной роли, хотя и требуется для ответа на некоторые возражения²⁶. Но его аргумент имеет здоровое ядро, содержащее, по сути, аргумент от отсутствия логической супервентности.

4. Эпифеноменализм ли это?*

Проблема, связанная с той позицией, которую я отстаиваю, состоит в том, что если сознание всего лишь естественно супервентно на физическом, то кажется, что ему не должна быть присуща каузальная действенность. Физический мир более или менее каузально замкнут в том плане, что любое физическое событие представляется имеющим физическое объяснение (с незначительной поправкой на квантовую неопределенность). Из этого вытекает, что нефизическое сознание не может выполнять какой-либо независимой каузальной работы. Оно кажется простым эпифеноменом, идущим от движка физической каузальности, но не производящим каких-либо различий в физическом мире. Оно существует, но — в той мере, в какой идет речь о физическом мире, — могло бы и не существовать. Подобный взгляд отстаивал Гексли (Huxley 1874), но многие находят его континуитивным и непривлекательным. И впрямь этих следствий оказалось достаточно, чтобы некоторые авторы (к примеру, Кирк (Kirk 1979) и Сигер (Seager 1991)) усомнились в выводах своих аргументов против материализма и обратились к рассмотрению той возможности, при которой сознание все же было бы логически супервентным на физическом.

Этот аргумент был по-разному, но в чем-то и сходно формализован Кирком (Kirk 1979), Хорганом (Horgan 1987) и Сигером (Seager 1991). Если мы допускаем, что физический мир каузально замкнут и что сознание каузально вызывает какие-то физические события, то при определенных естественных допущениях относительно каузальности из этого следует, что сознание должно быть логически (или метафизически) супервентным на физическом²⁷. Если это так, то при условии каузальной замкнутости физического мира только лишь естественная супервентность сознания означает, что сознание является чем-то эпифеноменальным. Главные черты этого аргумента вполне очевидны: если можно убрать феноменальное из нашего мира и при этом сохранить каузально замкнутый мир Z , то всё, что происходит в Z , имеет каузальное объяснение, независимое от феноменального, так как в Z нет ничего феноменального. Но всё, что происходит в Z , происходит и в нашем мире, и поэтому каузальное объяснение, применимое

к Z , в равной степени применимо и к нашему миру. Поэтому феноменальное каузально нерелевантно. Даже если бы сознательный опыт отсутствовал, это не изменило бы каузальных механизмов поведения.

Отвечая на это, я применю двойственную стратегию. Прежде всего, неочевидно, что одна лишь естественная супервентность должна влечь за собой эпифеноменализм в самом сильном смысле слова. Нельзя отрицать, что связанная с ней картина *напоминает* эпифеноменализм. Тем не менее сама природа каузальности весьма таинственна, и нельзя исключить, что, когда она будет понята лучше, мы сможем уяснить тот неприметный путь, на котором сознательный опыт может быть каузально релевантным. (По сути, может оказаться так, что фоновые допущения в вышеприведенных аргументах ложны.) Ниже я продемонстрирую несколько возможных вариантов такого анализа. С другой стороны, я рассмотрю *основания*, по которым эпифеноменализм мог бы быть сочен чем-то непривлекательным, и проанализирую, насколько они могут быть сильны в качестве аргументов. Если эти интуиции нельзя превратить в убедительные аргументы, то может оказаться, что та разновидность эпифеноменализма, которая вытекает из данной позиции, *всего лишь* континтуитивна, и что в конечном счете мы можем решиться на какую-то дозу эпифеноменализма.

Стратегии, нацеленные на избежание эпифеноменализма

Имеется немало путей, следуя которым можно попробовать сохранить тезис об отсутствии логической супервентности, но при этом избежать эпифеноменализма. Самый очевидный из них связан с отрицанием каузальной замкнутости физического и одобрением сильной разновидности интеракционистского дуализма, при котором ментальное заполняет провалы, имеющиеся в физических процессах. Думаю, что не стоит применять эту стратегию — по причинам, о которых вскоре пойдет речь. Есть, однако, и множество других, не столь очевидных возможностей, определяемых соответствующими представлениями о метафизике, и прежде всего о каузальности. Я рассмотрю четыре возможности такого рода.

- 1. Регулярностная каузальность.** Первая возможность состоит в том, чтобы принять радикальное юмовское представление о причинности, согласно которому для того, чтобы A было причиной B , требуется лишь единообразная регулярность между событиями типа A и B . Подобная позиция позволяла бы говорить о «каузальной» роли феноменального: простой факт того, что за болевыми ощущениями обычно следуют реакции уклонения, означал бы, что боль каузально вызывает реакции уклонения.

Близкий только что упомянутому неюмовский вариант отождествляет каузальную связь с любой *номической* (или закономерной) связью, даже если номическая регулярность есть нечто большее, чем единообразная регулярность. Идея естественной супервентности полностью совместима с существованием номической связи между опытом и поведением (к примеру, могла бы существовать закономерная связь между опытом и лежащим в его основе состоянием мозга, а также закономерная связь между этим состоянием мозга и поведением). И можно было бы сказать, что этого достаточно для причинения. Данный тезис можно было бы попробовать поддержать ссылкой на то, что контрфактическое высказывание «поведение оставалось бы тем же самым даже при отсутствии опыта» при наиболее естественной его интерпретации является *ложным*: если бы опыт отсутствовал, состояние мозга было бы иным, и поведение тоже было бы иным. Здесь контрфактическое высказывание оценивается путем рассмотрения *естественно возможных миров*, а не логически возможных миров.

Я нахожу обе эти позиции неправдоподобными. Я выдвигал аргументы против юмовского взгляда на каузальность в главе 2, и даже если говорить о неюмовской позиции, то все равно малоправдоподобно, что для причинения достаточно любой номической связи, — возьмем, к примеру, корреляцию, имеющуюся между цветом волос однояйцевых близнецов. Тем не менее подобные соображения по крайней мере дают нам представление о том, почему сознание *кажется* играющим каузальную роль. Между сознательным опытом и последующими физическими событиями имеются систематические регулярности самого разного рода, каждая из которых подталкивает нас к тому, чтобы *заключить* о каузальной связи. Принимая во внимание эти регулярности, мы могли бы ожидать, что люди будут извлекать из них выводы о каузальном отношении, опираясь на доводы, имеющие в широком смысле юмовский характер. Данное обстоятельство может, таким образом, позволить нам отделаться от некоторых наших интуиций относительно каузальной действенности сознания, что помогает в реализации второго аспекта избранной мной стратегии.

2. Каузальная сверхдетерминация. Быть может, мы могли бы сказать, что как физическое, так и феноменальное состояние, хотя они и совершенно различны, могли бы считаться причиняющими более позднее физическое состояние. Если физическое состояние P_1 связано с феноменальным состоянием Q_1 , то, возможно, истинно как то, что P_1 причиняет более позднее физическое состояние P_2 , так и то, что Q_1 причиняет P_2 . Это контринтуитивно: P_1 уже есть достаточная причина P_2 , и поэтому Q_1 казалось бы каузально избыточным. Тем не менее не очевидно, что Q_1 не могло бы находиться в каузальном отношении

к P_2 . Особенно правдоподобным это могло бы выглядеть при принятии нередуктивного взгляда на причинность (в духе концепции Тули (Tooley 1987)). Не исключено, что существует нередуцируемая каузальная связь между двумя физическими состояниями и отдельная от нее нередуцируемая каузальная связь между феноменальным состоянием и соответствующим физическим состоянием.

На такого рода каузальную сверхдетерминацию событий нередко смотрят с подозрением, хотя трудно убедительно показать, что здесь что-то не так. Природа причинности понимается до сих пор настолько плохо, что мы еще не можем исключить сверхдетерминацию. Сам я не буду развивать эту идею, но она тем не менее заслуживает серьезного внимания.

3. Несупервентность каузальности. Третья стратегия опирается на саму природу причинности. Мы видели в главе 2, что имеется два класса фактов, не являющихся логически супервентными на конкретных физических фактах: факты о сознании и факты о каузальности. Естественно предположить, что две эти недостатчи могли бы быть тесно увязаны и что между сознанием и причинностью имеется глубокая метафизическая связь. В конце концов, и то и другое весьма таинственно, и было бы лучше, если бы две эти тайны можно было упаковать в одну. Не исключено, к примеру, что сам опыт является чем-то вроде каузальной связи; возможно, он каким-то образом реализует юмовское «неизвестное каузальное отношение»; или, быть может, это отношение имеет более сложный характер. Подобное отношение могло бы указывать на то, что опыт играет более тонкую роль в причинности, чем ее обычная разновидность, — позволяющую тем не менее избежать самой сильной формы эпифеноменализма.

Концепция такого рода была разработана Розенбергом (Rosenberg 1996), доказывающим, что многие проблемы сознания в точности параллельны проблемам, возникающим в связи с каузальностью. Он доказывает, что вследствие таких параллелей дело может обстоять так, что в актуальном мире опыт *реализует* каузальность или какие-то из ее аспектов. Согласно этому представлению, каузальность нуждается в реализации чем-то для поддержания множества ее свойств, и опыт является естественным кандидатом на эту роль. Если так, то каузальные отношения могут существовать именно благодаря существованию опыта, и поэтому опыт оказывается каузально релевантным в указанном утонченном смысле.

Разумеется, эта концепция исключительно спекулятивна и сталкивается с рядом проблем. Кажется, к примеру, что она ведет к некоей разновидности панпсихизма — позиции, согласно которой все обладает сознанием — и которая многим кажется континтуитивной. Кроме

того, по-прежнему проблемой остается и зомбийный мир — кажется, что мы можем вообразить протекание всех этих каузальных процессов и без опыта, так что опыт по-прежнему мог бы выглядеть чем-то эпифеноменальным. В ответ можно было бы сказать, что каузальность должна быть чем-то реализована; в зомбийном мире она реализована чем-то другим, но опыт все же релевантен для нашего мира благодаря тому, что он реализует в нем каузальность. Для меня неочевидно, что каузальность *должна* быть реализована чем-то, имеющим какие-то дополнительные свойства; и если это так, то феноменальная природа каузальности все же могла бы оказываться избыточной. Но опять-таки метафизика каузальности еще очень неясна, и данная концепция, несомненно, заслуживает тщательного рассмотрения.

4. Внутренняя природа физического. Стратегия, которая кажется мне наиболее привлекательной, опирается на наблюдение о том, что физическая теория характеризует свои базовые сущности лишь *относительно*, в терминах их каузальных и иных отношений к другим сущностям. Фундаментальные частицы, к примеру, по большей части характеризуются в терминах их предрасположенности к взаимодействию с другими частицами. При этом, разумеется, конкретизируется их масса и заряд, но конкретизация массы в конечном счете сводится к предрасположенности получать определенное ускорение вследствие воздействия сил и т. д. Каждая сущность характеризуется через ее отношение к другим сущностям, а эти сущности характеризуются их отношениями к другим сущностям, и так далее до бесконечности (за исключением, быть может, тех сущностей, которые характеризуются их отношением к наблюдателю). Получающаяся в итоге картина физического мира — это картина громадного каузального потока, но она ничего не говорит нам о том, что *соотносится* этой причинностью. Отсылка к протону является отсылкой к вещи, каузально вызывающей определенного рода взаимодействия, определенным образом сочетающейся с другими сущностями и т. д.; но какая именно вещь вызывает это причинение и вступает в те или иные сочетания? Как отмечает Рассел (Russell 1927), в этом вопросе физическая теория хранит молчание²⁸.

Кого-то мог бы привлечь образ мира как чистого каузального потока без каких-либо дополнительных соотносимых причинностью свойств, но это приводило бы к странной несубстанциалистской картине физического мира²⁹. Он заключал бы в себе только каузальные и номические отношения между пустыми носителями, лишенными собственных свойств. Интуитивно кажется более разумным предположить, что базовые сущности, соотносимые всей этой каузальностью, имеют какую-то свою внутреннюю природу, какие-то *внутренние* свойства, так что мир не лишен субстанции. Но физика может в лучшем

случае отсылать к этим свойствам через их внешние отношения; она ничего напрямую не говорит о том, чем могли бы быть эти свойства. У нас есть некие туманные интуиции об этих свойствах, основанные на нашем опыте их макроскопических аналогов, — интуиции о самой «массивности» массы, к примеру, — но эти интуиции непросто выразить и, по правде говоря, неясно, есть ли в них вообще что-то.

Имеется лишь один класс внутренних, нереляционных свойств, с которым мы непосредственно знакомы, и это класс феноменальных свойств. Естественно предположить, что неопределенные внутренние свойства физических сущностей и известные нам внутренние свойства опыта могут быть как-то соотнесены или даже перекрываться. Быть может, как предполагал Рассел, хотя бы некоторые из внутренних свойств физического сами являются разновидностью феноменального свойства.³⁰ Эта идея на первый взгляд кажется дикой, но лишь на первый взгляд. В конце концов, у нас и впрямь *нет никакого представления* о внутренних свойствах физического. Их место вакантно, и феноменальные свойства выглядят не менее достойным кандидатом на их роль, чем какие-то другие.

Здесь, конечно, возникает опасность панпсихизма. Я не уверен, что эта перспектива так уж плоха, — если феноменальные свойства фундаментальны, то естественно предположить, что они могли бы иметь широкое распространение, — но данное следствие не является необходимым. Альтернатива состоит в том, что соответствующие свойства протофеноменальны. Тогда реализация такого свойства сама по себе не влечет опыта, хотя реализация множества таких свойств могла бы в их совместном действии делать это. Трудно вообразить, как все это могло бы работать (мы знаем, что этого не может происходить, когда речь идет об обычных физических свойствах), но эти внутренние свойства весьма далеки от наших привычных представлений. Эту возможность нельзя исключать априори.

Так или иначе, но подобная глубинная связь указывает на некую каузальную роль, которую могло бы играть феноменальное. Если у физического есть внутренние свойства, то физическая причинность в конечном счете соотносит то, что реализует эти свойства. Если они феноменальны, то можно говорить о феноменальной каузальности; если они протофеноменальны, то феноменальные свойства наследуют каузальную релевантность вследствие своего супервентностного статуса, так же как билльярдные шары наследуют каузальную релевантность молекул. В любом случае феноменология опыта человеческих агентов может наследовать каузальную релевантность каузальной роли внутренних свойств физического.

Конечно, подобная каузальная релевантность не будет столь прямолинейной, как в обычных случаях. Как и прежде, можно будет,

к примеру, вообразить устранение феноменальных свойств при сохранении паттерна каузального потока. Но теперь на это можно ответить, что, воображая подобный сценарий, мы в действительности изменяем внутренние свойства физических сущностей и заменяем их чем-то другим (трудность, разумеется, в том, что мы вообще не привыкли воображать внутренние свойства физического). Таким образом, мы просто обращаемся к миру, в котором каузальная работа производится чем-то другим. Если бы мог существовать мир, представляющий собой *один лишь каузальный поток*, данный аргумент не проходил бы, но подобный мир, похоже, логически невозможен, поскольку в нем нет ничего, что может соотноситься причинностью.

Эта позиция напоминает вторую позицию, охарактеризованную в параграфе 2, где речь шла о том, что у электронов есть какая-то скрытая сущность, отсылка к которой просто фиксируется физическими описаниями. Думаю, что по причинам, изложенными там, мы не должны *отождествлять* внутренние свойства с такими физическими свойствами, как масса. Кажется вполне разумным утверждение, что в зомбийном мире все же имеется масса, несмотря на различия в его внутренней природе. Если так, то масса является внешним свойством, которое может быть «реализовано» разными внутренними свойствами в разных мирах. Но какое бы семантическое решение мы ни приняли, данная позиция сохраняет сущностную дуальность свойств, с которыми напрямую работает физика, и скрытых внутренних свойств, которые конституируют феноменологию.

В определенном смысле этот взгляд может быть охарактеризован скорее как монизм, нежели как дуализм, но это не материалистический монизм. В отличие от физикализма, данная позиция предполагает фундаментальность некоторых феноменальных илиprotoфеноменальных свойств. В итоге она говорит о сети внутренних свойств, причем по крайней мере некоторые из них являются феноменальными или protoфеноменальными, а также о том, что они соотносятся друг с другом в соответствии с определенными каузальными/динамическими законами. Эти свойства «реализуют» внешние физические свойства, а законы, связывающие их, реализуют физические законы. В предельном случае, когда все внутренние свойства оказываются феноменальными, это воззрение могло бы быть наилучшим образом охарактеризовано в качестве одной из версий идеализма. Но это идеализм, мало похожий на идеализм Беркли. Мир не супервентен на сознании наблюдателя, а скорее представляет собой громадную каузальную сеть феноменальных свойств, лежащих в основании физических законов, провозглашаемых наукой. Если взять не столь крайний случай, в котором внутренние свойства protoфеноменальны или в котором некоторые

из них ни феноменальны, ниprotoфеноменальны, то его, быть может, лучше всего рассматривать в качестве разновидности расселовского нейтрального монизма. Базовые свойства мира не являются ни физическими, ни феноменальными, но физическое и феноменальное сконструировано из них. Феноменальное конструируется из комбинации их внутренних природ, а физическое — из их внешних отношений.

Согласно этой концепции, наиболее фундаментальными законами оказываются те, которые связывают базовые внутренние свойства. Привычные физические законы фиксируют реляционную форму этих законов, абстрагируясь при этом от внутренних свойств. Психофизические законы могут быть перетолкованы как законы, связывающие внутренние свойства (или свойства, сконструированные из них) с их реляционными профилями (или с комплексными реляционными структурами). Таким образом, эти законы не являются онтологическими «довесками» физических законов. Скорее, и те и другие оказываются следствиями подлинно фундаментальных законов. Эпистемологический порядок, однако, отличается от онтологического порядка: вначале мы сталкиваемся с реляционной структурой каузальной сети и лишь затем медленно приходим к лежащим в ее основании внутренним свойствам. Для повседневных объяснительных целей поэтому полезней всего продолжать думать об этой концепции в терминах сети каузальных законов, сопряженной с дополнительными принципами, связывающими физическое с феноменальным.

Ко всем этим метафизическим спекуляциям надо относиться с некоторой долей скептизма — но они показывают, что проблема эпифеноменализма не является чем-то банальным. Имеется немало тонких проблем, связанных с причинностью и с природой опыта, с которыми требуется лучше разобраться перед тем, как мы сможем наверняка сказать, является ли опыт чем-то эпифеноменальным. В любом случае сейчас я оставлю в стороне метафизические спекуляции и возвращусь на более приземленный уровень (хотя я вернусь к некоторым из этих вопросов в главе 8).

Так или иначе, но естественная супервентность *производит впечатление* чего-то эпифеноменалистического. Можно было бы сказать, что эта позиция является эпифеноменалистической *в первом приближении*: если она и допускает каузальную релевантность опыта, то неочевидным способом. Думаю, что мы можем передать этот аспект первого приближения, отметив, что указанная концепция делает опыт *объяснительно нерелевантным*. Мы можем объяснить поведение в чисто физических или вычислительных терминах — терминах, не отсылающих к феноменологии и не предполагающих такую отсылку. Если опыт неким

неявным образом и связан с каузальностью, данные объяснения могут абстрагироваться от этой неявной связи. Даже объяснительная нерелевантность могла бы вызывать тревогу; я подробно рассмотрю этот вопрос в следующей главе.

Некоторые надеялись избежать эпифеноменализма скачком на позицию «сильной метафизической необходимости», рассмотренную во втором параграфе этой главы. Если опыт не логически супервентен на физическом, то, как казалось некоторым, единственным способом сохранения его каузальной роли может быть провозглашение того, что он изначально тождествен или метафизически супервентен на каком-то физическом свойстве или свойствах. Однако помимо уже упомянутых мной трудностей эта позиция тоже имеет серьезные проблемы с объяснительной нерелевантностью. Сам факт представимости зомби показывает, что согласно этому воззрению поведение может быть объяснено в терминах, не отсылающих к существованию опыта и не предполагающих такую отсылку. Объяснительные отношения — это концептуальные отношения, так что сильная метафизическая необходимость в данном случае не имеет отношения к делу. Эта концепция по-прежнему оставляет поведение чем-то независимым от опыта в сильном смысле слова и обречена на столкновение с большинством из тех проблем, с которыми сталкивается и дуализм свойств. Поэтому мало что можно выиграть, заняв эту позицию.

Некоторые люди, которых убедили аргументы в пользу дуализма, но которые при этом считают, что феноменальное сознание должно играть значимую каузальную роль, могут соблазниться интеракционистской версией дуализма, согласно которой опыт заполняет каузальные пробелы в физических процессах. Уступка этому искушению, однако, порождает больше проблем, чем решает. Начать с того, что это заставляет делать большую ставку на будущее физики, которая в данный момент совсем не выглядит перспективной; физические события кажутся объяснимыми исключительно в терминах других физических событий. Данная позиция требует делать большую ставку и на будущее когнитивной науки, так как она предполагает, что обычные физические/функциональные модели окажутся недостаточными для объяснения поведения. Но самая глубокая проблема состоит в том, что эта концепция может справляться с проблемами эпифеноменализма ничуть не лучше той, которая предполагает каузальную замкнутость, — по причинам, которые мы вскоре обсудим.

Единственная форма интеракционистского дуализма, которая кажется хотя бы отдаленно соответствующей современным представлениям, эксплуатирует некоторые особенности квантовой механики.

Интеракционистский дуализм?

И здесь есть два возможных пути. Во-первых, некоторые апеллировали к существованию квантовой неопределенности и допускали, что нефизическое сознание могло бы быть ответственным за заполнение возникающих вследствие этой неопределенности каузальных пробелов, детерминируя значения ряда физических величин в кажущемся «вероятностным» распределении (напр., Eccles 1986). Хотя эти решения имели бы лишь незначительный непосредственный эффект, не исключено, что нелинейная динамика могла бы усиливать эти небольшие флуктуации и превращать их в масштабные макроскопические влияния на поведение.

Это смелая и интересная идея, но она сталкивается с рядом проблем. Во-первых, данная теория противоречит постулату квантовой механики о том, что эти микроскопические «решения» совершенно случайны, и из нее в принципе следует, что в них должен быть некий паттерн, который можно было бы отыскать, — а это допускающая проверку гипотеза. Во-вторых, для того чтобы эта теория позволяла говорить, что сознание выполняет *интересную* каузальную работу, поведение, порождаемое *этими* микроскопическими решениями, должно неким образом существенно отличаться от того поведения, которое порождается большинством других множеств решений, которые могли бы быть осуществлены в ходе чисто случайных процессов. Можно предположить, что такое поведение было бы более рациональным, чем в ином случае, и что оно вело бы к таким действиям, как высказывание «сейчас я вижу красное», которые не порождались бы случайными процессами. Это опять-таки в принципе доступно проверке, если бы удалось запустить симуляцию мозга, в которой такие решения вызывались бы действительно случайными процессами. Конечно, мы не знаем наверняка, к чему привели бы эти проверки, но утверждать, что при случайных процессах мы получили бы резкие поведенческие ухудшения, — значит делать ставку с очень малыми шансами на выигрыш.

Второй путь, на котором квантовая механика увязывается с вопросом о каузальной замкнутости, исходит из того факта, что согласно некоторым интерпретациям квантового формализма само сознание играет ключевую каузальную роль и требуется для осуществления так называемого *коллапса волновой функции*. Этот коллапс, как предполагается, происходит при любом акте измерения; и, по одной из интерпретаций, измерение можно отличить от того, что не есть измерение, только через присутствие сознания. Эта теория совершенно точно не является чем-то общепризнанным (начать с того, что она *предполагает*, что само сознание не является физическим, а это, конечно, противоречит взглядам большинства физиков), и я сам не принимаю ее, но в любом случае кажется, что та разновидность каузальной работы, которую выполняет сознание в данной ситуации, совершенно отлична от той ее

разновидности, которая требуется для того, чтобы сознание играло некоторую роль в управлении поведением³¹. Неясно, каким образом коллапс, происходящий во внешних объектах восприятия, позволяет сознанию влиять на физические процессы в мозге; подобные теории обычно ничего не говорят о том, что происходит в мозге во время коллапса. И даже если сознание каким-то образом вызывает коллапс состояния мозга, то это не отменяет высказанные выше замечания относительно кажущихся случайными процессов и их связи с поведением.

В любом случае все версии интеракционистского дуализма сталкиваются с концептуальной проблемой, связанной с тем, что они не так успешно избегают эпифеноменализма, как могло бы показаться; или, по крайней мере, что они не лучше справляются с ним, чем та концепция, которую я отстаивал. Даже если мы примем эти взгляды, окажется, что в определенном смысле феноменальное не является существенным. Мы всегда можем вычесть этот феноменальный компонент из любого объяснения, оставляя в нем чисто каузальный компонент. Вообразим (вместе с Экклзом), что «психоны» в нефизическом уме запускают физические процессы в мозге и что психоны являются носителями опыта. Мы можем рассказывать о каузальных отношениях между психонами и физическими процессами, а также о каузальной динамике психонов, вообще не упоминая того факта, что психоны обладают феноменальными свойствами. Как и в случае с физическими процессами, мы можем фантазийно вычесть *феноменальные* свойства психонов, получив ситуацию с изоморфной каузальной динамикой. Из этого следует, что тот факт, что психоны являются носителями опыта, не играет существенной роли в каузальном объяснении, и что даже согласно этой картине опыт объяснительно нерелевантен.

Кто-то мог бы возразить, сказав, что психоны (или эктоплазма, или что-то еще) полностью *конституируются* их феноменальными свойствами. Но даже в этом случае их феноменальные свойства в известном смысле оказываются нерелевантными для объяснения поведения; для описания их каузальной динамики значимы только их относительные свойства. А если кто-нибудь настаивает, что у них есть и дополнительные внутренние свойства, имеющие *каузальную* релевантность, то мы оказываемся в ситуации, сходной с той, с которой мы столкнулись выше, когда говорили о феноменальном характере внутренних свойств *физических* сущностей. В любом случае мы имеем дело с некоей каузальной релевантностью и объяснительной нерелевантностью. Так что мы и в самом деле не обретаем ничего особенного, отказываясь от каузальной замкнутости физического. Мы по-прежнему видим замкнутую, пусть и более широкую, каузальную сеть и не уходим от объяснительной избыточности феноменальной природы тех сущностей, которые входят в состав этой каузальной сети.

Мы можем даже вообразить, что если интеракционизм верен, то по причинам, совершенно независимым от сознательного опыта, мы в итоге оказались бы вынуждены постулировать психоны для объяснения поведения, чтобы заполнить наблюдаемые каузальные пробелы и объяснить имеющиеся в наличии факты. В таком случае психоны получили бы статус некоторых теоретических сущностей, подобных аналогичным сущностям в физике. Но ничто из того, что мы говорили бы о них, не предполагало бы опыт или отсылку к нему, который оказывался бы столь же объяснимительно избыточным, как и в обычном случае; мы по-прежнему могли бы рассуждать о зомби с психонами и т. д. Дополнительное наблюдение о том, что эти психоны могли бы обладать феноменальными свойствами, не лучше и не хуже отвечает эпифеноменализму, чем аналогичное наблюдение о том, что физические сущности (быть может, базовые или же имеющие весьма сложный состав) могли бы обладать феноменальными свойствами в дополнение к их внешним чертам. Поэтому отрицание каузальной замкнутости физического несущественно для избежания эпифеноменализма³².

Проблемы эпифеноменализма

Любая концепция, всерьез относящаяся к сознанию, должна будет иметь дело по крайней мере с ограниченной версией эпифеноменализма. Сам факт того, что опыт может быть непротиворечиво вычен из любого каузального объяснения, означает, что он избыточен для объяснения поведения независимо от того, присуща ли ему неявная каузальная релевантность. Не исключено, что он окажется каузально нерелевантным и в более сильном смысле; вопрос этот остается открытым. Поэтому мы должны обратиться ко второй стороне нашей стратегии и выяснить, в чем именно состоят проблемы, связанные с каузальной или объяснимой нерелевантностью опыта, и являются ли они в итоге фатальными. Гораздо подробнее мы обсудим это в главе 5, но сейчас можно кратко очертить ситуацию.

Самым распространенным возражением против эпифеноменализма является то, что он попросту контринтуитивен или даже «невыносим». Утверждение о контринтуитивности или невыносимости вывода, однако, не является достаточным основанием для отрицания этого вывода, особенно если последний получен на основании сильного аргумента. Эпифеноменализм может быть контринтуитивным, но он не очевидно ложен, так что если основательная аргументация ведет нас к нему, мы должны принять его. Разумеется, контринтуитивный вывод может подтолкнуть нас к тому, чтобы вернуться к аргументу и заново исследовать его, но мы в любом случае должны найти какую-то ошибку в аргументе, опираясь на независимые от его контринтуитивности основания. Если окажется, что ложность тезиса о логической

супервентности влечет эпифеноменализм, то логическая супервентность может быть *желаемой*, но мы не можем просто объявлять о ней как о неком изначальном факте. Для признания наличия логической супервентности нам нужно каким-то образом объяснить, как из физических фактов могли бы вытекать факты о сознании, но именно это, как я доказывал, мы и не в состоянии сделать.

Более конкретные возражения против эпифеноменализма разделяются на три группы: те, что связаны с отношением опыта к обыденному поведению; те, что связаны с отношением опыта к *суждениям* об опыте, и те, что связаны с общей картиной мира, возникающей, если принять эту идею.

Начнем с первой группы. Многие считают просто очевидным, что их переживания боли каузально вызывают то, что они убирают руку от огня, или что мое переживание головной боли не может не иметь никакого значения для объяснения того, почему я принимаю таблетки. Нельзя отрицать, что сильная интуиция действительно подталкивает нас к таким выводам. Однако мы можем без труда объяснить происхождение этой интуиции в терминах систематических регулярных связей между этими событиями. Мы гораздо более непосредственно знаем об опыте и о поведении, чем о лежащем в их основании состоянии мозга; при обнаружении систематических регулярных соотношений между опытом и поведением естественно делать вывод о наличии между ними сильной каузальной связи. Даже если бы связь между ними была лишь непрямой номической связью, возникающей вследствие отношений к фундирующему их состоянию мозга, мы все равно могли бы ожидать подобного вывода. Поэтому данную интуицию можнонейтрализовать. В любом случае подобное возражение не может быть фатальным для эпифеноменализма, так как эта интуиция не может быть напрямую превращена в аргумент. Здесь мы имеем дело с примером того, что является *всего лишь* континтуитивным.

Вторая группа возражений вызывает большее беспокойство. Кажется очень странным, что наш опыт не должен иметь значения для объяснения, к примеру, того, почему мы *говорим* о наших опытных переживаниях, или даже для наших внутренних *суждений* о таких переживаниях; это кажется гораздо более странным, чем простая нерелевантность моей боли для объяснения того, почему я убираю руку. Некоторые заявляют, что подобные проблемы не просто континтуитивны, но и фатальны. К примеру, можно было бы утверждать, что это несовместимо с нашим *знанием* об опыте или с нашей способностью *отсылать* к опытным переживаниям. Я считаю, что если четко сформулировать данные аргументы, то они в итоге не позволят сделать такой вывод, но нельзя отрицать, что эти вопросы действительно выглядят серьезными. Я посвящу главу 5 их обсуждению.

Возражения третьей группы касаются общих структурных характеристик данного воззрения. Одно из них состоит в том, что оно рисует непривлекательную и малоправдоподобную картину, в соответствии с которой опыт порождается физическим, «номологическими бездельниками», не интегрированными с другими законами природы. Думаю, что с этим возражением можно справиться на путях теории, ведущей к более интегрированной картине. Название «эпифеноменализм» наводит на мысль о представлении, согласно которому опыт бездельничает «где-то там наверху», паря вне зависимости от происходящих процессов; но лучшей картиной, все же совместимой с естественной супервентностью, будет та, которая изображает опыт внедренным в низовые каузальные трещины. В крайнем же случае мы можем попытаться сделать психофизические законы настолько простыми и элегантными, насколько это вообще возможно. К этой группе возражений относится и беспокойство по поводу того, как сознание могло бы эволюционировать при эпифеноменалистских предпосылках, но нетрудно увидеть, что это не составляет проблемы для той позиции, которую я отстаиваю; подробнее я обсужу это в конце данной главы.

В результате проверки оказывается, что серьезный ущерб эпифеноменализму наносит не так уж много аргументов. Наибольшую тревогу вызывает та группа аргументов, которая связана с суждениями об опыте и которую я рассмотрю в следующей главе. Если оставить в стороне аргументы, то у многих есть *интуитивное представление* о том, что эпифеноменализм должен быть ложен, но этой интуиции недостаточно, чтобы отрицать этот взгляд, опирающийся на сильные аргументы в его пользу.

Я не характеризую свою позицию как эпифеноменализм. Вопрос о каузальной релевантности опыта остается открытым, и для решения этой проблемы потребуется более детальная теория каузальности и опыта. Но из моей позиции вытекает по крайней мере слабая форма эпифеноменализма, а в итоге она может привести и к более сильной его разновидности. Однако даже если дело обернется именно таким образом, думаю, что аргументы в пользу естественной супервентности достаточно убедительны для того, чтобы мы приняли их. Эпифеноменализм контринтуитивен, но его альтернативы более чем контринтуитивны. Как мы уже видели и еще увидим, они попросту *неверны*. И мораль всего этого состоит в том, что если аргументы говорят об истинности естественной супервентности, то мы должны научиться жить с естественной супервентностью.

Некоторые тем не менее считают, что эпифеноменалистский характер данного воззрения является его фатальным дефектом. Мне в чем-то симпатична такая позиция, которую можно рассматривать как вы-

ражение парадокса сознания: когда идет речь о сознании, может показаться, что *все* альтернативы плохи. Думаю, однако, что проблемы, связанные с другими воззрениями, фатальны в гораздо более сильном смысле, чем континтуитивность этого. И раз какая-то из опций логического пространства должна быть верной, это воззрение кажется единственным разумным кандидатом.

5. Логическая география проблем

Аргумент в пользу моей позиции, навскидку, есть вывод из четырех посылок:

1. Сознательный опыт существует.
2. Сознательный опыт не является логически супервентным на физическом.
3. Если существуют феномены, не являющиеся логически супервентными на физических фактах, то материализм ложен.
4. Область физического каузально замкнута.

Из посылок (1), (2) и (3) с очевидностью вытекает ложность материализма. И это вместе с посылкой (4) и правдоподобным допущением о том, что физически идентичные существа будут иметь идентичный сознательный опыт, ведет к взгляду, который я назвал естественной супервентностью: сознательный опыт возникает из физического в соответствии с определенными законами природы, но сам не является физическим. Различные альтернативные позиции могут быть классифицированы в соответствии с тем, отрицают ли они посылки (1), (2), (3) или (4). Разумеется, некоторые из этих посылок можно отрицать разными способами.

Отрицание посылки (1):

- i. **Элиминативизм.** Согласно этому взгляду, позитивных фактов о сознательном опыте нет. Никто не сознателен в феноменальном смысле.

Отрицание посылки (2):

Посылку (2) можно отрицать по-разному в зависимости от того, как происходит выведение, о котором идет речь, — то есть в зависимости от того, какие физические свойства главным образом ответственны за выведение сознания. Я называю все эти взгляды «редуктивным материализмом», поскольку все они предполагают возможность такого анализа понятия сознания, который оказывается совместимым с редуктивным объяснением.

- ii. *Редуктивный функционализм.* Согласно этому воззрению, сознание концептуально выводимо из физического вследствие функциональных или диспозиционных свойств. Быть сознательным для какого-то состояния — значит играть определенную каузальную роль. В мире, физически идентичном нашему, исполнялись бы все релевантные каузальные роли, и поэтому все сознательные состояния были бы теми же самыми. Зомбийный мир, таким образом, является чем-то логически невозможным.
- iii. *Нефункционалистский редуктивный материализм.* В соответствии с этой позицией факты о сознании концептуально выводимы из физических фактов вследствие какого-то нефункционального свойства. Возможными кандидатами могут быть биохимические и квантовые свойства или же такие свойства, которые еще предстоит установить.
- iv. *Материализм новой физики.* Согласно этому взгляду, сейчас у нас нет представления о том, как физические факты могли бы объяснить сознание, но это связано с тем, что наши нынешние представления о физических фактах излишне заужены. Когда кто-то доказывает логическую возможность зомбийного мира, в действительности он доказывает, что все поля и частицы, взаимодействующие в пространственно-временном многообразии, постулируемом современной физикой, могли бы существовать и при отсутствии сознания. Но новая физика может внести изменения в эту картину. В радикально иной теоретической модели этих сущностей могло бы оказаться достаточно для выведения и объяснения сознания.

Отрицание посылки (3):

- v. *Нередуктивный материализм.* Это позиция, согласно которой, хотя факты о сознании и могут быть логически невыводимы из физических фактов, и поэтому сознание не может быть редуктивно объяснено, оно *все же* физично. Физические факты «метафизически вынуждают» факты о сознании. Несмотря на то что идея зомбийного мира совершенно непротиворечива, подобный мир является метафизически невозможным.

Отрицание посылки (4):

- vi. *Интеракционистский дуализм.* Эта позиция допускает нефизичность сознания, но отрицает, что физический мир каузально замкнут, так что сознание может играть самостоятельную каузальную роль.
- И есть моя позиция, принимающая посылки (1), (2), (3) и (4):
- vii. *Натуралистический дуализм.* Сознание естественно супервентно на физическом, но оно не является логически или «метафизически» супервентным.

Встречается и восьмая позиция, характеризующаяся общей недостаточной определенностью:

viii. *Ничего не понимающий материализм.* «Я ничего не понимаю в сознании. Оно кажется мне совершенно таинственным. Но оно должно быть физическим, так как материализм должен быть истинным». Этот взгляд широко распространен, но не в печатных версиях (хотя см. (Fodor 1992)).

Кратко подытожим ситуацию, как она видится мне. Вариант (i) кажется очевидно ложным; (ii) и (iii) опираются на неверный анализ понятия сознания и поэтому подменяют предмет обсуждения; (iv) и (vi) делают слишком большие и малоперспективные ставки на будущее физики, а также сталкиваются с фатальными концептуальными проблемами; а (v) либо неоправданно апеллирует к крипкеанской апостериорной необходимости, либо опирается на экстравагантные метафизические допущения. Я пытаю определенную симпатию к (viii), но эта позиция, судя по всему, в конечном счете должна трансформироваться в одну из более конкретных концепций, а ни одна из этих концепций, как кажется, не работает. В таком случае единственным приемлемым вариантом остается вариант (vii).

Теперь обсудим все это в более медленном темпе. Начнем с вариантов (iv) и (vi). Вариант (vi), интеракционистский дуализм, требует, чтобы в физике обнаружились некие пробелы, которые могут быть заполнены действиями нефизического ума. Имеющиеся факты указывают на то, что это едва ли произойдет. Вариант (iv) требует столь радикального изменения облика физики, что из нее можно было бы выводить факты о сознательном опыте; но никто не представляет, каким образом *какая бы то ни было* физика могла добиться такого результата. В самом деле, если согласиться, что физика в конечном счете имеет дело со структурными и динамическими свойствами, то представляется, что все, что можно будет когда-либо вывести из физики, будет иметь отношение к новым структурам и динамическим характеристикам, из которых (если не принимать другие редуктивные варианты) никогда не будет выводимо существование опыта.

Глубочайшей причиной, по которой надо отвергнуть варианты (iv) и (vi), является то, что они в конечном счете сталкиваются с той же самой проблемой, что и более привычная физика: феноменальный компонент может быть без противоречий вычен в них из каузального компонента. Если говорить об интеракционистском взорении, то мы видели, что даже если нефизические сущности имеют какой-то феноменальный аспект, мы можем непротиворечиво вообразить удаление этого феноменального компонента, оставаясь с чисто каузальной/динамической историей, характеризующей взаимодействие и по-

ведение соответствующих сущностей. Если же говорить о позиции, указывающей на роль новой физики, то даже если она недвусмысленно вводит в свой состав феноменальные свойства, факт *феноменальности* этих свойств не может играть существенной роли в каузальной/динамической истории; мы остались бы с когерентным физическим взорением, даже если бы мы вычли этот аспект. В любом случае динамика — это всё, что нам нужно для объяснения каузальных взаимодействий, и никакой набор фактов о динамике не сможет выдать факт о феноменологии. Значит, мы по-прежнему сможем рассказывать историю о зомби.

В ответ на это можно предпринимать самые разные маневры, но каждый из них может быть осуществлен и при стандартных физических представлениях. К примеру, быть может, абстрактная динамика упускает из виду тот факт, что нефизическое начало в интеракционистской концепции является сущностно феноменальным, и поэтому феноменальные свойства оказываются глубоко вовлеченными в каузальную сеть. Но равным образом возможно и то, что абстрактная динамика физики упускает из виду тот факт, что ее базовые сущности являются сущностно феноменальными (в конце концов, физика характеризует их только с внешней стороны) — и итог будет тем же. В любом случае, мы имеем дело с однотипной объяснительной нерелевантностью внутренних феноменальных свойств для каузальной/динамической истории. Шаг к интеракционизму или к новой физике поэтому не решает каких-либо проблем, присущих концепции дуализма свойств, которую я отстаиваю. В итоге они могут рассматриваться как более запутанные вариации той же позиции.

Что же касается варианта (iii), наиболее привлекательной выглядит та его разновидность, которая указывает на неизвестные и пока не замеченные нами свойства как на ключ к выводимости. Но в конечном счете мы сталкиваемся с той же проблемой: физика говорит нам только о структуре и динамике, а структура и динамика не ведут к феноменологии. Единственными доступными свойствами казались бы те свойства, которые характеризуют физическую структуру или функцию, или же свойства, сконструированные из них. Но структурные свойства явно не могут быть результатом анализа понятия сознания; ненамного лучше обстоит дело и с функциональными свойствами (хотя ниже я еще рассмотрю их). Любая подобная концепция в итоге приводит к подмене предмета обсуждения.

Остаются варианты (i), (ii), (v) и (vii), соответствующие позициям, вызывающим наиболее серьезное отношение в современной литературе: элиминативизму, редуктивному функционализму, нередуктивному материализму и дуализму свойств. Вариант (i) я отвергаю как находящийся в противоречии с очевидными фактами. Быть может, с помощью

какого-то экстраординарного аргумента и можно было бы показать, что сознательный опыт не существует, но я никогда не видел аргумента, имеющего хотя бы отдаленное отношение к доказательству этого тезиса. При отсутствии подобного аргумента останавливаться на варианте (i) — значит попросту уходить от проблемы путем отрицания интересующего нас феномена.

Вариант (v) зачастую оказывается привлекательным для тех, кто хочет относиться к сознанию всерьез и вместе с тем сохранить материализм. Я, однако, доказывал, что это попросту не работает. Нередуктивный материализм, отстаиваемый Серлом, сталкивается с внутренними проблемами и коллапсирует в одну из других концепций (скорее всего, в дуализм свойств). Другие защитники этого подхода обращаются к крипкеанской апостериорной необходимости, но та разновидность апостериорной необходимости, которая выявляется Крипке, не может спасти материализм. Вариант (v) можно последовательно принимать лишь при апелляции к сильной метафизической необходимости, выходящей за пределы крипкеанской апостериорной необходимости, и накладывании изначальных ограничений на пространство «метафизически возможных» миров. Мы видели, что у нас нет оснований верить в подобные ограничения или в некую третью, промежуточную разновидность возможности миров. Эти метафизические допущения не подкрепляются никакими другими феноменами, и трудно понять, как они *могли бы* подкрепляться ими.

Но даже если принять эту метафизику необходимости, при решении большинства объяснительных задач эта позиция будет походить на тот взгляд, который я отстаиваю. Из нее следует, что сознание не может быть редуктивно объяснено. Из нее следует, что сознательный опыт объяснительно нерелевантен в области физического. И из нее следует, что теория сознания должна прибегать к связующим принципам для соединения физического и феноменального, принципам, которые сами по себе не могут быть выводимыми из физических законов. В данной концепции эти принципы объявляются «метафизически необходимыми», но во всех практических отношениях ее итог оказывается тем же самым. Подобная теория будет иметь те же *очертания*, что и отстаиваемые мной дуалистические теории, и почти всё, что я буду говорить в ходе разработки нередуктивной теории в ряде последующих глав, будет в раной степени применимо и к ней.

Вариант (ii), редуктивный функционализм — самый серьезный из всех материалистических подходов. Если оставить в стороне различные безумные концепции, то, если материализм истинен, сознание является логически супервентным, и единственная имеющая хотя бы отдаленную видимость правдоподобия трактовка его возможной логической супервентности может быть дана на путях функционального

анализа. Таким образом, согласно этой позиции, сознательный опыт означает лишь исполнение определенной каузальной роли в какой-то системе. Феноменальные свойства трактуются точно так же, как психологические свойства, связанные, к примеру, с обучением или категоризацией.

Проблема этой позиции состоит, разумеется, в том, что она неверно интерпретирует смысл сознательного опыта или сознательности. Когда я интересуюсь, являются ли другие существа сознательными, я интересуюсь не их способностями или внутренними механизмами, о которых я уже могу располагать всей информацией; я интересуюсь тем, имеется ли у них нечто, выражющее то, каково это — быть такими существами. Есть много известных способов подкрепить эту мысль. Можно, к примеру, отметить, что даже после того как мы объяснили различные функциональные способности, это может не решить проблему объяснения опыта. Можно сослаться на то, что мы можем вообразить исполнение любой функциональной роли при отсутствии сознательного опыта. Можно также опереться на тот факт, что знание функциональных ролей автоматически не дает знания о сознании. Есть и другие возражения, выдвигавшиеся ранее и связанные с тем, что функционалистский анализ не может объяснить семантическую определенность атрибутирований сознания и приводит к коллапсу концептуального различия сознания и осведомленности.

В конечном счете редуктивный функционализм мало чем отличается от элиминативизма. Оба этих воззрения допускают существование различия, категоризации, доступности, способности дать отчет и т. п.; и в обоих отрицается существование чего-то еще, что нуждается в объяснении. Главное различие между ними состоит в том, что редукционисты полагают, что некоторые из объясняемых данных заслуживают названия «опыт», тогда как элиминативисты отрицают это. Помимо этого терминологического разногласия, суть этих взглядов очень близка. Часто отмечается, что граница между редукционизмом и элиминативизмом размыта и что редукция постепенно соскальзывает к элиминации по мере того, как мы оказываемся вынужденными модифицировать соответствующие понятия для осуществления редукции. Допускать существование сознания лишь в той степени, в какой оно определяется как какая-то функциональная способность, позиция редуктивного функционалиста наносит такой урон понятию сознания, что ее, возможно, лучше трактовать как одну из версий элиминативизма. Оба этих взгляда не принимают сознание всерьез.

Это оставляет нас с вариантом (vii), а именно с дуализмом свойств, который я отстаивал, как единственным приемлемым вариантом. Нельзя отрицать, что он кажется выводом из хорошо обоснованных

посылок. В определенных отношениях эта позиция континтуитивна, но это единственная концепция, лишенная какого-либо фатального дефекта. Кто-то может счесть неприемлемым ее дуалистический характер; но вскоре я попытаюсь доказать, что подобный дуализм не является столь уж неправдоподобным, как казалось многим, и что в его адрес можно высказать не так уж много серьезных возражений. Самая большая проблема, связанная с этим воззрением, состоит в том, что из него вытекает некая нерелевантность феноменальных свойств для объяснения поведения и что оно может приводить к эпифено-менализму, хотя и не автоматически. В следующей главе я, однако, буду доказывать, что эта объяснительная нерелевантность не имеет фатальных последствий. В итоге эта позиция позволяет создать непротиворечивую, натуралистичную, лишенную мистики картину сознания и того места, которое оно занимает в природном порядке.

Охватив более широким взглядом эту логическую географию, мы можем сказать, что существует три главных класса представлений о сознательном опыте.

Тип А, тип В
и тип С

Концепции *A-типа* признают — по различным функционалистским или элиминативистским соображениям, — что сознание в той мере, в какой оно существует, логически супервентно на физическом. В концепциях *B-типа* признается, что сознание не является логически супервентным и что нет априорного импликативного перехода от физического к феноменальному; при этом они говорят о своей материалистичности. Концепции *C-типа* отрицают как логическую супервентность, так и материализм.

Концепции *A-типа* существуют во множестве разновидностей — элиминативизм, бихевиоризм, разного рода редуктивный функционализм, — но у всех них есть нечто общее. Теоретик *A-типа* скажет, что (1) физические и функциональные дубликаты, лишенные того типа опыта, которым мы обладаем, непредставимы; (2) Мэри ничего не узнает о мире, когда она впервые видит красное (в лучшем смысле она обретает какое-то умение); и (3) всё, что надо объяснить относительно сознания, может быть объяснено объяснением реализации определенных функций. Среди образцовых теоретиков *A-типа* — Армстронг (Armstrong 1968), Деннет (Dennett 1991), Льюис (Lewis 1966) и Райл (Ryle 1949). К числу теоретиков *A-типа*, возможно, также относятся Дретске (Dretske 1995), Рей (Rey 1982), Розенталь (Rosenthal 1996), Смарт (Smart 1959), Уайт (White 1986) и Уилкис (Wilkes 1984).

Концепции *B-типа*, или нередуктивные разновидности материализма, обычно разрушаются под грузом внутренних проблем. Единственной концепцией *B-типа*, кажущейся хотя бы внутренне когерентной, является та, что придает ключевую роль сильной метафизической

необходимости. Занимая эту позицию, теоретик В-типа должен считать, что (1) зомби и инвертированные спектры представимы, но метафизически невозможны; (2) Мэри узнает что-то, когда она видит красное, но от этого знания можно отделаться с помощью анализа лоровского типа; и (3) сознание не может получить редуктивное объяснение, и тем не менее оно имеет физическую природу. Главная позиция В-типа никогда четко не формулировалась, но ближе всего к ней подошли Левин (Levine 1983, 1993) и Лор (Loar 1990). Среди других авторов, которые, как кажется, одобряют физикализм, но не логическую супервентность, — Бирн (Burge 1993), Фланаган (Flanagan 1992), Хилл (Hill 1991), Хорган (Horgan 1984b), Лайкан (Lykan 1995), Папино (Papineau 1993), Тай (Tye 1995) и ван Гуллик (van Gulick 1992).

Взгляды С-типа включают различные виды дуализма свойств, признающего ложность материализма и допускающего нередуцируемость некоторых феноменальных илиprotoфеноменальных свойств. Согласно этой позиции, (1) зомби и инвертированные спектры логически и метафизически возможны; (2) Мэри узнает что-то новое, и ее знание является знанием о нефизических фактах; и (3) сознание не может быть редуктивно объяснено, но могло бы нередуктивно объясняться в терминах дополнительных законов природы. Это позиция Кембелла (Campbell 1970), Хендриха (Honderich 1981), Джексона (Jackson 1982), Г. Робинсона (H. Robinson 1982), У. Робинсона (W. Robinson 1988), Сприджа (Sprigge 1994) и автора настоящей книги.

Стоит, пожалуй, отдельно упомянуть уже обсуждавшуюся позицию, отождествляющую феноменальные свойства с внутренними свойствами физических сущностей. Подобную позицию занимали Фейгл (Feigl 1958), Локвуд (Lockwood 1989), Maxwell (Maxwell 1978) и Рассел (Russell 1926). Я тоже симпатизирую ей. Я рассматриваю ее как разновидность С-воззрений, так как здесь признается фундаментальность феноменальных илиprotoфеноменальных свойств, хотя ей и присущи свойственные только ей метафизические очертания. В частности, она более монистична, чем наиболее естественные интерпретации С-воззрений. Возможно, стоит называть ее позицией типа С', хотя обычно я буду относить ее к типу С.

На пути А, В и С имеется две главные развики. Во-первых, является ли сознание логически супервентным (тип А против остальных)? Во-вторых, является ли истинным физикализм (тип В против типа С)? Если начать со второй развики, то я не испытываю особых колебаний относительно отрицания типа В. Хотя его достоинством является то, что мы видим здесь серьезное отношение к сознанию, позиции этого типа опираются на некогерентные или туманные метафизические допущения, к тому же во многом немотивированные; главной

мотивацией выступает просто стремление любой ценой избежать дуализма. В конечном счете это воззрение имеет ту же объяснительную форму, что и тип С, — но с налетом метафизической таинственности. Тип С характеризуется большей прямотой.

Главной разноткой является выбор между типом А и остальными типами. Редуктивный функционализм и элиминативизм кажутся мне ложными с такой очевидностью, что я с трудом представляю, как кто-то мог бы принимать воззрения, относящиеся к типу А. Мне кажется, что подобные воззрения можно было бы принимать лишь в том случае, если бы принимающие их считали, что с сознанием вообще не связано каких-либо значительных проблем. Тем не менее опыт показывает, что почти треть людей готова без особых колебаний принять позицию А-типа. Это указывает на Великое Разделение, упомянутое в предисловии, — грань между позициями, принимающими сознание всерьез, и теми, которые не делают этого.

Во многих отношениях грань, отделяющая тип А от остальных типов, является гораздо более глубокой, чем та, что имеется между типами В и С. Различие между двумя последними типами связано с весьма тонкими метафизическими проблемами, тогда как первое из них затрагивает самые что ни на есть фундаментальные интуиции. Несмотря на то что как тип В, так и тип А являются «материалистическими» воззрениями, воззрения типа В по своему духу оказываются гораздо ближе к воззрениям типа С. Обе этих позиции, в отличие от А-типа, признают глубину проблемы сознания.

В конечном счете при обсуждении этой проблемы мы не можем двигаться дальше, опираясь на аргументы. Если кто-то настаивает, что объяснение доступа и способности давать отчет объясняет всё; что Мэри ничего не узнает о мире, впервые получая опыт красного, и что функциональный изоморф, отличающийся сознательным опытом, непредставим, то я могу лишь заключить, что, когда речь заходит о сознании, мы оказываемся в разных плоскостях. Быть может, наша внутренняя жизнь драматически разнится. Быть может, один из нас «когнитивно замкнут» по отношению к интуициям другого. Более вероятно, однако, что один из нас запутался или находится под влиянием какой-то догмы. В любом случае как только диалектика доходит до этого места, мы оказываемся на мосту, который мы не можем пересечь с помощью аргументации. Здесь мы можем лишь грубо скрестить интуиции, как нередко бывает при обсуждении глубоких философских вопросов. Детальная аргументация в силах содействовать тому, чтобы обосновать и описать эту схватку, но не решить ее исход.

В начале этой работы я сказал, что мой подход предполагает серьезное отношение к сознанию. Теперь мы можем уточнить, о чем идет

речь. Серьезно относиться к сознанию — значит признавать, что помимо реализации различных функций имеется и еще что-то интересное и нуждающееся в объяснении³³. Это имеет статус *prima facie* посылки, которую мог бы опровергнуть лишь исключительно сильный аргумент. И ни один из известных мне аргументов даже не приближался к этому результату. В самом деле, теоретики А-типа обычно *не выдвигают аргументов* против этой посылки, а попросту отрицают ее. И наоборот, за какой-то чертой оказывается почти невозможным приводить аргументы *в пользу* этой посылки, как и в пользу существования сознательного опыта. В лучшем случае можно попробовать прояснить эти моменты в надежде на просветление.

После прояснения указанных моментов читатели могут решить для себя, стоит ли всерьез отнести к сознанию. Всё, что я утверждаю, — так это то, что если мы всерьез относимся к сознанию, то дуализм свойств оказывается единственным разумным вариантом. Если мы отвергли редуктивный функционализм и элиминативизм, то нам придется признать, что сознание не является логически супервентным на физическом. А если мы отвергли логическую супервентность, то перед нами прямой путь к дуализму свойств. Воззрения В-типа популярны, но, как кажется, не выдерживают серьезной философской проверки. И главный метафизический выбор, который остается сделать, — это выбор между стандартным воззрением С-типа и воззрением С'-типа. Мы не обязаны сразу решать этот вопрос — я сам не решил его для себя, — но в любом случае можно заключить, что, если мы хотим всерьез относиться к сознанию, мы должны признать фундаментальность феноменальных илиprotoфеноменальных свойств.

Некоторые из концепций, представленных в философской литературе, не укладываются с очевидностью в нарисованную мной схему. Но эта схема позволяет без труда локализовать их и проанализировать их проблемы. В примечаниях я кратко обсуждаю девять таких позиций: биологический материализм³⁴, физикалистский функционализм³⁵, психофункционализм³⁶, аномальный монизм³⁷, репрезентационализм³⁸, концепцию сознания как высокоуровневого мышления³⁹, редуктивный телеофункционализм⁴⁰, концепцию эмерджентной каузальности⁴¹ и мистерианство⁴².

6. Размышления о натуралистическом дуализме

Многие, в том числе и я сам в прошлом, полагали, что можно относиться всерьез к сознанию и одновременно оставаться материалистом. В этой главе я доказывал, что это невозможно — по очень ясным

причинам. Мораль в том, что если мы хотим иметь дело с этим феноменом, то мы обязаны в каком-то виде принимать дуализм. Можно было бы сказать, что мы не можем съесть материалистический пирог и закусить его сознанием.

Тем не менее многие озабочиваются поисками какой-то альтернативы той позиции, которую я представил, так как считают неприемлемым ее дуалистический характер. Такая реакция естественна, учитывая различные негативные ассоциации, возникающие в связи с дуализмом, но я подозреваю, что она основана всего лишь на каких-то современных догмах. Чтобы убедиться в этом, стоит рассмотреть различные основания, которые могли бы подтолкнуть к отрицанию дуализма в пользу материализма, и в объективном свете оценить аргументативный вес этих оснований.

Первое основание для предпочтения материализма — *простота*. Это хорошее основание. При прочих равных условиях мы должны предпочесть более простую теорию той, что является более расточительной в онтологическом плане. Бритва Оккама указывает нам на то, что мы не должны умножать сущности без необходимости. Но здесь нет равенства условий и *есть* необходимость. Мы видели, что материализм не может объяснить феномены, которые нуждаются в объяснении. Так же как Максвелл, вводя электромагнитные поля для объяснения определенных естественных феноменов, пожертвовал простой механистической картиной мира, мы должны принести в жертву простое физикалистское мировоззрение для того, чтобы объяснить сознание. Мы отдали должное Оккаму, признав, что для опровержения материализма нам нужны сильные аргументы. Но когда такие аргументы представлены, бритва Оккама не в силах спасти его.

Второе и, быть может, самое действенное основание в пользу материализма имеет индуктивный характер: материализм работал всегда и везде. Если взять такие феномены, как жизнь, познание или погода, то мы или уже располагаем их материалистическими объяснениями, или у нас есть серьезные основания полагать, что они где-то на подходе. Почему ситуация с сознанием должна чем-то отличаться от ситуаций такого рода?

Это основание, однако, легконейтрализовать. Как мы видели, существует простое объяснение успеха материалистических теорий в различных внешних областях. Когда речь идет о таких феноменах, как обучение, жизнь и погода, объяснены должны быть лишь структуры и функции. Если исходить из каузальной замкнутости физического, то следует ожидать физического объяснения таких структур и функций. Но уникальность случая с сознанием в том, что здесь надо объяснить не только структуры и функции, так что едва ли стоит ожидать сходства самого объяснения.

В самом деле, в главе 2 мы видели, что, принимая во внимание природу нашего доступа к внешним феноменам, мы должны *ожидать* успеха материалистического объяснения любых подобных феноменов. Наше знание об этих феноменах физически опосредовано — светом, звуком и другими перцептивными медиаторами. При условии каузальной замкнутости физического мы должны ожидать, что феномены, которые мы наблюдаем при их посредстве, логически супервентны на физическом — в ином случае мы ничего не узнали бы о них. Но наш эпистемический доступ к сознательному опыту является доступом совершенно иного рода. Сознание находится в самом центре нашего эпистемического универсума, и наш доступ к нему не является чем-то перцептивно опосредованным. Основания ожидать материалистического объяснения внешних феноменов, таким образом, не работают в случае сознания, и любой индуктивный вывод, который может быть сделан исходя из них, в лучшем случае будет шатким.

В-третьих, многие отдавали предпочтение материализму, чтобы *всерьез относиться к науке*. Считалось, что дуализм сам по себе бросает вызов науке. Согласно Черчленд (Churchland 1988), «дуализм не стыкуется с эволюционной биологией, а также с современной физикой и химией». Но это совершенно не так. Ничто в той дуалистической позиции, которую я защищаю, не требует от нас подвергать сомнению науку. Каузальная замкнутость физического сохраняется; физика, химия, нейронаука и когнитивная наука могут идти своим привычным путем. В своих областях физические науки в полной мере успешны. Они замечательно объясняют физические феномены; они лишь не могут объяснить сознательный опыт.

Черчленд указывает и на другие основания отказа от дуализма: (1) на систематическую зависимость ментальных феноменов от нейробиологических феноменов; (2) на современные вычислительные достижения, показывающие, что комплексные результаты могут достигаться и без нефизического гомункула; и (3) на отсутствие свидетельств в пользу дуализма, а также дуалистических объяснений и методологии. Первые два основания не работают против моей позиции. Что же касается третьего, то я уже привел аргументы в пользу дуализма, а дуалистические объяснения и методология будут проиллюстрированы в оставшейся части моей работы.

Четвертый мотив избегать дуализма возникает у многих вследствие разного рода спиритуалистических, религиозных, супранатуральных и других антинаучных спекуляций по поводу данного воззрения. Но они совершенно несущественны. Согласно той позиции, которую я отстаиваю, сознание управляет естественным законом, и не исключено, что в конечном счете о нем можно будет создать внятную

научную теорию. Не существует априорного принципа, говорящего о том, что все естественные законы должны быть физическими законами; отрицание материализма не означает отрицание натурализма. Натуралистический дуализм расширяет нашу картину мира, но не приводит к появлению на ней каких-то сил тьмы.

Сходное беспокойство вызывает то обстоятельство, что, как считали многие, одобрение дуализма привело бы к отказу от объяснений. Говоря словами Деннетта (Dennett 1991), «при любви дуализма к тайнам принять дуализм — значит сдаться» (с. 37). Не исключено, что это так относительно каких-то видов дуализма, но это не является его автоматическим следствием, как, надеюсь, станет ясно из оставшихся разделов книги.

Иногда можно услышать и пятое возражение против дуализма, состоящее в том, что он не может объяснить, как происходит взаимодействие физического и нефизического. Но в модели естественной супервентности ответ прост: они взаимодействуют благодаря психофизическим законам. Существует система законов, гарантирующих, что данная физическая конфигурация будет сопровождаться опытом такого-то рода, так же как существуют законы, диктующие то, что данный физический объект будет определенным образом гравитационно воздействовать на остальных.

Можно было бы возразить, что это не говорит нам, чем именно является эта связь или как некая физическая конфигурация порождает опыт. Но искать подобную связь было бы ошибкой. Даже в фундаментальных физических законах мы не можем отыскать того «связующего звена», которое и осуществляет всю работу. События просто происходят в соответствии с законом; и начиная с какого-то момента нельзя спрашивать «как». Как показал Юм, поиск подобных предельных связей не может привести к успеху. Если такие связи действительно существуют, они совершенно таинственны как в физическом, так и в психофизическом случае, и поэтому последний из них не создает здесь нам какой-то особой проблемы.

Примечательно, что противники Ньютона выдвигали сходное возражение против теории гравитации: *как* одно тело может оказывать воздействие на другое удаленное тело? Но со временем это возражение утратило силу. Мы научились принимать какие-то вещи за нечто изначальное.

Иногда беспокойство вызывает также и вопрос о том, как в дуалистической модели могло бы эволюционировать сознание: не происходило бы так, что здесь, словно в фокусе, в природе неожиданно появлялся какой-то новый элемент? Но это не проблема. Подобно фундаментальным законам физики, психофизические законы вечны

и существовали с начала времен. Дело могло обстоять таким образом, что на ранних этапах существования Вселенной не было ничего, что удовлетворяло бы физическим антецедентам этих законов, а поэтому не было и сознания, хотя это зависит от природы законов. В любом случае по мере развития Вселенной в результате эволюции возникли определенные физические системы, удовлетворяющие соответствующим условиям. При появлении этих систем сознательный опыт автоматически стал сопровождать их в результате действия тех самых законов, о которых мы говорили. Если признать реальность и вневременность психофизических законов, что и делает натуралистический дуализм, эволюция сознания не будет особой проблемой.

Короче говоря, лишь очень немногие из привычных оснований отвергать дуализм затрагивают отстаиваемую мной позицию. Главной остаточной мотивацией отрицать его могут быть просто негативные коннотации соответствующего термина, а также тот факт, что он идет вразрез с теми убеждениями, с которыми росли многие из нас. Но если мы отвлечемся от этих ассоциаций, то мы увидим, что нет оснований считать, что дуализм не может быть разумной и вполне приемлемой позицией. И я считаю, что очерченная мной концепция является той концепцией, с которой может научиться жить и которую может даже ценить тот, кто считает себя материалистом, но при этом хочет всерьез относиться к сознанию.

И впрямь моя концепция — это воззрение, которое и сейчас могут неявно разделять многие из тех, кто называет себя «материалистами». Всё, что я сделал, — так это раскрыл онтологические следствия естественной позиции, в частности той, что сознание «возникает» из физического. Некоторые дуалисты могут даже счесть мои воззрения слишком материалистическими на их вкус, но тут уж ничего не поделаешь. В идеале это позиция, использующая достоинства тех и других и лишенная их недостатков.

Этот дуализм, таким образом, не требует от нас отказа от множества *важных* элементов научного мировоззрения наших дней. Он требует лишь того, чтобы мы отбросили догму. В остальном эта позиция просто дополняет данное мировоззрение; это необходимое расширение, нужное для того, чтобы включить сознание в состав научного мировоззрения. Наше кредо таково: если это дуализм, то мы должны научиться любить дуализм.

Глава 5

Парадокс феноменального суждения

1. Сознание и познание

Пока мы главным образом акцентировали различия и грани между сознанием и познанием. Сознание таинственно; познание — нет. Сознание есть онтологическая новация; познание — бесплатное онтологическое приложение. Познание можно объяснить функционально; сознание сопротивляется такому объяснению. Познание подчиняется только законам физики; сознание отчасти подчинено независимым психофизическим законам.

Хотя фокусировка на этих различиях была необходима для разрешения множества тонких метафизических и объяснительных проблем, связанных с сознательным опытом, она может способствовать возникновению неверной картины ментального. Согласно этой картине, сознание и познание полностью оторваны друг от друга и живут независимой жизнью. У кого-то могло бы создаться впечатление, что теория сознания и теория познания практически не имеют отношения друг к другу.

Такая картина неверна. Наша ментальная жизнь не содержит того самоотчуждения, на мысль о котором наводит данная картина. Между сознанием и познанием существуют глубокие и фундаментальные связи. С одной стороны, содержание нашего сознательного опыта тесно соотнесено с содержанием наших когнитивных состояний. Когда у кого-то имеется феноменально индивидуализированное ощущение зеленого, у него имеется и соответствующее психологически индивидуализированное *восприятие* зеленого. С другой стороны, значительная часть познавательной активности может быть сосредоточена на сознательном опыте. Мы знаем о наших опытных переживаниях и выносим суждения о них; когда я пишу это, мои мысли по большей части заняты сознанием. И эти отношения между сознанием и познанием не случайны и произвольны, а систематичны.

Анализ этого систематического отношения может дать немало исходного материала для построения теории сознания. Действуя подоб-

ным образом, мы можем убедиться, что природа познания является не тем, что не имеет отношения к сознанию, а тем, что занимает центральное место в его объяснении. Конечно, теория познания не может провести всю объяснительную работу только своими силами, но тем не менее может сыграть важную роль. В конце концов, мы можем взяться за сознание только благодаря познанию. Детальное исследование связей между сознанием и познанием может стать той точкой опоры, в которой мы нуждаемся для конкретизации теории сознания и которая, быть может, в итоге позволит нам прийти к такой трактовке сознания, которая, с одной стороны, не будет чем-то мистическим, а с другой — не будет тривиализировать этот феномен.

В данной главе я заложу основы для исследования отношения между сознанием и познанием. В следующем параграфе я введу ряд понятий, занимающих центральное место в этом отношении. Оставшаяся часть главы имеет в основном оборонительный характер и посвящена обсуждению различных проблем, которые, как могло бы показаться, ставит перед нередуктивистской концепцией такое отношение между сознанием и познанием. В следующей главе я приступаю к решению задачи по построению позитивной теории, систематизирующей отношение между сознанием и познанием, для сопряжения этих аспектов в цельной картине ментального.

Феноменальные суждения

Отношение между сознанием и познанием обнаруживается прежде всего в той их связи, которая имеется в *феноменальных суждениях*.

Наш сознательный опыт не обретается в изолированном феноменальном вакууме. Мы осведомлены о нашем опыте и его содержании, мы выносим суждения об этом опыте, а затем можем высказываться о нем. Когда у меня имеется ощущение красного, я иногда формирую убеждение, что у меня имеется ощущение красного, которое может привести к верbalному отчету о нем. На более абстрактном уровне по итогам размышлений о загадках, которые загадывает нам сознание, размышлений, которым я предавался на протяжении этой книги, мы выносим суждения о сознании. На более конкретном уровне мы часто судим об *объектах* нашего сознательного опыта (к примеру, в нашем окружении), скажем, когда мы думаем: «Тут что-то красное». Различные суждения, связанные с сознанием, я называю *феноменальными суждениями* — не потому, что они сами по себе являются феноменальными состояниями, а потому, что они связаны с феноменологией или с объектами, имеющими феноменальный характер.

Феноменальные суждения зачастую отражаются в *высказываниях* о сознании — вербальных выражениях таких суждений. Высказывания людей о сознании могут быть самыми разными в разное время:

от «я чувствую пульсирующую боль» или «ЛСД вызывает у меня причудливые ощущения цвета» до «проблема сознания крайне сложна». Эти высказывания и суждения тесно связаны с нашей феноменологией, но в конечном счете они являются частью нашей психологии. Верbalные отчеты — это поведенческие акты, и поэтому они допускают функциональное объяснение. Сходным образом и феноменальные суждения сами по себе являются когнитивными актами и относятся к области психологии.

Часто утверждается, что убеждения должны пониматься как функциональные состояния, характеризуемые их каузальными связями с поведением, окружением и другими убеждениями, но этот взгляд не является общепринятым. Некоторые считают, что убеждение или его содержательная сторона могут отчасти конституироваться феноменальным опытом. Особенно спорным функционалистский взгляд кажется в его применении к убеждениям о сознании: если какие-то убеждения зависят от сознательного опыта, то лучшими кандидатами на эту роль оказываются убеждения о сознании. Поэтому я буду именовать подобные функциональные состояния менее нагруженным термином «суждения» и оставлю открытый вопрос, исчерпывается ли суждением о сознании убеждение о сознании. Мы можем представлять суждение в качестве того, что остается от убеждения после удаления любых связанных с ним феноменальных качеств.

Существование чисто психологических состояний, квалифицируемых в качестве таких суждений, не должно вызывать вопросов. Начать с того, что предрасположенность к верbalным отчетам определенного вида — это психологическое состояние; и мы по меньшей мере можем именовать «суждением» эту диспозицию. Кроме того, всякий раз, когда я формирую убеждение о моем сознательном опыте, это сопровождается разного рода функциональными процессами, как и в случае других убеждений. Эти процессы лежат в основе предрасположенности к вербальным отчетам, а также в основе различных диспозиций иного рода. Если кто-то убежден, что ЛСД вызывает причудливые ощущения цвета, сопровождающий это убеждение процесс может фундировать предрасположенность принимать ЛСД и в будущем или избегать этого и т. д. Мы можем использовать термин «суждение» как общее обозначение состояний и процессов, играющих каузальную роль, о которой идет речь. В первом приближении можно сказать, что некая система судит об истинности пропозиции, если она предрасположена утвердительно отвечать на вопрос о данной пропозиции, надлежаще вести себя в контексте ее других убеждений и желаний и т. д.

Суждения, не исключено, можно истолковать как то общее, что есть у меня и у моего зомбийного двойника. Мой зомбийный двойник

лишен какого-либо сознательного опыта, но он *высказывается* о том, что не лишен его; по крайней мере его детальные вербальные отчеты не отличаются от моих. При моем использовании термина «суждение» кажется естественным сказать, что мой зомбийный двойник выносит суждение о том, что у него есть сознательный опыт, и что его суждения в этом плане находятся в состоянии взаимно-однозначного соответствия с моими суждениями.

В конце этой главы я доказываю, что семантическое содержание моих феноменальных убеждений тонким образом отчасти конституируется самим сознательным опытом (к примеру, ощущения красного могут играть роль в конституировании содержания некоторых убеждений об ощущениях красного). Если это так, то содержательная сторона некоторых суждений зомби будет не столь богата, как содержательная сторона соответствующих убеждений, имеющихся у меня самого. Тем не менее они по меньшей мере будут находиться во взаимно-однозначном соответствии с моими суждениями, будут иметь ту же *форму* и будут *функционировать* точно таким же образом в плане своих поведенческих эффектов. Поэтому когда я говорю о суждении зомби о том, что у него есть ощущение красного, я рассуждаю о *чем-то* интересном в его психологии: по меньшей мере мои слова могут быть истолкованы как слова, дефляционным путем отсылающие к суждению, выражаемому им при помощи слов «у меня есть ощущение красного» (или слов с таким звучанием!). Аналогичным образом я буду говорить и о *высказываниях*, отвлекаясь от уточненных вопросов, которые связаны с их содержательной стороной.

Строго говоря, все описания высказываний и суждений о феноменальном в терминах их содержания (к примеру, отсылки к суждению о том, что кто-то имеет ощущение красного) должны быть истолкованы подобным дефляционистским способом. Феноменальные *убеждения* субъекта, несомненно, будут наделены всей полнотой атрибутируемого им содержания, но вопрос о содержании суждения не столь ясен именно потому, что неясно, какую роль играет сознание в конституировании содержания феноменального убеждения. Я не буду особо пользоваться этим различием на протяжении большей части этой главы, так как постараюсь сформулировать проблемы, которые ставят перед моей концепцией феноменальные суждения, так остро, как это только будет возможно. В конце этой главы я более детально рассмотрю эти вопросы о содержании.

Три типа феноменального суждения

Суждения, связанные с сознательным опытом, распадаются по меньшей мере на три группы. Я буду называть их феноменальными суждениями *первого порядка, второго порядка и третьего порядка*.

Обычно я буду отбрасывать уточняющую часть и говорить о «первопорядковых суждениях» и т. д., всегда подразумевая под этим феноменальные суждения.

Суждения *первого порядка* — это суждения, сопровождающие сознательные опытные переживания и касающиеся не опыта самого по себе, а *объекта опыта*. Когда у меня есть ощущение красного — к примеру, при взгляде на красную книгу, — оно обычно сопровождается явным или неявным суждением «тут есть что-то красное». Когда у меня есть опытное ощущение музыкального тона, его сопровождает психологическое состояние, связанное с этим музыкальным тоном. Кажется верным сказать, что любой объект, *сознательно переживаемый* в опыте, является также *когнитивно представленным*, хотя вопрос этим не исчерпывается. Любой сознательный опыт сопряжен с содержательным когнитивным состоянием. Это когнитивное состояние я и называю первопорядковым суждением. (Можно было бы попробовать доказать, что это состояние в некоторых отношениях отличается от убеждения или суждения, поскольку оно, к примеру, не должно приниматься в результате какой-то рефлексии. Более подробно я рассмотрю этот вопрос в следующей главе, но пока я буду говорить именно о «суждениях» — они таковы хотя бы в первом приближении.)

Содержание этих первопорядковых суждений мы можем трактовать как то, что составляет содержание *осведомленности*, где под осведомленностью понимается психологический коррелят сознания, упомянутый в главе 1: информация, относительно которой мы осведомлены, — это в общем информация, доступная когнитивной системе, доступная для верbalного отчета и т. п. Строго говоря, эти суждения не являются суждениями *о* сознании. Скорее, они *параллельны* сознанию и обычно представляют собой суждения *об* объектах и свойствах, существующих в окружении субъекта или даже в его голове. По сути, мы вправе сказать, что первопорядковое суждение — это суждение о том, что дано нам в соответствующем опыте. Там, где у меня есть опыт красной книги, имеется и соответствующее первопорядковое суждение о красной книге. В известном смысле мы можем поэтому сказать, что опыт и первопорядковые суждения — а поэтому также сознание и осведомленность — имеют общее содержание. В следующей главе я более детально объясню это отношение.

В этой главе меня в основном будут интересовать *второпорядковые* суждения. Они являются суждениями о сознательном опыте в более непосредственном плане. Когда у меня есть ощущение красного, я иногда отмечаю, что у меня есть ощущение красного. Я сужу о своей боли, о переживании определенных эмоциональных качеств и т. п. В общем, кажется, что относительно любого сознательного опыта верно, что если некто обладает соответствующими концептуальными ресурсами,

то он обладает хотя бы способностью судить о наличии у него соответствующего опыта.

О сознательных опытных переживаниях можно выносить и более детальные суждения. Человек может отмечать, что он ощущает особенно яркий оттенок фиолетового, или что его боль имеет всеохватный характер, или даже что зеленый остаточный образ — это третий подобный образ за сегодняшний день. Помимо суждений о конкретных сознательных переживаниях, второпорядковые суждения также включают суждения о конкретных *видах* сознательного опыта, когда, к примеру, кто-то отмечает, что определенный наркотик вызывает особенно интенсивные ощущения или что пощипывание перед чиханием является чем-то очень приятным.

Третьепорядковыми суждениями я буду называть суждения о сознательном опыте как некоем типе. Они выходят за пределы суждений о конкретных опытных переживаниях. Мы выносим третьепорядковые суждения прежде всего тогда, когда мы размышляем о том факте, что у нас имеются сознательные опытные переживания, и когда мы размышляем об их природе. На протяжении всего моего исследования я выносил третьепорядковые суждения. Типичным третьепорядковым суждением могло бы быть суждение «сознание загадочно, я не понимаю, как можно было бы редуктивно объяснить его». В числе других суждений такого рода — «сознательный опыт неуловим» и даже «сознательный опыт не существует».

Третьепорядковые суждения особенно распространены среди философов, а если говорить о философах — то среди тех, кто склонен рассуждать о тайнах опыта. Не исключено, что многие люди проживают жизнь, вообще не вынося каких-либо третьепорядковых суждений. Тем не менее подобные суждения характерны для значительного класса людей. Сам факт того, что люди выносят такие суждения, нуждается в объяснении.

Чтобы лучше запомнить эти различия, можно представить различные виды суждений, соотнесенных с сознанием, следующим образом:

- Первопорядковое суждение: *Это красное!*
- Второпорядковое суждение: *Сейчас у меня есть ощущение красного.*
- Третьепорядковое суждение: *Ощущения таинственны.*

2. Парадокс феноменального суждения

Существование феноменальных суждений вскрывает главную трудность нередуктивной теории сознания. И состоит она в следующем. Мы видели, что само сознание не может быть редуктивно объяснено.

Но феноменальные суждения относятся к области психологии и в принципе должны быть редуктивно объяснимы обычными методами когнитивной науки. Должно существовать физическое или функциональное объяснение, к примеру, того, почему мы предрасположены *высказываться о сознании так, как мы это делаем, а также того, каким образом мы выносим суждения о сознательном опыте*. А из этого следует, что наши высказывания и суждения о сознании могут быть объяснены в терминах, совершенно независимых от сознания. И даже более того: дело выглядит так, будто сознание *объяснимельно нерелевантно* для наших высказываний и суждений о сознании. Этот итог я называю *парадоксом феноменального суждения*.

Парадокс феноменального суждения, похоже, не привлекал к себе особого внимания, но он был живо изложен физиком Авшаломом Элитцуром (Elitzur 1989) как аргумент против концепций, в которых говорится о «пассивности» сознания; взамен он предлагает интеракционистский дуализм¹. Об этом парадоксе говорил и психолог Роджер Шепард (Shepard 1993), намекавший на то, что мы должны смириться с его существованием:

Одним словом, мы, похоже, по-прежнему остаемся с дилеммой: никакой анализ чисто физических процессов в мозге (или в компьютере), как кажется, не способен ухватить конкретного качества субъективного опыта, соответствующего этим процессам. Тем не менее анализ подобного рода наверняка должен быть в состоянии каузально объяснить, как индивид приходит к тому, чтобы высказывать мысли, подобные предыдущей. Быть может, нам придется смириться с тем, что, хотя как существование сознательных переживаний, так и отношения сходства между их квалиа имеют физическое воплощение с физическими причинами и действиями, сознательные переживания и сами квалиа не могут быть охарактеризованы как физические события и не могут сообщаться друг другу физическими системами (с. 242).

Как мы видели в предыдущей главе, вопрос о том, является ли сознание *каузально нерелевантным* для продуцирования поведения, — сложная метафизическая проблема, которую лучше оставить открытой. Но объяснительная *нерелевантность* сознания есть нечто более очевидное, и она связана со многими трудностями того же самого рода, что возникали бы в случае каузальной нерелевантности. Как бы ни обернулось дело с метафизикой каузальности, кажется достаточно ясным, что можно было бы дать такое физическое объяснение поведения, которое не отсылало бы к существованию сознания и не подразумевало бы его существование.

Когда в беседе я говорю, что «сознание — это самая таинственная вещь», это поведенческий акт. Когда я писал в одной из предыдущих глав, что «сознание не может быть редуктивно объяснено», это был

поведенческий акт. Когда я комментирую появление особенно интенсивных фиолетовых квалиа, переживаемых мной, это поведенческий акт. Как и все поведенческие акты, данные акты в принципе могут быть объяснены в терминах внутренней каузальной организации моей когнитивной системы. Можно рассказать некую историю о паттернах активации нейронов, которая объясnit осуществление этих актов; а на более высоком уровне, возможно, можно было бы рассказать историю о когнитивных репрезентациях и их высокоуровневых отношениях, выполняющих соответствующую объяснительную работу. В настоящее время мы, конечно, не знаем деталей этого объяснения, но если область физического каузально замкнута, то какое-то редуктивное объяснение в физических и функциональных терминах будет существовать.

При этом объяснении моих высказываний в физических или функциональных терминах нам никогда не нужно будет ссылаться на существование самого сознательного опыта. Физическое или функциональное объяснение будет независимым и в равной степени применимым как к зомби, так и к по-настоящему сознательному субъекту опыта. Кажется поэтому, что сознательный опыт нерелевантен для объяснения высказываний о феноменальном и феноменальных суждений, несмотря на то что подобные высказывания и суждения напрямую относятся к сознательному опыту!

Один из путей сопротивления этому выводу состоял бы в попытке доказать, что *содержательная сторона* моих высказываний и суждений о феноменальном не может быть во всей ее полноте редуктивно объяснена, так как сознание играет какую-то роль в конституировании данного содержания. Можно было бы, к примеру, попробовать доказать, что высказывания и убеждения зомби — это *другие* высказывания и убеждения (хотя они выглядят и звучат точно так же!), потому что зомби не располагал бы полным понятием сознания. Но будет оставаться по меньшей мере странным, что сознание должно быть нерелевантным по отношению к тем *звукам*, которые мы издаем, говоря о сознании, к моим движениям пальцев в данный момент и т. п.; поэтому такой ответ не может полностью устраниТЬ данную трудность. Так что пока я оставлю в стороне этот подход и продолжу размышлять о высказываниях и суждениях в «дефляционистском» ключе, допускающем возможность их редуктивного объяснения.

Сопротивление указанному выводу можно было бы оказать и иным способом — доказывая, что для *любого* высокоуровневого свойства, относительно которого можно было бы думать, что оно релевантно для объяснения, можно найти такое низкоуровневое объяснение, которое не ссылается на существование данного свойства. Можно было бы

утверждать, что объяснительно нерелевантным является такое психологическое свойство, как память, так как можно дать такие нейрофизиологические объяснения действий, при которых вообще не упоминается память; можно даже было бы попытаться показать объяснительную нерелевантность температуры в физике, так как объяснительные ссылки на температуру могут быть в принципе заменены объяснением на молекулярном уровне (Ким (Kim 1989) называет это проблемой *объяснительного исключения*). И это могло бы означать, что сознание в плане своей объяснительной нерелевантности ничуть не хуже любых других высокоуровневых свойств. Если сознание сходно с памятью и температурой, то это совсем не плохая компания.

Мы видели, однако, что все высокоуровневые свойства, такие как температура и память, логические супервентны на физическом. Из этого следует, что когда мы объясняем некое действие в нейрофизиологических терминах, это не делает память чем-то объяснительно нерелевантным. Память может *унаследовать* объяснительную релевантность вследствие имеющегося у нее логически супервентностного статуса. Когда мы объясняем желание какого-то мужчины найти себе женщину, ссылаясь на то, что он мужчина и не женат, это не делает объяснительно нерелевантным тот факт, что он холостяк! Общим принципом здесь является то, что, когда два множества свойств *концептуально* соотнесены, существование объяснения в терминах одного множества не делает другое множество объяснительно нерелевантным. В известном смысле одно из объяснений может быть своего рода пересказом другого — вследствие указанного концептуального отношения между соответствующими терминами.

Когда мы рассуждаем о взаимодействии наших воспоминаний, в определенном смысле мы пересказываем физическую историю на более высоком уровне абстракции. На этом более высоком уровне мы опускаем многие детали физической истории и поэтому зачастую даем гораздо более удовлетворительное объяснение (все те детали могут быть нерелевантным шумом), но этот уровень тем не менее логически соотнесен с низкоуровневой историей. Это же верно и для температуры. Существование низкоуровнего объяснения столь же мало делает такие высокоуровневые свойства объяснительно нерелевантными, как молекулярные процессы в бильярдном шаре делают нерелевантным быстроту его движения. В общем, высокоуровневые свойства такого рода будут более экономными *переописаниями* того, что описывается низкоуровневым объяснением. Можно было бы сказать, что зачастую низкоуровневое описание *имплицитно* — пусть и не эксплицитно — будет даже содержать высокоуровневые свойства из-за их логически супервентностного статуса. Там, где есть логическая супервентность, не существует проблемы объяснительной нерелевантности.

Проблемы, связанные с сознанием, гораздо более серьезны. Сознание не является логически супервентным на физическом, и поэтому мы не можем утверждать, что физическое или функциональное объяснение имплицитно предполагает сознание или что сознание наследует объясняющую релевантность благодаря логической супервентности на свойствах, используемых при таком объяснении. Физическое или функциональное объяснение поведения независимо от сознания в гораздо более сильном смысле. Подобное объяснение может быть осуществлено в терминах, которые даже и не *предполагают* существование сознательного опыта. Сознание концептуально независимо от того, что используется при объяснении наших высказываний и суждений о сознании.

Это не означает, что мы *вообще* не можем обращаться к сознательному опыту при объяснении поведения. Совершенно оправданным было бы объяснение того факта, что кто-то отшатывается от огня, ссылаясь на то, что он испытал боль. В конце концов, даже нередуктивная концепция не отрицает существования закономерных регулярностей между опытом и последующим поведением. Подобные регулярности, однако, в итоге зависят от регулярностей на физическом уровне. С любым объяснением поведения, обращающимся к ощущению боли, скоррелировано и более фундаментальное объяснение в чисто физических/функциональных терминах — возможно, в терминах психологической боли или восприятия боли, — которое не ссылается на какие-либо свойства опыта и не предполагает их. Опыт обретает некую косвенную объясняющую релевантность благодаря его номической связи с теми физическими и функциональными процессами, но все же остается избыточным относительно базового объяснения.

Чтобы с особой силой прочувствовать эту проблему, представьте моего зомбийного двойника, живущего в соседнем универсуме. Он постоянно рассуждает о сознательном опыте — кажется даже, что он зациклен на нем. Он не отходит от своего компьютера, набирая главу за главой о тайнах сознания. Часто он отпускает комментарии о том, как ему приятно переживать какие-то чувственные квалиа, признаваясь в особой любви к насыщенным зеленым и фиолетовым оттенкам. Нередко он вступает в полемику с зомбийными материалистами, доказывая, что их позиция не может отдавать должное реальностям сознательного опыта.

Но при этом у него *вообще* нет сознательного опыта! В его универсуме правда на стороне материалистов, а он ошибается. Большинство его высказываний о сознательном опыте абсолютно неверно. Несомненно, однако, что существует физическое или функциональное объяснение того, почему он говорит то, что говорит. В конце концов, его

универсум безраздельно управляет законами, и там не происходит чудес — так что его высказывания должны иметь *какое-то* объяснение. Причем это объяснение в конечном счете должно быть дано в терминах физических процессов и законов, так как *только* такие процессы и законы имеются в его универсуме.

(Как и прежде, можно было бы по праву доказывать, что, используя слово «сознание», зомби не отсылает к сознанию в полном смысле. Но пока рассуждения о высказываниях и суждениях зомби о сознании должны пониматься в указанном выше дефляционном ключе. И даже если у него нет такого полного понятия, несомненно, что он выносит суждение о том, что у него имеется *какое-то* свойство, выходящее за пределы его структурных и функциональных свойств, — свойство, которое он именует «сознанием», — и проблема с неменьшей остротой возобновляется в этом виде.)

Конечно, мой зомбийный двойник — это всего лишь логическая, а не эмпирическая возможность, и нам не стоит *слишком* переживать по поводу тех странных вещей, которые случаются в логически возможных мирах. И все же повод для беспокойства здесь есть. В конце концов, любое объяснение поведения моего двойника будет вместе с тем объяснением и *моего* поведения, так как процессы, протекающие в его теле, точно соответствуют процессам, которые протекают в моем. Объяснение *его* высказываний очевидным образом не зависит от существования сознания, поскольку в его мире вообще нет сознания. И из этого следует, что объяснение моих высказываний также не зависит от существования сознания.

Чтобы усилить ощущение парадокса, отметим, что сам мой зомбийный двойник предается точно таким же рассуждениям. Известно, что он сокрушался по поводу судьбы *своего* зомбийного двойника, тратящеего всё свое время на разрешение проблем, связанных с сознанием, хотя его у него и нет. Он озабочен тем, что тот должен говорить об объяснительной нерелевантности сознания в своем собственном универсуме. При этом он остается полностью уверенным в существовании сознания и в невозможности дать ему редуктивное объяснение. Но в его случае всё это большая иллюзия. В его универсуме *нет* сознания — и правда всецело на стороне элиминативистов. Несмотря на тот факт, что его когнитивные механизмы функционируют так же, как и мои, *его* суждения о сознании абсолютно неверны.

Эта парадоксальная ситуация вызывает и восхищение, и беспокойство. Она не является очевидным образом фатальной для нередуктивной позиции, но мы тем не менее должны как-то отреагировать на нее. Это, несомненно, главная трудность, с которой сталкивается нередуктивная теория, и если подобная теория по крайней мере не готова

к встрече с ней, то она не может быть полностью удовлетворительной. Мы должны тщательно исследовать последствия этой ситуации и отделить то, что является всего лишь континтуитивным, от того, что несет угрозу самому существованию нередуктивного представления о сознании.

Ницше сказал: «То, что не убивает нас, делает нас сильнее». Если мы сможем справиться с этим парадоксом, мы, быть может, узнаем нечто важное об отношении между сознанием и познанием. Я посвящу оставшуюся часть данной главы разрешению этого парадокса; что же касается смежных вопросов относительно связи между сознанием и познанием, то к ним мы будем возвращаться в последующих главах. Таким путем мы сможем поместить теорию сознания на гораздо более надежный фундамент.

(Кто-то мог бы счесть, что от этого парадокса можно избавиться, опираясь на то, что выше я назвал позицией В-типа, согласно которой сознание оказывается супервентным с метафизической, но не концептуальной необходимостью, или на взгляды С'-типа, в соответствии с которыми феноменальные свойства конституируют внутреннюю природу физического. Но парадокс почти с такой же силой возникает и в случае принятия этих воззрений. Даже если они и спасают некую каузальную релевантность сознания, они по-прежнему ведут к объяснительной нерелевантности, так как объяснительная релевантность должна опираться на концептуальные связи. Даже в соответствии с этими воззрениями можно давать редуктивное объяснение феноменальных суждений, но не самого сознания, делая тем самым сознание объяснительно нерелевантным для таких суждений. Можно было бы найти такое процессуальное объяснение этих суждений, которое ни на какой своей стадии не отсылало бы к существованию опыта и не предполагало бы его существование; присутствие какой-то дополнительной «метафизически необходимой» связи или внутренних феноменальных свойств будет с концептуальной стороны совершенно независимым от того, что используется при объяснении поведения.)

Убедиться в этом можно и иным способом: в соответствии с подобными воззрениями, зомби все же представимы и их поведение будет в полной мере объяснимым. Поскольку это объяснение применимо к зомби, существование сознания не будет играть в нем существенной роли. Но то, что происходит с зомби, происходит и с нами, так что такое объяснение будет в равной степени применимо и к нам. Поэтому даже при указанных взглядах будет иметься такое объяснение наших феноменальных суждений, при котором сознание окажется чем-то совершенно избыточным.)

При объяснении *большей части* нашего поведения факт объяснительной нерелевантности сознания может быть хоть и континтуитивным, но не таким уж и парадоксальным. Для объяснения того, что я потянулся за книгой передо мной, не нужно ссылаться на мое феноменальное *ощущение* этой книги; достаточно вместо этого сослаться на мое *восприятие*. Когда любитель музыки вздыхает при каком-то особенно изысканном переходе, можно было бы подумать, что главным для объяснения его поведения могло бы быть переживаемое им качество звуковых ощущений, но в действительности объяснение может быть дано исключительно в терминах звукового восприятия и функциональных реакций. Даже при объяснении того, почему я убираю руку от огня, может оказаться достаточно функционального объяснения в терминах психологического понятия боли.

В общем, оказывается, что там, где, как можно было бы подумать, мы нуждаемся в отсылке к феноменальным свойствам для объяснения поведения, обычно можно ограничиться психологическими свойствами. В главе 1 мы видели, что в основе каждого феноменального состояния лежит некое психологическое состояние. Там, где можно было бы сослаться на ощущение, мы ссылаемся на перцептивные регистрации; там, где можно было бы сослаться на феноменальное качество эмоции, мы ссылаемся на соответствующее функциональное состояние; там, где можно было бы сослаться на мысль, нам нужно лишь сослаться на содержание этой мысли. Именно это соответствие между феноменальными и психологическими свойствами делает объяснительную нерелевантность феноменальных свойств в целом *не очень* серьезной проблемой. Вначале это кажется континтуитивным, но и только. По крайней мере в случае поведения, напрямую не связанного с сознательным опытом, дело представляется таким образом, что нет особой нужды ссылаться при объяснении на феноменальные свойства.

Но объяснительная нерелевантность сознательного опыта начинает вызывать беспокойство, когда речь заходит о наших высказываниях и суждениях о сознании. Объяснительная нерелевантность сознания по отношению к *первопорядковым* феноменальным суждениям, таким как «это красная вещь», впрочем, может и не вызывать особой тревоги. Кажется вполне разумным, что они должны объясняться исключительно в терминах восприятия и других психологических процессов; в конце концов, такие суждения напрямую связаны не с сознательным опытом, а с состоянием мира. Однако в случае второпорядковых и третьепорядковых феноменальных суждений объяснительная нерелевантность, похоже, ставит перед нами реальные проблемы. Эти суждения являются суждениями о сознательном опыте, и именно они ответственны за наши рассуждения о наших ощущениях и за беспокойство

**Навстречу
парадоксу**

философов относительно тайн сознания. Одно дело — признавать, что сознание нерелевантно для объяснения того, как я расхаживаю по комнате, другое — что оно нерелевантно для объяснения того, почему я рассуждаю о сознании. Мы наверняка были бы склонны думать, что факт того, что я наделен сознанием, будет частью объяснения того, почему я говорю, что я наделен сознанием, или почему я сужу, что наделен сознанием; и тем не менее кажется, что это не так.

В самом деле, объяснение того, почему мы говорим и судим о том, что где-то есть вода, будет включать в качестве своей части тот факт, что в этом месте в действительности есть вода. Аналогичным образом кажется, что существование звезд и планет почти наверняка является объяснимым релевантным для наших суждений о существовании звезд и планет. Правило таково, что когда мы выносим истинное и достоверное суждение о том, что P , тот факт, что P является истинным, обычно играет центральную роль при объяснении данного суждения. Существуют суждения, объекты которых объяснимы нерелевантны для самих суждений. Таковы, к примеру, религиозные убеждения или убеждения относительно НЛО; похоже, что их можно объяснить без ссылок на богов или НЛО. Но вполне вероятно, что это ложные убеждения, и они не могут рассматриваться в качестве очевидных примеров знания. По контрасту с этим мы знаем, что обладаем сознанием.

Здесь мы оказываемся в непростой ситуации: как знание о сознании может быть согласовано с тем фактом, что сознание объяснимо нерелевантно для феноменальных суждений? Если феноменальные суждения возникают по причинам, независимым от самого сознания, не означает ли это, что они являются необоснованными? В этом, в конце концов, и состоит главная трудность, связанная с парадоксом феноменального суждения, и позже в этой главе я сделаю ее предметом подробного рассмотрения.

Данный парадокс является следствием тех фактов, что (1) область физического каузально замкнута; (2) суждения о сознании логически супервентны на физическом; (3) сознание не является логически супервентным на физическом; и (4) мы знаем о своем сознании. Из посылок (1) и (2) следует, что суждения о сознании могут быть редуктивно объяснены. В сочетании с посылкой (3) это означает, что сознание объяснимо нерелевантно для наших суждений, что не стыкуется с посылкой (4). Так и возникает парадокс. Можно было бы попробовать избежать его, отрицая одну из посылок. Я кратко рассмотрю каждый из этих спасительных путей.

Некоторые дуалисты будут отрицать посылку (1). Традиционно картезианский интеракционистский дуализм мотивировался той идеей, что только так можно отдать должное релевантности сознания для действий. В самом деле, Элитцур (Elitzur 1989) прямо заключает от су-

ществования высказываний о сознании к тому, что законы физики не могут быть исчерпывающими и что сознание играет активную роль в управлении физическими процессами (он допускает, что второй закон термодинамики мог бы оказаться ложным). Но я уже доказывал, что интеракционистский дуализм мало что дает для избежания проблемы объяснительной нерелевантности.

Кто-то мог бы соблазниться отрицанием посылки (2), но вспомним, что мы *определили* суждения как функциональные состояния, логически супервентные на физическом. Можно было бы, правда, попробовать доказать, что функциональных состояний, хотя бы отдаленно напоминающих то, как мы представляем себе суждение, вовсе не существует. Но даже если это и так, мы можем просто отступить к *высказываниям* о сознании, являющимся поведенческими актами и поэтому логически супервентным более непосредственным образом — и ставящим перед нами почти столь же серьезные проблемы. Даже если бы кто-то заявил, что поведенческие акты не являются исключительно физическими (они могли бы попытаться доказать, что сознательный опыт требуется для того, чтобы нечто могло считаться высказыванием, а не просто шумом, или считаться высказыванием о сознании), все равно удивительно, что сознание объяснительно нерелевантно по отношению к производимым нами таким звукам и записям, которые могут быть систематически истолкованы как то, что имеет отношение к сознанию. Так что аналогичные проблемы возникнут вне зависимости от того, как мы определим соответствующие состояния. Тем не менее подобные соображения могут играть по крайней мере вспомогательную роль при разрешении данного парадокса, так как, похоже, именно *убеждения*, а не *высказывания*, наиболее тесно связаны с знанием, и какое-то содержание феноменального убеждения может быть конституировано самим опытом. Позже в этой главе я еще вернусь к данному вопросу.

Редукционисты и элиминативисты, конечно, будут отрицать посылку (3) или (4). Я уже привел исчерпывающие аргументы в пользу (3), так что я не буду повторять их здесь. Отрицание посылки (4) аналогичным образом ведет к элиминативизму — позиции, которую я уже отверг. Тем не менее вскоре я рассмотрю вопрос о том, как редукционист мог бы воспользоваться парадоксом феноменального суждения.

Думаю, что самой разумной установкой в данном случае было бы признать, что все перечисленные посылки, скорее всего, верны; и посмотреть, как их можно согласовать друг с другом. Мы знаем о существовании сознательного опыта; физическое почти наверняка каузально замкнуто; и ранее мы выяснили, что сознание не является логически супервентным на физическом. Фокус в том, чтобы научиться жить с такой комбинацией.

3. Объяснение феноменальных суждений

Учитывая сказанное ранее, объяснение того, почему мы говорим о сознании то, что мы фактически говорим о нем, кажется серьезным и интересным проектом для когнитивной науки. Эти высказывания — поведенческие акты, и они должны быть доступны для объяснения не меньше, чем любые другие поведенческие акты. И любой когнитивный ученый, вступающий на этот путь, вправе рассчитывать на богатую добычу. Объяснение наших высказываний и суждений о сознании может быть трудным делом, но оно не будет столь же трудным, как объяснение самого сознания. Конечно, такое объяснение не будет автоматически давать объяснение сознания, но оно может направить нас в верную сторону.

Мы можем не ограничиваться допущением возможности подобного объяснения в качестве интеллектуального вывода из каузальной замкнутости физического и логической супервентности поведения. Существуют и независимые основания считать, что феноменальные суждения окажутся естественными спутниками определенных когнитивных процессов, и, поразмышляв, мы увидим, что можно даже *ожидать* подобных суждений от системы, обладающей интеллектом и устроенной определенным образом. Если это так, то объяснение таких высказываний и суждений может быть не столь трудным, как можно было бы подумать; они могут оказаться следствием неких основополагающих принципов когнитивного устройства.

Здесь я дам лишь очень краткий ответ на вопрос о том, почему можно было бы думать подобным образом; более детально я рассмотрю его в главе 8. Чтобы прочувствовать ситуацию, вообразим, что мы создали компьютерный интеллект в виде автономного агента, который воспринимает свое окружение и способен рационально обдумывать воспринятое. Какой будет подобная система? Будет ли у нее какое-то понятие сознания или какие-либо иные понятия, родственные последнему?

Чтобы удостовериться в этой возможности, отметим, что при самом естественном устройстве такой системы она наверняка имела бы понятие о себе — к примеру, она могла бы отличать себя от остального мира и от других сущностей, похожих на нее. Кажется также вполне правдоподобным, что она могла бы иметь гораздо более прямой доступ к ее собственным когнитивным ресурсам, чем к ресурсам других систем. Если бы она была способна размышлять, то она, вероятно, была бы напрямую осведомлена о содержании своих собственных мыслей и могла рассуждать об этом факте. Более того, такая система, скорее всего, имела бы прямой доступ к перцептивной информации наподобие нашей когнитивной системы.

Если бы мы спросили такую систему, каким было ее восприятие, что она ответила бы? Сказала бы она «никаким»? Могла ли она сказать: «Да, я знаю, что здесь есть красный трехколесный велосипед, но у меня нет представления о том, *как* я это знаю; эта информация просто появилась в моей базе данных»? Не исключено, но вряд ли. Система, устроенная подобным образом, была бы весьма неэффективной и неестественной; ее доступ к ее собственным перцептивным содержаниям был бы до странности опосредованным. Кажется гораздо более вероятным, что она ответила бы: «Я знаю, что здесь находится красный трехколесный велосипед, потому что я *вижу* его здесь». А если мы спросим, как она знает, что она видит этот велосипед, ответ, скорее всего, будет таким: «Я просто *вижу* его, и всё».

Система, которая ответила бы: «Я знаю, что я *вижу* его, потому что сенсоры 78–84 активированы таким-то и таким-то образом», была бы странной. Как указывает Хоффштадтер (Hofstadter 1979), нет необходимости обеспечивать системе столь детальный доступ к ее низкоуровневым частям. Даже программа Т. Винограда SHRDLU (1972) не знала о коде, на котором была написана, несмотря на то что она могла воспринимать виртуальный мир, делать выводы об этом мире и даже в определенной степени обосновывать свое знание. Подобное дополнительное знаниеказалось бы совершенно необязательным и лишь осложнило бы процессы, ведущие к ее осведомленности о чем-то или формирующие делаемые ею выводы.

Так что представляется вероятным, что подобная система имела бы схожую с нами установку относительно ее перцептивных данных, знание которых было бы у нее прямым и непосредственным, по крайней мере в той степени, в какой затрагивается сама эта система. Если бы мы спросили, как она знает о том, что она видит красный трехколесный велосипед, эффективно сконструированная система сказала бы: «Я просто *вижу* его!» Если бы мы спросили, как она знает, что этот велосипед красный, она ответила бы примерно так, как отвечаем и мы: «Он просто *выглядит* красным». Если бы подобная система обладала способностью к размышлению, она могла бы заинтересоваться вопросом о том, почему вещи кажутся красными, и о том, почему красное выглядит *именно так*, а синее вот так — иначе. С позиции самой системы это просто изначальный факт, что красное и синее выглядят так, как они выглядят. Разумеется, с нашей более высокой позиции мы знаем, что дело лишь в том, что красное приводит систему в одно состояние, а синее — в другое; но с позиции машины это не помогает.

По мере своих размышлений она могла бы задуматься о самом факте того, что она, как кажется, имеет доступ к тому, о чем она думает, и что у нее есть чувство самой себя. Рефлексивная машина, устроенная так, чтобы иметь прямой доступ к содержанию своих восприятий и мыслей,

очень скоро могла бы озабочиться тайнами сознания (Хофштадтер (Hofstadter 1985a) обстоятельно обсуждает эту идею): «Почему этот жар *чувствуется* именно так?»; «Почему я есть я, а не кто-то другой?»; «Я знаю, что процессы, идущие во мне, — это лишь движение тока по электрическим цепям, но как это объясняет мое *переживание* мысли и восприятия?»

Разумеется, не стоит так уж всерьез относиться к тем спекуляциям, которым я здесь предавался, но они помогают осознать *естественность* факта наших суждений и высказываний о том, что мы обладаем сознанием, при соответствующем рациональном устройстве нас самих. Странной была бы когнитивная система, лишенная какого бы то ни было представления о том, о чем идет речь, когда мы спрашиваем, каково это — быть *ею*. Тот факт, что мы мыслим и говорим о сознании, может быть следствием совершенно естественных характеристик нашего устройства — так же как в этих системах. И при объяснении того, почему эти системы мыслят и говорят так, как они делают это, нам точно никогда не потребуется ссылка на *сознание*, если говорить о нем в полном смысле этого слова. Быть может, эти системы действительно обладают сознанием, быть может — нет, но объяснение работает независимо от этого факта. Все объяснения, касающиеся функционирования этих систем, могут быть даны исключительно в вычислительных терминах. И очевидно, что в таком случае привидение в машине не может играть какой-то объяснительной роли.

Всё это и означает (развивая тезис из главы 1), что, хотя сознание и является чем-то удивительным, высказывания о сознании не таковы. Хотя сознание и является такой чертой мира, существование которой мы не смогли бы предсказать исходя из физических фактов, вещи, которые мы говорим о сознании, — совершенно заурядный когнитивный феномен. Тот, кто обладал бы достаточным знанием о когнитивной структуре, смог бы напрямую предсказать вероятность таких высказываний, как «я *ощущаю* себя сознательным так, как не смог бы ни один физический объект», или даже декартовского «*cogito ego sum*». В принципе, редуктивное объяснение в терминах внутренних процессов должно было бы делать высказывания о сознании, по сути, не более удивительными, чем любой другой аспект поведения. Выше я указал, как можно было бы дать подобное объяснение, и я более детально рассмотрю этот вопрос в одной из следующих глав.

Позже мы увидим, что детали надлежащего объяснения могут оказаться очень полезными как отправная точка для теории сознания. Впрочем, отношение между объяснением феноменальных суждений и объяснением сознания — отношение весьма тонкого свойства. И прежде чем продолжить, я рассмотрю не столь утонченную реакцию на нашу ситуацию.

В этот момент многим читателям, особенно склонным к редукционизму, возможно, пришла в голову естественная мысль: если мы объяснили, почему мы *говорим*, что наделены сознанием, и почему мы *выносим суждения* о том, что наделены сознанием, то не объяснили ли мы всего, что должно быть объяснено? Почему попросту не прекратить наши поиски теории сознания, объявив само сознание химерой? Еще лучше: почему не объявить теорию наших суждений о собственном сознании полноценной теорией сознания? Вполне можно было бы предположить, что теория наших суждений — это всё, что нам нужно для теории сознания.

Достаточно ли объяснения суждений?

Эта позиция может быть подкреплена рядом соображений относительно суждений, формируемых в других областях. Можно было бы подумать, что широко распространенная вера в богов, которую можно обнаружить в самых разных культурах, дает нам самое серьезное основание полагать, что боги существуют. Есть, однако, и альтернативное объяснение этой веры — в терминах общественных и психологических факторов. Атеисты могли бы сослаться на психологическую неуверенность людей перед лицом космоса, на потребность в общей отдушине для духовной и эмоциональной экспрессии и на самовоспроизводящуюся по самой своей сути природу определенных систем идей для объяснения неизбежности широкой распространенности религиозных верований при условии тех природных характеристик и обстоятельств, которыми мы обладаем и в которых находимся. Можно даже указать на существование ряда весьма правдоподобных, но при этом ошибочных аргументов в пользу существования Бога, таких как аргумент от целесообразности и космологические аргументы. Хотя эти аргументы и ошибочны, они *не очевидно* таковы (в частности, аргумент от целесообразности мог по праву считаться убедительным в додарвиновские времена), и нетрудно понять, почему они должны были в целом укреплять мысль о естественности религиозной веры.

Представление о том, что широко распространенные религиозные верования могли бы быть объяснены подобным образом, без отсылки к существованию каких-то богов, обычно считается дополнительным свидетельством в пользу того, что богов в действительности не существует. Согласно этой интерпретации, атеистическая гипотеза не только может не хуже теистической объяснить сложное устройство природы; она может также объяснить причины столь большой популярности теистической гипотезы. Это действенный способ подорвать фундамент противоположного воззрения. В случае религиозной веры данный аргумент выглядит очень сильным. Он выглядит заманчиво и тогда, когда речь идет о сознании.

Нет сомнений, что это самый мощный аргумент в пользу редуктивного или элиминативистского взгляда на сознание. Но он недостаточен. Аналогия не проходит. Объяснение наших суждений о сознании и близко не подходит к устранению тайн сознания. Почему? Потому что само сознание является *тем, что нуждается в объяснении*. Гипотеза о существовании Бога выдвигалась, похоже, главным образом для объяснения разного рода очевидных фактов о нашем мире, таких как его упорядоченность и его кажущаяся целесообразность. Когда выяснилось, что эти данные могут столь же успешно объясняться альтернативной гипотезой, отпала необходимость выдвигать гипотезу о Боге. Нет такого особого феномена, как *Бог*, на который мы могли бы указать и сказать, что он нуждается в объяснении. В лучшем случае здесь можно говорить лишь о косвенных свидетельствах². Сходным образом идея о существовании НЛО часто выдвигается для объяснения странных небесных явлений, отметин на земле, исчезновений в Бермудском треугольнике, заявлений людей, «выживших» после встречи с НЛО, и т. п. Если эти данные могут быть объяснены без допущения существования НЛО, то мы лишаемся основания верить в НЛО.

Сознание, однако, не является объяснительным конструктом, вводимым для содействия объяснению поведения или каких-то событий в мире. Скорее, это исходный экспланандум, самостоятельный феномен, нуждающийся в объяснении. Не имеет поэтому значения то, что, как может оказаться, сознание вообще не требуется для объяснения других феноменов. Ведь наши свидетельства о сознании и не были связаны с этими другими феноменами. Даже если наши суждения о сознании получают редуктивное объяснение, *всё*, что из этого следует, — так это то, что наши суждения могут получить редуктивное объяснение. Проблема «сознание — тело» состоит не в том, чтобы объяснить наши суждения о сознании. Если бы это было так, она была бы весьма тривиальной проблемой. Скорее, она связана с объяснением самого сознания. Если указанные суждения могут быть объяснены без объяснения сознания, это интересно, а может и удивительно, но данное обстоятельство не устраивает проблему «сознание — тело».

Позицию, согласно которой нам достаточно объяснить наши суждения о сознании (так же как достаточно объяснить наши суждения о Боге), естественнее всего истолковать в качестве элиминативистской позиции относительно сознания (аналогичной элиминативистской позиции относительно Бога). В качестве таковой она сталкивается со всеми теми проблемами, с которыми естественным образом сталкивается элиминативизм. В частности, она предполагает отрижение свидетельств нашего собственного опыта. Такие вещи могут делать только философы — или те, кто запутался в интеллектуальных узлах.

Наши опытные переживания красного не исчезают после такого отрицания. Сохраняется нечто, выражющее то, каково это — быть нами, и это нечто по-прежнему нуждается в объяснении. Отбрасывать само сознание в результате парадокса феноменального суждения означало бы выплеснуть вместе с водой самого ребенка.

В позиции, указывающей на достаточность объяснения феноменальных суждений, есть определенная интеллектуальная привлекательность. Она кажется смелым шагом, четко разрешающим все проблемы и выставляющим повергнутую путаницу на всеобщее обозрение. Тем не менее уже через полминуты подобное «решение» утрачивает убедительность. Завершив наши размышления, мы осознаем, что объяснили лишь определенные аспекты нашего поведения. Мы объяснили, почему мы говорим то-то и то-то и почему мы расположены к этому, но мы даже не прикоснулись к главной проблеме, а именно к самому сознательному опыту. По истечении тридцати секунд мы обнаруживаем, что смотрим на красную розу, вдыхаем ее аромат и спрашиваем: «Почему я переживаю это *именно так?*» И мы осознаем, что совершенно не объяснили этого.

Если не признавать эту позицию разновидностью элиминативизма, то ее можно было бы попробовать истолковать как вариацию функционалистского воззрения, в соответствии с которым понятие сознания конструируется как «нечто ответственное за суждения о сознании». Но эта дефиниция столь же неадекватна, как и любые другие функциональные дефиниции сознания. Вне зависимости от того, является ли сознание *фактически* ответственным за суждения о сознании, это не кажется концептуальной истиной. В конце концов, хотя бы *логически* возможна ситуация — неважно, правдоподобна она или нет, — при которой можно было бы объяснить указанные суждения, не объяснив при этом сознания; и этого достаточно для того, чтобы показать ложность подобной конструкции сознания.

Есть и другие вариации аргументов такого рода. К примеру, можно было бы попробовать доказать, что можно дать сугубо редуктивное объяснение того, почему я считаю, что сознание нельзя редуктивно объяснить, почему я думаю, что сознание не является логически супервентным, или того, почему, на мой взгляд, оно не может быть функционально определено. Мы могли бы даже редуктивно объяснить, почему я считаю сознательный опыт экспланандумом. Можно подумать, что это полностью подрывает мои аргументы в предыдущих разделах, открывая путь к редуктивному воззрению на сознание. Но опять же такой взгляд может нравиться лишь в пылу интеллектуального сражения. А в итоге нам все-таки придется объяснять, почему существование в качестве интеллектуального агента переживается

именно так. Объяснение поведения или какой-то каузальной роли — это объяснение не того, что нужно объяснить. Это могло бы показаться ослиным упрямством, но оно основано на простом принципе: наши теории должны объяснять то, что взвывает к объяснению.

Эта линия аргументации — возможно, самая интересная из тех, что доступны редукционисту или элиминативисту. Если бы я был редукционистом, я был бы редукционистом именно такого рода. Но в итоге она сталкивается с той же проблемой, с которой сталкиваются и другие подобные воззрения: она не объясняет то, что нуждается в объяснении. Несмотря на всю заманчивость этой позиции, в конечном счете она не принимает проблему всерьез. Нельзя так просто решить загадку сознания³.

Деннет о феноменальных суждениях

Одним из сторонников взгляда, что нам нужно лишь объяснить наши суждения о сознании, является Дэниел Деннет. В статье 1979 г. он пишет:

Я считаю, что подобные суждения *исчерпывают* наше непосредственное сознание, что наш индивидуальный поток сознания не содержит ничего, кроме таких пропозициональных эпизодов, или лучше: что такие потоки сознания, состоящие исключительно из подобных пропозициональных эпизодов, есть та реальность, которая инспирирует множество неверных описаний, которые выдаются за теории сознания, как дилетантские, так и академические... Мой взгляд, грубо говоря, состоит в том, что наши отчеты не соотносятся ни с каким феноменальным многообразием. Существуют публичные отчеты, которые мы даем, существуют эпизоды нашей пропозициональной осведомленности, наши суждения, и все; за ними — если говорить об интроспекции — лишь темнота (Dennett 1979, с. 95).

На этот счет я могу лишь заметить, что интроспекция Деннета очень сильно отличается от моей. В ходе собственной интроспекции я наблюдаю ощущения, болевые и эмоциональные переживания и много других разных вещей, которые, хотя и *сопровождаются* суждениями, не являются *только* суждениями — если только не *переопределить* понятие суждения или «эпизодов нашей пропозициональной осведомленности» таким образом, чтобы они включали подобные переживания. Если так переопределить данные термины, то позиция Деннета будет вполне основательной, но у нас больше не будет причин считать, что наши суждения могут быть редуктивно объяснены. Если же сконструировать суждения в качестве функционально индивидуализированных состояний, таких как предрасположенность к тому, чтобы давать отчет, — как, я думаю, и хочет сделать Деннет, — то его тезис становится неправдоподобным. Он попросту состоит в отрицании данных, которые должна объяснять теория сознания.

Что могло бы стоять за заявлениями, что интроспекция показывает только суждения? Быть может, Деннет — зомби⁴. Быть может, он вкладывает необычный смысл в слово «суждение». Скорее всего, однако, он принял за интроспекцию что-то другое — то, что можно было бы назвать *экстроспекцией*, процесс наблюдения за своими собственными когнитивными механизмами, так сказать, «со стороны» и размышления над происходящим. Наблюдая за чьими-то *механизмами*, легко прийти к заключению, что вся работа выполняется суждениями. При соответствующих когнитивных процессах происходят лишь многочисленные категоризации, различия и реакции. Вполне вероятно, что процессы, связанные с моим восприятием желтого объекта, могут быть полностью объяснены в терминах определенной чувствительности сетчатки, трансформации во внутренние представления, а также категоризации и именования этих представлений. Но это не объясняет содержательную сторону интроспекции, а объясняет лишь *процессы*, протекающие в ее ходе. Экстроспекция — это не интроспекция, хотя нетрудно понять, как философ, склонный к размышлению о своих внутренних механизмах, мог бы принять одно за другое. Сознательный опыт остается незатронутым этим методом объяснения. (Не исключено, впрочем, что такие описания могли бы стать великолепным объяснением феноменологии *слепого зрения* (о котором пойдет речь в главе 6), пусть и не обычного сознания!)

Сходный шаг делается Деннетом и в, возможно, центральном аргументе его «Объясненного сознания» (Dennett 1991). После изложения теории способности давать отчет Деннету нужно показать, что она объясняет всё, что нуждается в объяснении, и в частности — что она объясняет опыт в той мере, в какой он нуждается в объяснении. Приведя долгие предварительные маневры, он выдвигает ключевой довод (с. 363–364), утверждая, что теория опыта должна объяснить то, почему вещи *кажутся такими*, какими они кажутся нам. И он доказывает, что его теория может объяснить, почему вещи кажутся такими, какими они кажутся нам. Поэтому он заключает, что его теория объясняет всё, что нуждается в объяснении.

Это элегантный аргумент, и он, в отличие от многих редукционистских аргументов о сознании, имеет видимость правдоподобия. Но его элегантность основана на использовании неявной двусмыслистности понятия «кажимости», балансирующего на грани, отделяющей феноменальное от психологического. «Кажется» можно истолковать в феноменальном смысле, и в этом смысле казаться — значит *переживаться* так-то и так-то. Но есть и психологический смысл слова «кажется», в соответствии с которым вещи кажутся нам такими-то, если мы расположены высказывать суждения о том, что они именно

таковы. Теория опыта должна объяснять кажимость в первом смысле. Деннетовская же теория объясняет ее во втором⁵.

После выявления этой тонкой уловки аргумент утрачивает большую часть своей силы. Когда Деннет говорит, что его теория объясняет кажимость вещей, в конечном счете это означает, что она объясняет то, почему мы *говорим*, что вещи именно таковы, и почему мы соответствующим образом ведем себя в других отношениях. (Как отмечает сам Деннет, его теория сознания основана на его квазибихевиористской теории содержания.) Но объяснение такого рода далеко не соответствует тому, что должна объяснять теория сознания. И в итоге получается, что, называя такую теорию теорией сознания, мы предрешаем все ключевые вопросы.

В общем, если мы начинаем с феноменальных суждений как с экспланандумов теории сознания, мы неизбежно приходим к редуктивному возврению. Но предельными экспланандумами являются не суждения, а сами опытные переживания. Никакое объяснение поведенческих диспозиций не сможет объяснить наличия переживания сознательности агента.

4. Аргументы против объяснительной нерелевантности

Мы видели, что парадокс феноменального суждения ведет к континтуитивным следствиям. Но пока это всё, что мы видели. Кто-то сочтет, что эти следствия не только континтуитивны, но и фатальны. Чтобы утверждать это, он нуждается в *аргументе*. Такой аргумент показал бы, почему сознание просто не может быть объяснительно нерелевантным.

Аргументы такого рода на удивление трудно найти, но их все-таки можно сформулировать. Общая идея состоит в доказательстве того, что объяснительная нерелевантность несовместима с каким-то надежно установленным фактом о нас самих. Я вижу здесь три возможных пути. Можно было бы попробовать доказать, что объяснительная нерелевантность несовместима с фактом нашего *знания* о наших сознательных переживаниях; или что она несовместима с фактом нашей *памяти* о наших сознательных переживаниях; или что она несовместима с фактом наших *отсылок* к нашим сознательным переживаниям. Я не думаю, что какой-либо из этих аргументов убедителен, но все они поднимают интересные проблемы, и все они заслуживают изложения.

Некоторые из этих аргументов естественнее всего выражать в терминах скорее *каузальной* нерелевантности, чем объяснительной нерелевантности. Чтобы полностью раскрыть их потенциал, я на время

соглашусь с тезисом о каузальной нерелевантности опыта с целью посмотреть, пройдут ли эти аргументы. Не исключено, что сходные аргументы можно сформулировать только в терминах объяснительной нерелевантности, но они окажутся более запутанными. В любом случае я по крайней мере допускал, что опыт *мог бы* оказаться каузально нерелевантным, и интересно посмотреть, ведет ли это к фатальным последствиям.

Чтобы в полной мере раскрыть потенциал возражений оппонентов, я также в дальнейшем иногда буду говорить не о «суждениях», а об «убеждениях». Как я отмечал выше, моя главная линия обороны не связана с проведением границы между убеждениями и суждениями, так что здесь я не буду активно прибегать к этому. Этот момент, однако, все же может сыграть вспомогательную роль. В ходе последующего обсуждения следует по крайней мере помнить, что (1) при рассуждении об убеждениях и суждениях зомби мы исходим из их дефляционных понятий и что (2) мои собственные феноменальные убеждения, в полном смысле слова, могут быть отчасти конституированы сознательным опытом.

5. Аргумент от самопознания*

Наиболее сложной проблемой, возникающей вследствие объяснительной нерелевантности, является уже обсуждавшаяся мной проблема нашего знания о наших осознанных опытных переживаниях. Кажется, что мы не просто *судим*, что у нас есть опытные переживания; мы *знаем*, что они есть у нас. Но если нередуктивный взгляд на сознание верен, то опыт объяснительно нерелевантен для вынесения этого суждения; то же самое суждение было бы вынесено и при отсутствии опыта. Трудно поэтому, как может показаться, понять, как это суждение может считаться *знанием*.

Сказанное можно было бы просто выразить в качестве *задачи*: если опыт объяснительно нерелевантен, как мы можем знать об опыте? Эта задача важна как таковая, и она является одним из главных вопросов, связанных с сознательным опытом. Существуют, однако, и другие сложные вопросы такого рода, и не исключено, что мы не сможем ответить на них до того, как мы разработаем детальную теорию сознания. В более сильном виде сказанное может быть выражено и в качестве *аргумента*: если опыт объяснительно нерелевантен, то мы *не могли бы* знать о том, что у нас имеются опытные переживания. Здесь я попытаюсь найти ответ именно на такие аргументы. Я также выскажу ряд идей относительно того, как можно было бы решить указанную задачу, но к данному проекту я еще вернусь.

Я вижу два близких пути развития подобного аргумента. Во-первых, его можно было бы развернуть, напрямую отталкиваясь от возможности моего зомбийного двойника. Мой зомбийный двойник выносит те же самые феноменальные суждения, которые выношу я. Там, где я сужу, что нахожусь в сознании, он тоже выносит суждение о том, что находится в сознании. Более того, его суждения производятся теми же самыми *механизмами*, что и мои суждения. Если суждение, как нередко считается, по праву обосновывается исключительно механизмами его формирования, то суждения, выносимые зомби, будут столь же обоснованы, как и мои. Ясно, однако, что его суждения вообще необоснованы. В конце концов, они являются полностью и систематически ложными. Но кажется, что из этого следует, что и мои суждения тоже не могут быть обоснованными. Они производятся теми же механизмами, которые ответственны за ложные суждения зомби, и поэтому они точно не могут считаться знанием.

Если мои феноменальные суждения не являются более обоснованными, чем суждения зомби, то это подрывает фундамент нередуктивной позиции. Исходный тезис нередуктивной позиции, наше знание о факте опыта, оказался бы разрушен. Это означает, что в данном случае мы имеем дело как с задачей, так и с аргументом. С задачей: как мои суждения могут быть более обоснованными, чем суждения зомби, если признать, что они формируются одними и теми же механизмами? С аргументом: если мои суждения формируются теми же самыми механизмами, что и суждения зомби, то они не могут быть более обоснованными.

Второй аргумент апеллирует к *каузальной теории знания*. Часто утверждается, что ключевым фактором при обосновании некоего убеждения о какой-то сущности является надлежащая каузальная связь между данным убеждением и соответствующей ему сущностью. К примеру, мои убеждения о столе, на который я смотрю, хотя бы отчасти обосновываются тем фактом, что данный стол каузально ответствен за эти мои убеждения. Сторонники каузальной теории полагают, что суждение о каком-то объекте или положении дел должно иметь каузальное отношение к этому объекту или положению дел, если оно должно считаться знанием (возможным исключением являются области априорного знания, к примеру концептуального или математического). Кажется очевидным, что если мое убеждение, что Джон находится в бассейне, не имеет каузального отношения к Джону или бассейну, то я не знаю о том, что Джон находится в бассейне.

Опыт, однако, каузально нерелевантен — или по крайней мере я пока признаю его таковым. Сознательный опыт не играет каузальной роли в формировании суждения о таком опыте. Если каузальная теория

знания верна, то из этого следует, что мы не можем ничего знать о наших опытных переживаниях. Здесь перед нами опять-таки и задача, и аргумент. Задача: как я могу знать об опыте — при условии, что опыт каузально не порождает мои суждения? Аргумент: если опыт не играет каузальной роли в формировании моих суждений, то они не могут считаться знанием.

Шумейкер (Shoemaker 1975a) использует подобные доводы для аргументации в пользу материалистического взгляда на сознание, а также для аргументации в пользу редуктивного функционализма. Шумейкер недвусмысленно одобряет каузальную теорию знания, доказывая, что если мы вообще должны знать об опыте, то он должен каузально порождать наши интроспективные убеждения относительно опыта. Он также использует одну из разновидностей аргумента с зомби для подкрепления редуктивного функционализма. Если зомби или их функциональные эквиваленты логически возможны, то опыт недоступен для интроспекции: зомби наделены теми же самими интроспективными механизмами, что и мы, и поэтому данные механизмы не позволяют нам определить, являемся ли мы зомби или нет. И Шумейкер заключает, что зомби и их функциональные эквиваленты должны быть логически невозможными.

Думаю, что ответ на все эти аргументы достаточно очевиден. Дуалист свойств должен доказать, что каузальная теория знания неприменима к нашему знанию о сознании и что обоснование наших суждений о сознании не связано с механизмами формирования этих суждений. Знание о сознательном опыте во многих важных отношениях резко отличается от знания в других областях. Наше знание о сознательном опыте состоит не в каузальном отношении к опыту, а в совершенно другом отношении.

Этот вывод может быть подкреплен независимыми доводами. Один из таких доводов предполагает изначальное рассмотрение другой возможной реакции со стороны дуалиста свойств, а именно реакции, предполагающей использование *релиабилистской* теории знания. Такая реакция может поначалу казаться перспективной, но, думаю, можно показать, что релиабилистская теория бесполезна в случае нашего знания о сознании. И выясняется, что каузальная теория не работает по той же причине.

Согласно релиабилистской теории, убеждения о чем-то обоснованы в том случае, если они формируются *надежным* путем, то есть в процессе того, что обычно порождает истинные убеждения. Перцептивные убеждения, к примеру, обоснованы, если они возникают в результате оптической стимуляции со стороны окружающих объектов, то есть в результате процесса, обычно порождающего истинные

убеждения; они необоснованы, если порождены галлюцинацией, являющейся очень ненадежным механизмом. С нередуктивной теорией опыта полностью совместима надежность наших феноменальных суждений в актуальном мире: по крайней мере если говорить о номической корреляции, кажется вероятным, что, когда некто выносит суждение, что он имеет визуальный опыт, он имеет визуальный опыт. А вот феноменальные суждения моего зомбийного двойника абсолютно недостоверны; в целом они ложны.

Могло бы, таким образом, показаться, что релиабилистская теория позволяет разрешить наши трудности: из нее следует, что наши суждения об опыте могли бы быть обоснованными даже и при отсутствии прямой каузальной связи, и она обладает ресурсами для объяснения факта обоснованности моих суждений при необоснованности суждений моего зомбийного двойника. Многие, однако, считают, что обращение к релиабилистской теории всё же не может удовлетворить наши запросы; возникает чувство, что мы совершаляем здесь какой-то сомнительный маневр, с помощью которого нельзя решить те задачи, которые надо было решить. Знание, которое отводит нам релиабилистская теория, кажется слишком эфемерным, чтобы соответствовать тому знанию, которое мы имеем о сознательном опыте. И если подумать, то нетрудно увидеть, почему это так.

Проблема в том, что, если наши убеждения о сознании обосновывались бы только релиабилистской связью, мы не могли бы быть *полностью уверены* в том, что обладаем сознанием. Одного лишь существования релиабилистской связи недостаточно для полной уверенности, так как мы не можем исключить возможность того, что эта релиабилистская связь отсутствует и что на другом ее конце вообще нет сознания. Обрести такую уверенность можно было бы лишь при наличии некоторого *дополнительного* доступа к другому концу этого соединения; но это означало бы, что наше знание о сознании опирается и на какое-то другое основание. Такая ситуация нередко допускается в случае нашего знания о внешнем мире: мы не обязаны быть полностью уверенными, что стулья существуют, чтобы *знать* (в обыденном смысле слова), что они существуют, так что нет проблемы в том, что мы полностью не уверены в наличии релиабилистской связи между стульями и нашими суждениями о стульях. Мы, однако, полностью уверены в том, что обладаем сознанием; по крайней мере эта уверенность составляет основу той позиции, которую я отстаивал. Быть может, знание о нашем сознании может быть поставлено под вопрос различными «философскими» доводами, но не напрямую — по аналогии с сомнением в нашем знании о внешнем мире, — но если бы наши убеждения обосновывались только релиабилистской связью, это было бы возможным.

Убеждения, обоснованные только релиабилистской связью, всегда совместимы с наличием *скептических гипотез*. Такие гипотезы связаны со сценариями, при которых вещи кажутся субъекту точно такими же, но соответствующие убеждения оказываются ложными из-за отсутствия релиабилистской связи. В случае перцептивного знания, к примеру, можно сконструировать ситуацию, где отсутствует релиабилистская связь — скажем, когда субъект является мозгами в бочке, — притом что этому субъекту всё по-прежнему кажется точно таким же. Ничто в исходном эпистемическом положении субъекта не исключает этого сценария. В случае же сознания такие скептические гипотезы не могут быть сконструированы. Наше исходное эпистемическое положение уже *включает* наш сознательный опыт. Нет такой ситуации, в которой всё кажется нам точно таким же, но в которой мы не обладаем сознанием, так как наш сознательный опыт (по крайней мере отчасти) конституирует то, какими нам кажутся вещи.

Примечательно, что при создании скептических сценариев, применимых к другим видам знания, таким как наше знание о внешнем мире, философы никогда не забывают о допущении того, что эти сценарии *в том, что касается опыта*, тождественны изначальному сценарию. Как отмечал Декарт, скептицизм может простираться лишь до этого предела. Если скептический сценарий сущностным образом предполагает наличие совершенно иного набора опытных переживаний — к примеру, множества ярких вспышек желтого и зеленого и оглушительного шума, — то он автоматически исключается. Мы *знаем* (в гораздо более сильном смысле, чем раньше), что подобная ситуация не является той, в которой мы находимся.

Из этого следует, что релиабилистская концепция знания не может снабжать нас знанием, достаточно прочным для того, чтобы ему можно было приписать тот характер, который свойствен нашему знанию о сознательном опыте, и поэтому эта концепция неприменима к данному случаю. Однако все сказанное мной о релиабилистском объяснении знания применимо и к каузальной концепции знания. Там, где есть каузальность, имеется и контингентность: существующая каузальная связь могла бы и не существовать. Если единственным источником обоснования убеждения о X является каузальная связь с X , то субъект не может наверняка знать, что данная каузальная связь существует. Он мог бы наверняка знать об этом лишь в том случае, если у него был бы какой-то *независимый* доступ к X или к соответствующей каузальной цепи, но это означало бы, что его знание базируется не только на самой этой каузальной цепи. Всегда нашелся бы такой скептический сценарий, при котором все казалось бы этому субъекту точно таким же, но указанная каузальная связь отсутствовала бы, и X не существовал;

поэтому субъект не может наверняка знать об *X*. Но мы наверняка знаем, что обладаем сознанием; поэтому каузальное объяснение данного знания оказывается неадекватным.

Конечно, оппонент мог бы просто не согласиться с тем, что наше знание о сознании является достоверным, и заявить о *существовании* скептических сценариев, которые не могут быть исключены нами, — к примеру, сценария с зомби. Но всякий, кто избирает этот путь, скорее всего, изначально был элиминативистом (или редуктивным функционалистом) в вопросе о сознании. Если кто-то признает, что наши непосредственные данные не исключают возможности того, что мы зомби, то он должен заключить, что мы *действительно* зомби: ведь это дает гораздо более простую картину мира. Причина того, что проблема сознания вообще существует, однако, в том, что наши непосредственные данные *исключают* указанную возможность. Принимать сознание всерьез — значит соглашаться с наличием у нас таких непосредственных данных, которые исключают его несуществование. Разумеется, всё это опять-таки можно оспорить, но суть в том, что у нас нет особых причин завязывать спор об этом на *данном этапе* аргументации. Мы давно уже расстались с элиминативистами и редуктивными функционалистами. Если мы принимаем сознание всерьез, то у нас есть серьезное основание считать неприемлемым каузальное или релиабилистское объяснение нашего феноменального знания.

Что обосновывает феноменальные суждения?

Главная проблема упомянутых концепций в том, что они делают наш доступ к сознанию *опосредованным*, подобно тому как опосредован — каузальной цепью или надежным механизмом —

наш доступ к окружающим нас объектам. Подобное опосредование уместно при наличии провала между нашей базовой эпистемической ситуацией и теми феноменами, о которых идет речь, как в случае внешнего мира: мы дистанционно связаны с окружающими объектами. Интуитивно ясно, однако, что наш доступ к сознанию вообще ничем не опосредован. Сознательный опыт находится в центре нашего эпистемического универсума; мы имеем *прямой* доступ к нему.

Это приводит к вопросу: что именно обосновывает наши убеждения относительно наших опытных переживаний, если не каузальная связь с ними и не механизмы формирования этих убеждений? Думаю, что ответ на этот вопрос очевиден: данные убеждения обосновываются *наличием* соответствующих опытных переживаний. К примеру, сам факт того, что в данный момент у меня имеется опытное переживание красного, обосновывает мое убеждение, что у меня имеется опытное переживание красного. Поменяйте это переживание на другое или

уберите его вовсе, и главный источник обоснования этого моего убеждения будет устранен. Когда я убежден, что мой опыт состоит в том, что я слышу сильный шум, главной основой этого убеждения является мое опытное переживание сильного шума. В самом деле, какая еще основа могла бы здесь иметься?

Это обстоятельство можно подчеркнуть, отметив, как и раньше, что опыт является частью нашей базовой эпистемической ситуации. Замените мои переживания ярко-красного цвета переживаниями тусклого-зеленого, и вы поменяете мои свидетельства в пользу наличия некоторых из моих убеждений, включая убеждение в том, что я переживаю нечто ярко-красное. Это отражается в факте невозможности создания такого скептического сценария, в соответствии с которым я находился бы в качественно идентичной эпистемической ситуации, но мои переживания были бы радикально иными. Мои переживания — часть моей эпистемической ситуации, и одно лишь их наличие свидетельствует в пользу ряда моих убеждений.

Всё это означает, что в опыте есть что-то внутренне эпистемичное. Переживать — значит автоматически находиться в тесном эпистемическом отношении к переживанию, которое можно было бы назвать «знакомством». Даже концептуально невозможно наличие у субъекта такого переживания красного без *какого-то* эпистемического контакта с ним: переживать — значит подобным образом соотноситься с этим переживанием.

Заметим, что я не утверждаю, что переживать — значит автоматически знать об этом опытном переживании в том смысле, в каком знание требует убежденности. Думаю, что такой тезис был бы ложным: у нас имеется немало опытных переживаний, относительно которых у нас нет никаких убеждений, а поэтому и никакого знания. Кроме того, можно переживать, совершенно не концептуализируя это переживание. Переживать и, следовательно, быть знакомым с этим переживанием — значит находиться в таком отношении к нему, которое является чем-то более изначальным, чем убеждение: оно *свидетельствует* в пользу наших убеждений, но само по себе не составляет какого-либо убеждения.

В самом деле, из сказанного мной не следует, что все убеждения об опытных переживаниях непогрешимы в том смысле, что каждое из таких убеждений автоматически оказывается в полной мере обоснованным. Поскольку убеждения об опытных переживаниях находятся на некоей дистанции от самих опытных переживаний, они могут формироваться по самым разным причинам — и подчас будут формироваться необоснованные убеждения. Если кто-то, к примеру, недостаточно внимателен, он может выносить совершенно неверные суждения

о собственных переживаниях. Мой тезис состоит не в том, что наличие опытного переживания — *единственный* релевантный фактор при обосновании убеждения относительно опыта или при отсутствии такого обоснования. Речь идет лишь о том, что это является *одним* из факторов — возможно, главным из них — и потенциальным источником обоснования, отсутствующим при отсутствии опыта.

Кто-то мог бы подумать, что всё это — не более чем *ad hoc* конструкция, созданная для спасения проблематичной теории, но я не думаю, что она такова. Совершенно независимо от любых соображений, связанных с объяснительной нерелевантностью, у нас имеется очень весомое основание говорить об особой природе эпистемологии опыта и о том, что она по своей сути сильно отличается от эпистемологии других областей. Многие рассуждали о нашем «прямом знании» опыта или «знакомстве» с ним без какого бы то ни было принуждения делать это в оборонительных целях. Многие даже заявляли, что знание об опыте — это основание всего знания именно потому, что мы находимся в таком прямом эпистемическом отношении к нему. Тезис о том, что всё знание проистекает из знания об опыте, может быть преувеличением, но общая идея, состоящая в том, что в нашем знании об опыте есть нечто особое, никогда не была опровергнута⁶.

Аналогичным образом несложно независимо мотивировать и тезис о том, что сами опытные переживания обосновывают наши убеждения относительно опыта. Так, в ходе тщательного обсуждения нашего знания о нашей собственной ментальной жизни Сиверт (Siewert 1994) — всерьез принимающий сознание, но не выказывающий какой-либо симпатии к позиции, согласно которой сознание объяснительно нерелевантно, — детально обосновывает концепцию, признающую наличие «гарантии от первого лица» по отношению к нашим убеждениям о наших переживаниях, по крайней мере отчасти основанной на наличии у нас опытных переживаний. Так что не нужно совершать каких-то неловких движений для того, чтобы истолковать опыт в качестве прямого источника обоснования.

Отвечая на аргументы

При учете сказанного можно дать прямой ответ на аргументы против объяснительной нерелевантности. В ответ на аргумент от каузальной теории знания мы отмечаем, что существует независимое основание полагать, что каузальная теория не годится для объяснения нашего знания об опыте: наше знание об опыте базируется на более непосредственном отношении. А в ответ на аргумент от моего зомбийного двойника мы отмечаем, что обоснование моих убеждений относительно опыта включает не только указание на механизмы формирования этих убеждений — существенным образом оно включает отсылку и к самим

опытным переживаниям. Поскольку мой зомбийный двойник лишен опытных переживаний, он находится в совершенно иной эпистемической ситуации по сравнению со мной, и его суждения лишены соответствующего обоснования.

Соблазнительным могло бы показаться такое возражение: если мое убеждение относится к области физического, то его обоснование тоже должно быть физическим — но такой вывод был бы неверным. Из факта отсутствия физического обоснования можно было бы заключить, что для *физической* части меня (к примеру, для моего мозга) это убеждение необоснованно. Вопрос, однако, состоит в том, обоснованно ли это убеждение для *меня*, а не для моего мозга, и если дуализм свойств верен, то я — это не только мозг. Я состою как из физических, так и из нефизических свойств, и нельзя полностью охарактеризовать меня, отсылая только к одной половине. В частности, обоснование моего убеждения осуществляется благодаря не только моим физическим, но и нефизическими характеристикам, а именно благодаря отсылке к самим опытным переживаниям.

Если упорствовать в возражении и сказать: «Данное убеждение все равно сформировалось бы даже при отсутствии опыта!», то на это мы ответим: «И что?» В *данном* случае у меня есть *свидетельство* в пользу моего убеждения, а именно мое непосредственное знакомство с опытом. В ином случае такое свидетельство отсутствует. Указание на то, что в ином случае это убеждение могло бы быть сформировано в отсутствие такого свидетельства, не означает, что это свидетельство не обосновывает это убеждение в данном случае⁷. Я знаю, что наделен сознанием, и это знание базируется исключительно на моем непосредственном опыте. Тезис о том, что опыт безразличен для процессов моего психологического функционирования, неравнозначен утверждению, что он безразличен для *меня*.

Наконец, в подобных ситуациях всё время воспроизводится мысль: «Но твой зомбийный двойник сказал бы то же самое!» Если я говорю, что я знаю, что наделен сознанием, то отмечается, что и мой зомбийный двойник говорит это. Если я говорю, что мое убеждение обосновано моим непосредственным знакомством с опытом, то отмечается, что и мой зомбийный двойник говорит это. Ответ на это опять-таки: «Ну и что?» В лучшем случае это показывает, что с позиции *третьего лица* мой зомбийный двойник и я идентичны, так что *вы* не можете быть уверены, что я обладаю сознанием; но мы всегда знали об этом. Из этого, однако, никак не следует, что я не могу знать о своем сознании с позиции *первого лица*. С позиции первого лица мой зомбийный двойник сильно отличается от меня. У меня есть опытные переживания, у него — нет. Поэтому у меня есть свидетельство в пользу моего убеждения, а у него нет. Несмотря на то что он говорит то же,

что и я, я знаю, что я отличен от него (хотя *вы* можете и не быть уверены в этом), из-за своего прямого знакомства от первого лица с моими опытными переживаниями. Поначалу это может звучать несколько парадоксально, но на самом деле речь идет об очевидных вещах — о том, что наш опыт сознания позволяет нам знать о наличии у нас сознания.

Даже если возразить, что мой зомбийный двойник был бы *убежден* в тех же вещах, что и я, это никак не может сделать правдоподобной скептическую гипотезу от первого лица о том, что я мог бы быть зомби. В основе подобного возражения может лежать неявное допущение того, что сами убеждения являются изначальными детерминантами моей эпистемической ситуации; поэтому в ситуации, в которой я оказываюсь убежденным точно в том же, в чем я убежден сейчас, такая ситуация в плане наличных свидетельств неотличима от моей нынешней ситуации. Но это, конечно, не так. Свидетельство в пользу моих убеждений по поводу опытных переживаний гораздо более изначально, чем сами убеждения. Изначален сам опыт, убеждения же — по большей части вторичный феномен.

Надо также помнить, что мы исходим из *дефляционного* (то есть функционального) понятия убеждения, и поэтому утверждение, что мой зомбийный двойник убежден в том же, в чем и я, по-прежнему не выходит за рамки констатации того общего, что есть у нас с позиции третьего лица: он предрасположен делать такие же заявления, что и я, такие же выводы и т. п. Здесь не идет речь о том, как вещи выглядят изнутри. Ощущение того, что тезис «зомби обладал бы такими же убеждениями» выступал бы в качестве возражения, может быть связано с допущением *инфляционного* понятия убеждения, согласно которому убеждение — по крайней мере отчасти — является опытным феноменом. Только при таком смысле дело могло бы обстоять так, что идентичность убеждений делала бы ситуации неотличимыми с позиции первого лица; но, разумеется, при таком смысле вообще нет оснований признавать наличие у зомби таких же убеждений.

Итогом всего этого оказывается то, что аргументы от самопознания не дают основания отвергать ту концепцию, которую я отстаиваю. Если мы принимаем сознание всерьез, то у нас уже имеется серьезное основание принимать такую эпистемологию сознания, которая лишает силы эти аргументы. Хотя можно выдвинуть немало соблазнительных аргументов, ни один из них, как представляется, не выдерживает критической проверки.

Конечно, еще предстоит немало сделать для уточнения эпистемологии сознания от первого лица. В лучшем случае я набросал лишь контуры того контекста, в котором можно осмыслить всё это; и остается еще много спорных вопросов. В частности, желательным был бы

анализ того, как именно опыт обосновывает убеждение; того, какие еще факторы релевантны для обоснования убеждений относительно опытных переживаний; того, при каких именно обстоятельствах убеждение об опыте оказывается полностью обоснованным, и т. п. Всё это важные вопросы, которые заслуживают детального рассмотрения при исследовании эпистемологии сознания. Но все они относятся к упомянутой *задаче*, и не видно оснований считать, что эта задача не может быть решена. Важно то, что не проходят те *аргументы*, которые, как могло бы показаться, могли быть выдвинуты против нередуктивного взгляда на опыт на основании самопознания.

6. Аргумент от памяти*

Второе возражение относительно каузальной или объяснительной нерелевантности опыта состоит в том, что оно несовместимо с тем фактом, что мы *помним* о наших опытных переживаниях. Несомненно то, что мне кажется, будто я часто вспоминаю мои прежние переживания, когда, к примеру, мне припоминается терпкий запах нафталиновых шариков от моли в шкафу из моего детства, или когда я вспоминаю особенно яркий оранжевый оттенок закатного неба, на которое я смотрел вчера. Но, как часто утверждается, для воспоминания чего-то надо находиться в надлежащем каузальном отношении к воспоминаемому; подобные утверждения иногда обозначаются как *каузальная теория памяти*. Но если опытные переживания каузально нерелевантны для осуществления моих психологических функций, то дело представляется таким образом, что мои прежние опытные переживания каузально не соотнесены с каким-либо из моих нынешних состояний. И если так, то мы вообще не могли бы помнить о них.

Каузальная теория памяти, однако, не выбита на скрижалях. Она является результатом анализа того, что кажется правдоподобным в разного рода ситуациях. Как и в случае с знанием, дело может обстоять так, что каузальная теория окажется уместной во многих, но не во всех областях. Неочевидно, в частности, что она может быть уместной в сфере опыта. В случае памяти каузальные теории могут быть *не столь* неуместными, как в случае знания, так как нельзя сомневаться в опосредованности нашего отношения к тем опытным переживаниям, о которых мы помним, и весьма вероятно, что это опосредование в значительной своей части содержит каузальную цепь. Но это не значит, что всё здесь сводится к такой цепи.

В случае опытных переживаний, о которых мы помним, несомненно, будет существовать каузальная связь на *психологическом* уровне: базовое когнитивное состояние, имевшееся во время изначального

опыта, будет каузально связано с когнитивным состоянием того времени, когда мы помним о нем. И представляется вполне вероятным, что для того, чтобы помнить об опыте, требуется лишь соответствующая каузальная связь такого рода. К примеру, может существовать каузальная связь между феноменальным *убеждением*, имевшимся ранее, и убеждением, имеющимся позже; и если сказанное в предыдущем параграфе верно, то это изначальное убеждение, будучи обосновано знакомством с самим опытом, может считаться *знанием*. Такая каузальная связь между убеждением, обоснованным знакомством, и более поздним убеждением выглядит вполне достаточной, чтобы счесть это более позднее убеждение примером памяти. Кажется, таким образом, что есть серьезное основание считать, что каузальная связь с опытом не требуется для наличия памяти об этом опыте.

Разумеется, вопрос о том, что считать «памятью», а что — всего лишь обоснованным истинным убеждением о прошлом, во многом связан с тем, какие семантические решения мы принимаем в этих случаях. Важно то, что нередуктивная теория может сохранить лицо, указывая на механизм, посредством которого формируются истинные убеждения о прошлых опытных переживаниях. И если она может сделать это, то это нейтрализует *аргумент*, который можно выдвинуть против такой теории исходя из памяти. Если кто-то *настаивает*, что для памяти требуется каузальная связь с объектом, то мы можем просто сказать, что вместо этого мы обладаем «псевдопамятью» об опытных переживаниях и т. п., и ничего важного не будет потеряно. Но в любом случае мне кажется, что каузальной связи с соответствующим изначальным психологическим состоянием вполне достаточно для того, чтобы считать эти убеждения воспоминаниями.

7. Аргумент от референции*

Третий аргумент против каузальной или объяснительной нерелевантности сознания состоит в том, что она несовместима с нашей способностью *отсылать* к нашим сознательным опытным переживаниям. Кажется несомненным, что мы можем мыслить наши сознательные опытные переживания и говорить о них — я занимался этим на протяжении всей книги. Но иногда утверждается, что отсылка к некоей сущности требует каузальной связи с ней; эти утверждения известны как каузальная теория референции. Если это действительно так, то было бы невозможно отсылать к каузально нерелевантным опытным переживаниям.

Кажется, однако, что нет никакой существенной причины, по которой отсылка к некоей сущности *требовала бы* каузальной связи с ней.

Референция нередко подразумевает каузальную связь, но не очевидно, что дело не может обстоять иначе. При отсылке к какой-то сущности требуется лишь, чтобы наши понятия обладали *интенсионалами* (прежде всего первичными интенсионалами), которым могла бы удовлетворять эта сущность. К примеру, мое понятие «самая большая звезда во Вселенной» имеет первичный интенсионал, указывающий на референт в любом центрированном мире. В актуальном мире этот интенсионал указывает на какую-то звезду *S* — неважно, связан ли я каузально с ней или нет, — и поэтому *S* является референтом этого понятия. При наличии первичного интенсионала, которому могла бы удовлетворять некая сущность в актуальном мире, мы обладаем всеми базовыми ингредиентами, нужными для референции.

Первичные интенсионалы многих наших понятий действительно подразумевают каузальную характеристику: в данном центрированном мире они указывают на соответствующую сущность, каузально связанную с центром. На этой идее основана так называемая каузальная теория референции. Но нет оснований считать, что первичный интенсионал *должен* работать подобным образом. Имеется множество других функций, которые в гипотетическом центрированном мире указывают на сущность, лишенную каузальной связи с центром. Подобные функции могли бы создавать безупречные первичные интенсионалы с безупречным референтом.

Кроме того, *существование* первичного интенсионала — даже в тех случаях, где он характеризуется каузально, — никак не зависит от каузальной связи с референтом. Первичный интенсионал независим от тех событий, которые происходят в актуальном мире. Каузальная связь нередко может играть какую-то роль при *оценке* первичного интенсионала в мире, но эта роль сильно отличается от роли, играемой при определении самого первичного интенсионала. В самом деле, некоторые наши понятия (к примеру, «Санта-Клаус») вообще лишены референта, и тем не менее у них есть первичный интенсионал — интенсионал, который *мог бы* указывать на референт, если бы мир оказался устроен надлежащим образом.

Зачастую каузальная связь с референтом играет определенную роль при *получении* понятия, то есть при формировании первичного интенсионала. Можно было бы попробовать доказать, что даже в случае «самой большой звезды во Вселенной» каузальные связи с миром со-действуют получению базовых понятий, из которых формируется это комбинированное понятие. Но представляется, что здесь опять-таки нет серьезных причин утверждать, что существование первичного интенсионала *требует* каузальной связи с соответствующим предметом. Даже мозг в бочке — несмотря на его каузальную изоляцию — мог бы

обладать понятиями с первичными интенсионалами (хотя большинство из них было бы интенсионалами, которым не удовлетворяло бы ничего из существующего в мире). Еще раз отметим, что *состав* первичного интенсионала независим от таких каузальных связей.

Каузальное отношение занимает центральное место во множестве наших понятий по той естественной причине, что мы обычно отсылаем к тому, о чем мы *знаем*, а знаем мы, как правило, о тех вещах, с которыми мы каузально связаны. Но мы уже видели, что имеется весомое основание отвергнуть каузальную модель знания, по крайней мере тогда, когда речь идет о сознании: в последнем случае наше знание является более непосредственным. Поэтому чтобы отсылать к сознанию, нам не нужно отсылать к нему через интенсионал, который указывает на то, что каузальным образом связано с данным центром; мы можем делать это через интенсионал, указывающий на то, с чем этот центр непосредственно знаком.

В любом случае важно то, что (1) мое понятие «сознания» может иметь первичный интенсионал независимо от существования каузальной связи с референтом (так как существование первичного интенсионала никогда не зависит от подобной каузальной связи); и что (2) первичный интенсионал может указывать на референт независимо от существования каузальной связи с референтом (так как нет причин считать, что первичный интенсионал может указывать на референт только благодаря каузальной связи). Этот интенсионал специфицирует безупречную функцию от центрированных миров к характеристикам этих миров; в нашем мире есть то, что удовлетворяет этому интенсионалу, и поэтому у моего понятия есть референт. Как мы видели, в сознании есть что-то от базового понятия (так же как, возможно, в пространстве и времени), так что не стоит надеяться найти такую детальную характеристику этого интенсионала, какую можно найти для ряда других понятий; вместе с тем нет основания считать, что оно не годилось бы для указания на референт в каком-то мире.

8. Содержание феноменальных убеждений*

Даже если согласиться с тем, что дуализм свойств можно совместить с отсылками к сознанию, остается немало интересных головоломок, связанных с содержанием наших феноменальных понятий и убеждений. Во-первых, возникают вопросы о природе интенсионалов наших понятий — как общих понятий, таких как «сознание», так и более конкретных, таких как «переживание красного»: на что

именно они указывают в некоем данном мире? И во-вторых, возникают вопросы о том, что *составляет* содержание наших понятий: конституируется ли это содержание только нашей психологической природой или же психологической и феноменальной, и какую роль играет каждая из них? У меня нет окончательных суждений на этот счет, но я попробую хотя бы затронуть эти вопросы.

Интересный путь работы с некоторыми из этих вопросов подразумевает рассмотрение того, имеется ли различие между содержанием моих феноменальных убеждений и соответствующих убеждений зомби, и если да, то в чем оно состоит? Я возвращаюсь здесь к рассуждениям скорее об «убеждениях», чем о «суждениях», так как вопрос состоит именно в том, присутствует ли в содержании феноменального убеждения некий элемент, выходящий за пределы содержания феноменального суждения, являющегося, как мы договорились, исключительно психологической сущностью. По крайней мере в целях обсуждения я допущу, что зомби обладает убеждениями (хотя очевидно, что его убеждения не выходят за пределы его суждений). Вопрос в том, существует ли какое-то различие между содержанием его и моих убеждений. В частности, в чем разница — если она вообще существует — между условиями истинности моих и его убеждений и между интенсионалами соответствующих им понятий?

Один из возможных путей состоит в полном отождествлении содержания наших убеждений и понятий. Согласно этой позиции, зомбийные понятия «сознания» и «переживания красного» имеют те же первичные интенсионалы, что и мои понятия об этом, а его убеждения — те же самые (первичные) условия истинности. Его понятие «сознающее существо», к примеру, тоже указывает на сознающие существа в данном центрированном мире. Дело лишь в том, что в его мире отсутствуют такие существа или по крайней мере что он не таков. Поэтому его убеждение «я обладаю сознанием» имеет те же условия истинности, что и мое убеждение; разница лишь в том, что его убеждение *ложно*, а мое истинно.

Чтобы оценить эту линию рассуждений, надо рассмотреть ряд частных вопросов. Во-первых, когда зомби говорит: «Я обладаю сознанием», говорит ли он нечто ложное? Кто-то мог бы сказать, что нет: мы должны снисходительно интерпретировать его замечание, так, будто его понятие отсылает к некоему функциональному свойству, которое он реализует, и тогда его высказывание истинно⁸. Но ведь зомби (по крайней мере если это мой зомбийный двойник) наверняка будет настаивать, что его понятие не является функциональным: он имеет в виду свойство, выходящее за пределы его способности различать, категоризировать, давать отчеты и т. п. И кажется, что у нас

нет оснований не верить ему на слово. Разумеется, мы можем интерпретировать его фразы «дефляционно», и тогда такие заявления, как «ко мне вернулось сознание», могли бы оказываться истинными в обыденных контекстах, но по крайней мере в философских контекстах кажется оправданным оценивать его понятия с позиции более высоких стандартов, на которые он ориентируется, и тогда его убеждения оказываются ложными. И дело не в том, что он пребывает в состоянии концептуальной путаницы, которая могла бы быть устранена более тщательным концептуальным анализом. (Если бы он мог сделать это, я тоже мог бы сделать это; но данное обсуждение предполагает, что я не нахожусь в состоянии подобной концептуальной путаницы.) Поэтому кажется, что в определенном отношении мы вправе говорить о ложности его высказываний о сознании⁹.

Этого, однако, недостаточно для демонстрации того, что понятие, которое есть у зомби, имеет *тот же* интенсионал, что и мое понятие. Быть может, это лишь показывает, что его понятие выглядит примерно так: «свойство, выходящее за пределы каких-либо физических и функциональных свойств», полноценным же понятием сознания он не обладает. Реальной проверкой является то, имеются ли какие-то центрированные миры, в которых идентичны все релевантные обстоятельства, но в которых истинностное значение зомбийного убеждения отлично от моего. К примеру, что если мы оба говорим с каким-то сознающим существом? Я говорю: «Вы обладаете сознанием», и мое высказывание истинно. Но если он скажет: «Вы обладаете сознанием», то истинным ли будет его высказывание? Кто-то мог бы дать отрицательный ответ, так как у него нет непосредственного знакомства с сознанием, которое дало бы ему полноценное понятие. Другая интуиция, однако, говорит, что мы вправе сказать и «да».

Быть может, наиболее релевантные примеры связаны с гипотетическим существом, обладающим каким-то неструктурным, нефункциональным внутренним свойством, которое *не* является феноменальным свойством — если такое свойство возможно, — таким, что ему были бы присущи те же отношения, которые обычно присущи феноменальным свойствам. Когда я говорю такому существу: «Вы обладаете сознанием», мое высказывание ложно; но не обстоит ли дело так, что если мой зомбийный двойник говорит это, то, если его высказывание было истинным в предыдущем случае, оно является истинным и здесь? При отсутствии какого-либо знакомства с сознанием трудно понять, как понятие, которое есть у зомби, могло бы быть достаточно специфицированным, для того чтобы провести различие между этими двумя случаями. Но если это так, то понятие сознания, имеющееся у зомби, не является полноценным понятием.

Если наше знакомство с сознанием играет роль в конституировании первичного интенсионала нашего понятия, то это означает, что понятие «сознание» интересным образом отличается от других понятий, таких как «вода». В случаях таких понятий первичный интенсионал независим от актуального референта; когнитивная система с иным референтом или вообще без референта могла бы обладать тем же самым первичным интенсионалом. Но сознание — это необычный референт: наше отношение к нему непосредственно, и оно находится в центре нашей ментальной жизни — и поэтому могло бы играть необычно важную роль в конституировании первичного интенсионала. В любом случае я оставляю этот вопрос открытym.

А как обстоит дело с *конкретными* феноменальными понятиями, такими как понятие «переживание красного»? Ситуация с ними несколько более запутанна, так как с такими терминами может быть связано не одно понятие. Один из способов разобраться с этими понятиями предполагает рассмотрение вопроса об описании индивидов с инвертированным спектром — когда, к примеру, человек, смотрящий на красные вещи, имеет такие переживания, какие имеются у меня, когда я смотрю на зеленые вещи. Когда я говорю, что такой человек (глядя на красную розу) имеет опыт красного, в каком-то смысле мое высказывание можно счесть истинным: в этом смысле «переживание красного» означает примерно «переживание такого типа, который обычно вызывается (у индивида, обладающего опытом) красными вещами». Не исключено, что существует публично-языковое понятие «переживания красного», которое работает примерно таким образом, но в любом случае я абстрагируюсь от его существования.

Более естественным, возможно, является тот смысл, в котором подобное высказывание оказывается ложным: тот человек имеет не опыт красного, а опыт зеленого. Согласно одному из способов разъяснения этого смысла, первичный интенсионал моего понятия «переживание красного» мог бы означать примерно «переживание такого типа, который обычно вызывается (во мне) красными вещами». Этому смыслу присущи некоторые интересные свойства: к примеру, если мы с вами обладаем инвертированными относительно друг друга спектрами, то ваше понятие «переживание зеленого» будет указывать на то, что я называю переживаниями *красного*. Из этого следует, что ваше замечание «трава вызывает у меня переживания зеленого» могло бы быть истинным, несмотря на то что мое замечание «трава вызывает у вас переживания зеленого» ложно. Иногда такое положение дел объясняется малопривлекательным¹⁰, но оно является естественным следствием индексикальности данного понятия (сходный феномен связан с «Я»-высказываниями, а также с замечаниями о «воде», которые делаю я и мой двойник с Двойника Земли). Если мы хотим избежать

такого рода вещей, всегда можно прибегнуть к охарактеризованному выше более «публичному» понятию¹¹.

Первичный интенсионал этого понятия — «переживание красного» — должен быть совершенно недвусмысленным, если заранее договориться об общем понятии опыта. Ваше понятие «переживание красного» может иметь тот же самый первичный интенсионал, что и мое, даже если наши спектры инвертированы: оба они указывают на те же самые сущности в данном центрированном мире (переживания такого типа, который обычно вызывается в центре красными вещами), хотя наши понятия, разумеется, будут иметь разные *референты*, так как мы населяем разные центрированные миры (мой мир центрирован на мне, а ваш — на вас). Из-за этого наши понятия могут также иметь разные вторичные интенсионалы: мои указывают на переживания красного в контрафактическом мире, а ваши — на переживания зеленого. Даже зомби мог бы располагать тем же первичным интенсионалом, что и мы оба, по крайней мере в той степени, в какой он вообще обладает понятием об опыте, хотя его интенсионал, конечно, не будет ни на что указывать в его центрированном мире, и его понятие «переживание красного» не будет осуществлять какой-либо референции.

Дело, однако, не ограничивается сказанным. Эта реляционная конструкция понятия «переживание красного» всё же относительно периферийна. Между тем с ним связано понятие — быть может, самое важное понятие из тех, что используются в этих дискуссиях, — которое не может быть исчерпано реляционной характеристикой такого рода. А именно, у нас имеется понятие *качества*, к примеру, переживаний красного. Ничто в данной выше реляционной характеристике не ухватывает понятия этого качества, о чем свидетельствует тот факт, что охарактеризованный выше первичный интенсионал совместим со многими различными качествами. Тот, чей спектр был бы инвертирован, тоже имел бы этот интенсионал, но мое понятие об этом качестве — назовем его «*R*» — *отлично от* соответствующего понятия — назовем его «*G*» — у моего инвертированного двойника.

На первый взгляд, можно было бы счесть достаточным ухватывание этого качества *вторичным* интенсионалом моего понятия «переживание красного» при той его характеристике, которая была дана выше. Как мы видели, мой инвертированный двойник и я располагаем различными вторичными интенсионалами, соответствующими различным качествам, на которые указывают наши понятия «переживание красного». Но ухватывания этого качества вторичным интенсионалом недостаточно. Чтобы убедиться в этом, отметим *информационность* узнавания того, что переживания красного (то есть переживания, вызываемые красными объектами) имеют свое специфическое качество.

Иными словами, априори далеко не очевидно, что переживания красного должны быть *R*. Узнавая об этом, мы сужаем нашу модель актуального мира: тот тип опыта, который вызывается красными вещами, мог бы быть *таким* или *таким*, но он *вот такой*. И подобная информативность требует различия в первичном интенсионале: когда два понятия имеют один и тот же первичный интенсионал, априорно истинным является то, что они имеют один и тот же объем¹².

Иначе в этом можно убедиться, заметив, что, когда Мэри впервые ощущает красное, она узнает нечто *отличное* от того, что узнает ее инвертированный двойник, ощущающий зеленое там, где она ощущает красное. Мэри узнает, что красные вещи вызывают переживания *вот такого* рода, а ее двойник узнает, что они вызывают переживания *вот такого* рода. Их модели мира сужаются по-разному: Мэри теперь принимает одно множество центрированных миров, а ее двойник — другое. Из этого следует, что их понятия об интересующих нас качествах должны иметь различные первичные интенсионалы. У Мэри первичный интенсионал указывает на переживания одного типа (*вот этого* типа) в любом данном центрированном мире, у ее двойника — на переживания другого типа.

Мое квазитативное понятие «*R*» играет незначительную роль в коммуникативных практиках. Этим оно напоминает витгенштейновского «жука в коробке»¹³. Мой инвертированный двойник имеет иное родственное понятие «*G*», но другие понимают его так, будто он говорит то же самое, что, как они трактуют, говорю и я, — если допустить, что ситуация, в которой они находятся, остается неизменной в этих двух сценариях. Это отражает ту «невыразимость», о которой я говорил в главе 1: несмотря на богатство внутренней природы переживаний красного, я почти не могу передать это различие словами, разве что пришпиливая его посредством разного рода реляционных свойств и допуская при этом, что у других с ними связаны те же самые переживания. Похоже, таким образом, что коммуникативное бремя несет на себе именно реляционное понятие «переживания красного». Эта невыразимость может быть истолкована как косвенное подтверждение объяснительной нерелевантности опыта: факт, что мы мало что можем сказать о внутреннем качестве опыта, гармонирует с тем фактом, что это качество напрямую не участвует в когнитивных процессах¹⁴.

(Разумеется, мы по-прежнему можем говорить об этих качествах, что я подчас и делал в этой книге. Я могу выражать свои феноменальные убеждения в языке; дело лишь в том, что мой язык будет передавать полное содержание моих убеждений другим, лишь если они сами располагают соответствующими качествами и находятся в таких же отношениях.)

Это явный случай того, когда содержание наших понятий и убеждений конституируется чем-то выходящим за пределы нашей физической и функциональной структуры, — так что редуктивное объяснение содержаний убеждения невозможно¹⁵. Мой инвертированный двойник и я могли бы быть физически идентичными, но соответствующие квалитативные понятия различны, и не только в плане их референции, но и относительно их первичного интенсионала. Перед нами случай — даже более явный, чем случай с «сознанием», — где содержание феноменального убеждения конституируется самой феноменологией. Здесь, несомненно, происходит нечто интересное: некое переживание, которое, как можно было бы подумать, является референтом квалитативного понятия, каким-то образом попадает внутрь понятия и конституирует его смысл (смысл отождествляется здесь с первичным интенсионалом). Это совершенно не похоже на обычные случаи, в которых объект понятия мог бы играть определенную роль в конституировании вторичного, но не первичного интенсионала¹⁶. Это возможно лишь потому, что опыт — *сердце* ментального.

Мы видим, таким образом, что феноменальное убеждение действительно заключает в себе нечто большее, чем феноменальное суждение, по крайней мере в наших случаях. Не исключено, что это могло бы содействовать пониманию эпистемологии сознания. Так, наиболее специфическим случаем подобного конститутивного отношения был бы тот, при котором в конституировании феноменального понятия *S* («опыт этого рода»¹⁷) участвовало бы одно лишь переживание *S*. Прямое конститутивное отношение — способ, каким опыт, так сказать, проникает внутрь понятия, — могло бы помочь уяснению того, как сам опыт мог бы обосновывать убеждение о том, что данный опыт является *S*. И впрямь оно тесно увязывает опыт и убеждение того вида, которое можно было бы счесть подходящим для данной ситуации. И при наличии таких специфических обоснованных феноменальных убеждений можно понять, как из них могли бы вытекать более общие обоснованные феноменальные убеждения (такие как убеждение, что некто обладает сознанием). Я не буду здесь развивать эту тему, так как она полна неясностей, но отношение между переживаниями и феноменальными понятиями дает богатую пищу для размышлений.

(Не исключено, что можно было бы выдвинуть тезис о том, что убеждение, что опыт есть *S*, когда понятие *S* конституируется самим опытом и тем способом, который был выше описан, всегда является обоснованным. Это не отрицало бы существования необоснованных феноменальных убеждений, но они возникали бы в случаях иного отношения между понятием и опытом, к примеру когда понятие, конституированное одним переживанием — или одним набором пережи-

ваний, или внешним описанием, — применялось бы к другому переживанию. Этот сравнительно слабый тезис мог бы, возможно, ухватывать правдоподобный элемент стандартных тезисов о «некорректируемости», оставляя при этом место для всех обычных контрпримеров; он мог бы даже послужить основой детальной концепции эпистемологии сознания. Я, впрочем, не уверен в сказанном.)

Тот факт, что в моем убеждении есть элемент, отсутствующий в соответствующем убеждении моего зомбийного двойника, может помочь такженейтрализовать интуиции, поддерживающие упомянутые ранее эпистемологические аргументы. Идентичными являются *только* наши (функционально сконструированные) суждения; и неверно говорить, что даже при отсутствии опыта имелись бы те же самые убеждения¹⁸. Но ведь здесь, в конце концов, ключевое значение имеют именно убеждения. Мы видели, что подобная различенность убеждений не требуется для опровержения эпистемологических аргументов (и я не предполагал ее наличия в предшествующих обсуждениях), но она все же может способствовать устраниению сомнений, которые могли бы оставаться в этом вопросе.

Естественно поинтересоваться тем, как далеко могло бы простираться такое конституирование содержания посредством опыта. Тот факт, что оно имеет место в случае конкретных феноменальных понятий, дополнительно подкрепляет идею о том, что так обстоит дело и с более общим понятием «сознания», хотя вопрос об отношении между моим понятием и тем понятием, которое есть у зомби, и остается неясным. А далее можно спросить, может ли, как считали некоторые философи, опыт играть какую-то роль в конституировании содержания *нефеноменальных* понятий, таких как понятия о внешних видах. Такое расширение неочевидно, но не исключено, что тот факт, что опыт играет эту роль в одном случае, может давать основание считать, что он может играть ее и в других случаях.

В любом случае ничто не требует здесь ссылаться на каузальную теорию референции. В самом деле, каузальная связь с опытом, похоже, не могла бы служить основой того прямого отношения, которое обнаруживается нами между переживаниями и первичными интенсионалами феноменальных понятий: во всех обычных случаях, в которых имеется каузальная связь (к примеру, в случае с «водой»), подобное отношение отсутствует. Напротив, кажется, что такая конституция обусловлена нашим непосредственным знакомством с опытом. Поэтому каузальная связь с опытом не требуется для конституирования соответствующих первичных интенсионалов, и уж точно никакие каузальные связи не требуются для того, чтобы первичные интенсионалы указывали на референт в таких случаях; поэтому в отсылке к опыту нет

никакой принципиальной трудности, даже если вы занимаете позицию дуализма свойств.

Представляется, таким образом, что, хотя объяснительная нерелевантность опыта по отношению к физическому поведению и может поначалу быть континтуитивной, против нее нет сильных аргументов. Те аргументы, которые могли бы показаться сильными, к примеру от эпистемологии и референции, оказываются просто задачами. Рассмотрение этих моментов поднимает много интересных вопросов, но в итоге мы увидели, что есть серьезное основание считать, что эпистемология и семантика опыта не могут быть сущностно каузальными, а должны истолковываться в иных терминах. Я мало сказал здесь о том, как это могло бы пониматься с позиции дуализма свойств. Исчерпание этих вопросов потребовало бы большого самостоятельного исследования; но надеюсь, что сказанного мной достаточно для ясного понимания того, что нередуктивное воззрение создает естественный контекст для прояснения этих проблем.

Часть III

НА ПУТИ К ТЕОРИИ СОЗНАНИЯ

Глава 6

Когерентность сознания и познания

1. На пути к нередуктивной теории

Даже если сознание нельзя объяснить редуктивно, теория сознания все же может существовать. Нам нужно просто сместиться к *нередуктивной* теории. Мы можем отказаться от проекта, состоящего в попытке объяснить существование сознания исключительно в терминах чего-то более основополагающего, допустив вместо этого, что оно само является фундаментальным, и объяснить, как оно соотносится со всем остальным в мире.

Подобная теория будет сходна по типу с физическими теориями материи, движения или пространства и времени. Физические теории не выводят существование этих параметров из чего-то более фундаментального, но при этом содержат основательные и детальные характеристики таких параметров и их соотношения, благодаря которым в нашем распоряжении оказываются удовлетворительные объяснения многих конкретных феноменов, связанных с массой, пространством и временем. Они добиваются этого, вводя простой и действенный набор *законов*, содержащих различные параметры, из которых вытекают самые разные частные феномены.

Аналогичным образом, краеугольным камнем теории сознания будет набор *психофизических законов*, отвечающих за отношение между

сознанием и физическими системами. Мы уже допустили, что сознание является чем-то естественно (хоть и не логически) супервентным на физическом. Эта супервентность должна гарантироваться психофизическими законами; характеристика этих законов как раз и укажет нам на то, *каким образом* сознание зависит от физических процессов. При наличии физических фактов о какой-то системе подобные законы позволят нам заключить, какого рода сознательный опыт будет связан с этой системой, если с ней вообще будет связан какой-то сознательный опыт. Эти законы как часть базового устройства универсума будут дополнять законы физики.

Из этого следует, что хотя эта теория не будет объяснять существование сознания в смысле ответа на вопрос «почему существует сознание?», она будет в состоянии объяснить конкретные случаи его существования в терминах базовой физической структуры и психофизических законов. Это опять-таки аналогично объяснению в физике, позволяющему понять, почему конкретные случаи материи или движения обладают теми характеристиками, которыми они обладают, путем отсылки к общим базовым принципам в сочетании с определенными локальными свойствами. Все возможные макроскопические физические феномены можно объяснить в терминах базовых физических законов; сходным образом, мы могли бы ожидать, что все возможные «макроскопические» опытные феномены могли бы объясняться в теории сознания психофизическими законами.

В таких законах не надо искать какой-то особой сверхъестественности. Они составляют часть базового устройства природы точно так же, как и законы физики. Конечно, в них будет нечто «элементарное». На каком-то уровне эти законы надо будет принять за истину и счесть далее необъяснимыми. Но то же самое справедливо и для физики: предельные законы природы на каком-то этапе всегда будут казаться чем-то произвольным. Именно это делает их законами природы, а не законами логики.

В науке мы никогда не получаем ничего просто так: что-то и где-то всегда должно приниматься за аксиому. Примечательно, что в большинстве областей науки за аксиому в конечном счете должны приниматься лишь законы физики и, возможно, некоторые ограничительные условия. Но нет оснований считать, что законы физики должны иметь абсолютную привилегию в этом отношении. Если оказывается, что при исследовании сознания надо принять за аксиому некий аспект отношения между физическими процессами и сознанием, то так тому и быть. Такова цена теории.

Тем не менее мы, конечно, хотим вводить минимально возможное количество аксиом. Предельная теория не будет оставлять данную

связь на уровне «состояние мозга X порождает состояние мозга Y » для громадного множества сложных физических состояний и связанных с ними переживаний. Она будет систематизировать эту связь посредством основополагающей объяснительной схемы, устанавливая простые базовые законы, благодаря которым осуществляется эта связь. Физика не желает сводиться к простому множеству наблюдений о положении, скорости и заряде различных объектов в разные времена; она систематизирует эти наблюдения и показывает, каким образом они вытекают из основополагающих законов, настолько простых и действенных, насколько это вообще возможно. То же самое должно быть верно и относительно теории сознания. Мы должны пытаться объяснить супервентность сознания на физическом в терминах наиболее простых из возможных наборов законов.

В конечном счете мы будем стремиться получить набор *фундаментальных законов*. Физики ищут набор базовых законов, настолько простых, что их можно было бы выписать на футболке; того же надо ожидать и в теории сознания. В обоих случаях мы ищем базовую структуру универсума, и у нас есть серьезное основание считать, что она весьма проста. Однако открытие фундаментальных законов может быть далекой целью. В физике мы вначале имели законы, характеризующие макроскопические регулярности, и только после этого подошли к базовым фундаментальным законам. Аналогичным образом, и в теории сознания можно было бы ожидать, что мы начнем с *небазовых* законов, характеризующих отношение между физическими процессами и сознательным опытом на достаточно высоком уровне. Но даже такие высокоуровневые принципы могли бы пока стать для нас важным объяснительным обретением, так же как принципы термодинамики были весьма полезными до установления принципов статистической механики, которые лежат в основании первых. И после обнаружения таких высокоуровневых законов мы наложим серьезные ограничения на любые лежащие в их основании фундаментальные законы, направляя себя, таким образом, в поисках фундаментальной теории.

После того, как мы в конце концов получим в свое распоряжение фундаментальные теории физики и сознания, мы можем обрести то, что с полным правом можно будет назвать теорией всего. Фундаментальные физические законы будут объяснять характер физических процессов; психофизические законы будут объяснять связанные с ними сознательные опытные переживания; а все остальное будет следствием этого.

Разумеется, может оказаться так, что в процессе поиска подобных теорий откроется нечто, что изменит наше представление о предельной теории. Может, к примеру, случиться так, что мы найдем всеобъемлющие законы, включающие как физические феномены, так и феномены сознания в состав более масштабной теории, так же как мы нашли

теорию, объединившую электричество и магнетизм, или как сейчас физики ищут теорию, унифицирующую все основные физические силы. Не исключено, что откроется нечто и еще более удивительное. Но та схема, которая имеется у нас сейчас, по крайней мере может быть отправной точкой для поисков теории, которая могла бы послужить в качестве первого приближения и трамплина для любых более радикальных последующих теорий.

Как мы могли бы построить теорию сознания?

Кто-нибудь мог бы сказать, что весь этот метафизический пафос, конечно, хорош, но как все это выглядит на практике? В частности, как мы можем *найти* те психофизические законы, которые

и будут конституировать теорию сознания? В конце концов, теория сознания сталкивается с грандиозной проблемой, не актуальной для физической теории, а именно с отсутствием данных. Поскольку сознание не может напрямую наблюдаться в экспериментальных контекстах, мы не можем попросту поставить эксперименты, измеряющие опытные переживания, связанные с различными физическими процессами, подтверждая или опровергая ими разные психофизические гипотезы. Не может ли быть так, что эти законы, даже если они существуют, так и останутся неизвестными? И впрямь, могло бы показаться, что непроверяемость любой теории сознания, которую мы могли бы выдвинуть, низведет такие теории до статуса псевдонауки.

Эта обеспокоенность, конечно, не беспочvenна: именно в силу указанных обстоятельств с теорией сознания дело обстоит не так просто, как с физическими теориями. Но это не значит, что мы должны вообще отказаться от поисков теории сознания. Начать с того, что каждый из нас имеет доступ к богатому источнику данных в своем собственном случае. Мы знаем о своих детализированных и конкретных осознанных переживаниях, и нам также известно о фундирующих их физических процессах, так что в нашем распоряжении уже находится большой набор регулярных связей. Исходя из этих регулярностей, мы можем, используя принцип наилучшего объяснения, некоторым образом предположить существование как можно более простых базовых законов, которые могли бы порождать указанные регулярности. В настоящий момент нам вообще неизвестно о каком-либо наборе законов, который мог бы выполнять эту задачу, так что это не является заведомо тривиальным ограничением теории. Вполне могло бы оказаться, что есть только один относительно простой набор законов, позволяющий получать нужные результаты — и тогда у нас было бы серьезное основание считать, что они являются частью верной теории.

Можно упорствовать в возражениях, сказав, что с данными от первого лица совместимы самые разные теории: от солипсизма (согласно

которому только я обладаю сознанием) до панпсихистских теорий (в соответствии с которыми все наделено сознанием); от биохимикалистских концепций (в которых утверждается, что сознание порождается исключительно биохимической организацией определенного рода) до вычислительных теорий (согласно которым для порождения сознания достаточно надлежащей организации вычислительных процессов); в том числе и такие экстравагантные теории, как та, что утверждает, будто люди обладают сознанием только в нечетные годы (в данный момент это 1995 г.). Как мы можем исключить какую-либо из этих теорий, если мы не в силах проникнуть в сознание других людей для измерения их сознательного опыта?

Все подобные теории логически совместимы с имеющимися данными, но этого недостаточно, чтобы сделать их правдоподобными. Солипсистские теории, к примеру, крайне неправдоподобны из-за их в высшей степени произвольного характера (почему сознательной должна быть только эта личность?) и громадной пространственно-временной гетерогенности (мой сознательный опыт систематическим образом связан с моим физическим устройством, но мой физический двойник, находящийся в каком-то другом месте, вообще будет лишен сознания). Соображения правдоподобия играют определенную роль — наряду с ролью эмпирических свидетельств — при создании наших теорий в самых разных областях. Мы, к примеру, одобляем теорию эволюции в отличие от теории, согласно которой мир был сотворен пятьдесят лет назад вместе со всеми воспоминаниями существовавших тогда людей и окаменелостями. Мы также одобляем некоторые простые теории квантовой механики, но не одобляем другие, эмпирически эквивалентные, но крайне искусственные теории. Для создания теории недостаточно эмпирических свидетельств; в расчет должны приниматься также принципы правдоподобия, простоты и эстетичности, а также и другие соображения.

Роль, которую играет, в частности, простота, нельзя переоценить. Без этого ограничения научное теоретизирование в целом было бы удручающе бесконтрольным. Ведь для любой научной теории можно легко придумать эмпирически эквивалентную ей *ad hoc* гипотезу. Но никто не примет такую гипотезу — именно потому, что она неоправданно сложна. Поэтому, если мы сможем отыскать простой набор базовых законов, совместимых с имеющимися у нас данными, у нас будет серьезное основание отвергать более сложные альтернативы.

Содействовать созданию теории сознания могут и другие вероятностные ограничения. Наиболее очевидным из них является принцип, на который мы опираемся всякий раз, когда мы принимаем чей-то вербальный отчет за индикатор его сознательного опыта: принцип,

согласно которому отчеты людей об их переживаниях в общем и целом точно отражают содержание этих переживаний. Мы не можем доказать его истинность, но антецедентно он гораздо правдоподобнее его альтернативы. В определенной мере это правдоподобие основано на заключении от нас самих, но он может рассматриваться и как методологическое ограничение при создании теории сознания. Если бы этот принцип оказался совершенно неверным, это привело бы к полному неведению: тогда мир попросту был бы лишен рациональности, и мы не смогли бы построить теорию сознания. При построении любой теории мы допускаем рациональность мира, допускаем, что планеты не возникают ниоткуда с готовыми окаменелостями и что его законы не сложны и не сверстаны так, чтобы воспроизводить предсказания более простых законов. В ином случае дозволено будет все.

Исходя из подобных вероятностных допущений, мы получаем в свое распоряжение очень полезное ограничение теории сознания, а на деле еще и богатый источник данных даже с позиции третьего лица: чтобы установить, сознательно ли человек переживает стимул, надо просто спросить его! Этот принцип позволяет нам делать гораздо более сильные выводы относительно связи между сознательными переживаниями и их физическими основами. Разумеется, это допущение настолько правдоподобно, что исследователи постоянно опираются на него, и мало кто задумался бы о том, чтобы поставить его под сомнение. Полезную роль играют и родственные ему допущения, такие, как принцип, что воспоминания людей об их переживаниях, в целом, не являются чем-то радикально неверным. Конечно, исключения из этих правил иногда встречаются, но мы, по крайней мере, предполагаем, что отчеты и воспоминания скорее всего будут точными отражениями опыта — при отсутствии серьезных оснований думать иначе.

В число правдоподобных допущений могли бы также входить: гомогенность фундаментальных законов в пространстве и времени; зависимость сознательного опыта только от внутреннего физического состояния организма; маловероятность отражения в сознательном опыте таких случайных факторов, как распределение молекул внутри нейрона, быть может, за исключением того случая, когда они влияют на его работу; и т. д. Разумеется, ложность любого из этих допущений логически возможна, но, при отсутствии оснований думать иначе, разумно допустить истинность каждого из них. Взятые вместе, эти правдоподобные допущения накладывают сильные ограничения на теорию сознания и могут оказать нам существенную помощь при создании соответствующей теории.

Так как же быть с беспокойством по поводу того, что теория сознания не может быть проверена? Оно станет актуальным, лишь если окажется, что существуют две в равной степени простые теории, иде-

ально соответствующие имеющимся данным и релевантным правдоподобным ограничениям. Этого вполне может и не случиться: одна из созданных теорий может оказаться явно предпочтительнее всех ее конкуренток. Но если мы столкнемся с двумя одинаково хорошими теориями, мы можем оказаться перед сложным выбором, — но даже в таком случае мы значительно улучшим свое понимание сознания путем резкого сужения альтернатив. В любом случае ясно, что не стоит заранее беспокоиться по поводу непроверяемости — пока у нас нет даже одной теории, которая могла бы хоть сколько-нибудь удовлетворительно объяснить интересующие нас феномены.

Конечно, эта опора на данные от первого лица и на правдоподобные ограничения означает, что теория сознания будет иметь спекулятивный характер, не свойственный теориям в большинстве других научных областей. Из-за невозможности строгой интерсубъективной проверки мы не сможем обрести полную уверенность, что наши теории находятся на верном пути. По этой причине наука о сознании, по-видимому, всегда будет лишена сильных эмпирических мандатов, имеющихся у других наук, и прожженные исследователи по большей части будут сторониться ее. Но сознание — настолько важный феномен, что лучше хоть как-то понимать его, чем не понимать вовсе: если можно выдвинуть осмысленную теорию сознания и показать, что она превосходит все конкурирующие теории, то это будет важным достижением, даже если эта теория никогда не получит абсолютно убедительного подтверждения. Попросту говоря, такова лодка, в которой мы находимся, пытаясь понять мир: мы берем те материалы, которые имеются у нас в наличии, и мы работаем с этими материалами.

В этой и последующих двух главах я сделаю несколько первых шагов на пути создания теории сознания. В первых двух из упомянутых глав обсуждаются возможные *небазовые* психофизические законы и обосновываются некоторые принципы, выражющие высокую регулярности, обнаруживающиеся в контексте зависимости сознания от физических процессов. Третья глава содержит спекулятивные рассуждения о характере лежащих в их основании фундаментальных законов. Все это не более чем предварительные разыскания, но мы должны начинать с чего-то.

2. Принципы когерентности

Перспективнее всего начать разработку теории сознания, обратив внимание на примечательную *когерентность* сознательного опыта и структуры нашего познания. Феноменология и психология ментального не оторваны друг от друга; они систематически соотнесены.

Множество закономерных отношений между сознанием и познанием может, по сути, послужить отправной точкой для теории сознания. Лучше всего браться за их соотношение, сконцентрировавшись на феноменальных суждениях. Эти суждения имеют психологический характер, но при этом они тесно связаны с феноменологией, и в качестве таковых наводят мост между этими областями. Размышляя об этих суждениях и о том, как они функционируют в нашем собственном случае, мы можем сформулировать целый ряд принципов, связывающих феноменальное и психологическое.

Наиболее очевидным из этих принципов является принцип, который был упомянут мной в первом параграфе: наши второпорядковые суждения о сознании в общем и целом верны. Мы можем назвать его принципом *достоверности*. Когда я высказываю суждение о том, что у меня есть слуховое ощущение, оно обычно имеется у меня. Когда я думаю о том, что только что испытал боль, я, как правило, только что испытал боль. Имеется и обратный принцип, который мы могли бы назвать принципом *детектируемости*: если в наличии есть какое-то опытное переживание, мы обычно можем высказать о нем некое второпорядковое суждение. Конечно, многие переживания ускользают от нашего внимания, но мы, как правило, способны заметить их: переживание, которое в принципе не могло бы быть замечено нами, было бы странной разновидностью опыта¹.

Указанные мной принципы не абсолютны. Наши второпорядковые суждения иногда могут быть ошибочными, демонстрируя тем самым исключения из принципа достоверности. Причиной могло бы быть невнимание (в рассеянном состоянии я мог бы подумать, что только что испытал боль, тогда как в действительности я всего лишь услышал громкий звук), неверная трактовка релевантных категорий (как в случае, когда я ошибочно обозначаю ощущение малинового как ощущение каштанового), психическое заболевание или нейрофизиологическая патология (как в случаях отрицания слепоты, когда субъекты высказывают ложные суждения о своих опытных переживаниях) и множество других обстоятельств. Если говорить об обратных случаях, то можно утверждать, что опытные переживания могут быть незаметными, к примеру, во сне или если они слишком мимолетны, чтобы можно было обратить на них внимание. Тем не менее эти принципы по крайней мере ухватывают значимые регулярности. В типичном случае второпорядковое суждение, как правило, будет верным, а опытное переживание будет таким, что его можно будет заметить. Эти регулярности не абсолютные законы, но они слишком часты, чтобы быть простым совпадением. Мы имеем дело с чем-то систематическим.

Я не буду детально обосновывать эти утверждения, так как я сосредоточу внимание на других принципах когерентности. Но рассмотрим

рение ситуации с позиции первого лица показывает правдоподобность этих принципов, по крайней мере, в этой области, и их естественно распространить на другие случаи, воспользовавшись принципами гомогенности и простоты. Эти принципы одобряются и здравым смыслом, что имеет определенный вес; конечно, здравый смысл может быть пересилен при наличии убедительных оснований, но, при прочих равных условиях, надо оставаться скорее на стороне здравого смысла, чем противоречить ему. Наконец, как я отмечал выше, указанные принципы имеют статус некоего методологического ограничения при разработке теории сознания. Если бы второпорядковые суждения всегда были недостоверны или если бы большинство наших опытных переживаний совершенно нельзя было бы заметить, то наши суждения об опыте так мало соотносились бы с реальностью, что к разработке теории сознания нельзя было бы и приступить.

Наиболее фундаментальный принцип когерентности сознания и познания не содержит указаний на второпорядковые феноменальные суждения. Скорее он касается отношения между сознанием и первопорядковыми суждениями. Принципы, с которыми мы будем иметь здесь дело, имеют отношение к когерентности сознания и осведомленности. Вспомним, что осведомленность — психологический коррелят сознания, и его можно приблизительно охарактеризовать как состояние, в котором некая информация оказывается напрямую доступна для намеренного контроля за поведением и может быть использована для верbalного отчета. Содержательные моменты осведомленности соответствует содержательным моментам первопорядковых феноменальных суждений (с оговоркой, о которой еще пойдет речь), содержательным состояниям — таким как «этот объект красный» — которые не направлены на сознание, а параллельны ему.

Там, где есть сознание, есть и осведомленность. Мое визуальное переживание красной книги на моем столе сопровождается функциональным восприятием книги. Оптические стимулы обрабатываются и преобразуются, и мои перцептивные системы регистрируют то обстоятельство, что на столе находится объект такой-то формы и такого-то цвета, причем эта информация оказывается доступной для контроля за поведением. Это же справедливо и для конкретных деталей переживаемого. Каждая деталь когнитивно представлена в плане осведомленности о ней. Чтобы убедиться в том, что каждая деталь должна быть представлена подобным образом, достаточно обратить внимание на то, что я могу высказываться по поводу этих деталей и выстраивать свое поведение в зависимости от них; к примеру, я могу указывать

Когерентность сознания и осведомленности

на соответствующие части этой книги. Подобная систематическая доступность информации предполагает существование некоего внутреннего состояния, являющегося носителем этого содержания.

Это внутреннее состояние есть первопорядковое феноменальное суждение — по крайней мере в первом приближении. Я делаю это уточнение потому, что можно было бы усомниться, вправе ли мы, строго говоря, вообще называть его «суждением». Содержательная сторона этого состояния не обязана быть чем-то таким, что было бы рефлексивно признано субъектом, а на деле она вообще могла бы быть неконцептуализирована им. Такое состояние можно было бы считать суждением только в слабом смысле; возможно, было бы лучше говорить о нем как о своего рода информационной регистрации или в лучшем случае как об имплицитном или субперсональном суждении. Позже в данной главе я еще поговорю более подробно об этой проблеме, пока же мои рассуждения об этих состояниях как суждениях должны пониматься в широком смысле как указание на класс репрезентативных состояний, которые не обязаны рефлексивно признаваться субъектом и могут не иметь концептуализированного содержания.

То, что верно для визуального опыта, в равной степени верно для любого чувственного опыта. Звуковые переживания репрезентированы в нашей слуховой системе таким образом, что позднейшие процессы имеют доступ к ним при осуществлении контроля за поведением; в частности, соответствующие данности доступны для вербального отчета. В принципе, тот, кто ничего не знает о сознании, мог бы исследовать наши когнитивные процессы и удостовериться в этих данностях, относительно которых имеется осведомленность, наблюдая за той ролью, которую играет информация в плане упорядочения последующих процессов. Таким же образом мы можем трактовать галлюцинации, а также другие случаи, в которых имеются ощущения, но нет реальных ощущаемых объектов. Хотя реальный объект, с которым связано содержание восприятия, и отсутствует, в нашей перцептивной системе все же имеется репрезентация. У Макбета было первопорядковое когнитивное состояние, содержательная сторона которого — «здесь кинжал» — сопровождала его переживание кинжала, несмотря на отсутствие воспринимаемого и переживаемого кинжала.

Под этот колпак подпадает даже и неперцептивный опыт. Хотя у переживания боли может и не быть *объекта*, его содержание в духе «что-то саднит» — или, быть может, вернее «что-то не так» — все же когнитивно репрезентируется. Это ясно из самого факта того, что мы можем высказываться о боли и соответствующим образом направлять свое поведение. Как и в ситуации с визуальным восприятием, здесь имеется осведомленность, хотя объект этой осведомленности

и не столь четко определен. Сходные вещи можно сказать и по поводу переживания эмоций, равно как и других «внутренних» переживаний. Во всех этих случаях переживаниям соответствуют некие когнитивные состояния; если бы дело обстояло иначе, содержание опыта вообще не могло бы отражаться в поведении.

Заметим, что принцип, о котором идет речь, состоит не в том, что всякий раз, когда у нас есть сознательное переживание, мы осведомлены *о нем*. Здесь важны суждения первого, а не второго порядка. Суть этого принципа в том, что, если у нас есть опытное переживание, мы осведомлены относительно его *содержания*. Когда в опыте нам дана книга, мы осведомлены о ее наличии; когда мы переживаем боль, мы осведомлены о присутствии чего-то негативного; в моменты мыслительного опыта мы осведомлены о том, о чем мы мыслим. Речь здесь не идет о том, что за опытным переживанием следует особое суждение, как могло бы быть в случае второпорядковых суждений; нет, эти первопорядковые суждения сопровождают опытные переживания и существуют вместе с ними.

Связь между опытными переживаниями и суждениями второго порядка имеет намного более косвенный характер: хотя мы *способны* замечать наши переживания, по большей части мы замечаем только содержание наших переживаний, а не сами переживания. Лишь иногда мы отстраняемся и подмечаем свое *переживание книги*; обычно мы просто думаем о ней. Редкость суждений второго порядка контрастирует с повсеместностью первопорядковых суждений. Наиболее непосредственной поэтому оказывается связь между сознанием и первопорядковыми суждениями.

Пока я пытался доказать, что сознание сопровождается осведомленностью. Но верно и обратное. При наличии осведомленности обычно имеется и сознание. Если мы осведомлены о чем-то в нашем окружении, и некие данные, о которых можно отчитаться, направляют наше поведение, то в такой ситуации, как правило, имеется и соответствующий сознательный опыт. Когда моя когнитивная система препрезентирует лающую собаку, мой опыт свидетельствует о лающей собаке. Когда я осведомлен о высокой температуре в том месте, где я нахожусь, я чувствую жар. И так далее.

Тут возникают небольшие сложности. Может показаться, что некоторые виды осведомленности не сопровождаются соответствующими им опытными переживаниями. Примером может быть осведомленность, связанная с хранимой в памяти информацией. Я осведомлен, что Клинтон является президентом — в том смысле, что у меня есть доступ к этой информации, что я могу вербально отчитаться о ней и могу использовать ее, осуществляя преднамеренный контроль за своим поведением. Но если я актуально не думаю об этом, то кажется, что

соответствующий сознательный опыт здесь отсутствует; или, если он есть, то он исключительно слаб. Сходным образом я могу быть осведомлен (без актуализации), что внизу стоит велосипед, но при этом не иметь никакого переживания на сей счет. Подобная безопытная осведомленность наиболее заметна в случае пропозициональной осведомленности — я осведомлен, *что* мой велосипед внизу — хотя она может быть приписана и объектуальной осведомленности, так как мы, как кажется, вправе сказать, что я осведомлен *о своем велосипеде*.

Мы могли бы оставить все как есть, но лучше наложить такие ограничения на понятие осведомленности, чтобы придать ему больший параллелизм с сознанием. Кажется правдоподобным допущение, что между процессами, протекающими в одном случае, и процессами, протекающими в другом, имеется *некое* функциональное различие — об этом свидетельствует хотя бы тот факт, что я могу отчитаться о различии между ними. И именно это функциональное различие нам необходимо выделить.

Быть может, самым очевидным различием является то, что при наличии сознательной осведомленности имеется некий *прямой* доступ, отсутствующий в случаях, где осведомленность не сопровождается сознанием. К примеру, информацию о том, что Клинтон — президент, надо «востребовать» чтобы она оказала влияние на преднамеренный контроль за поведением, по крайней мере тогда, когда мы актуально не думаем об этом. Она не так непосредственно готова для оказания влияния на этот процесс контроля, как когнитивные состояния, связанные с переживаниями и актуальными мыслями. Иными словами, когнитивный доступ к этой информации в данном случае не столь прям. Именно это обстоятельство указывает на функциональное различие актуальных и неактуальных мыслей.

Мы, таким образом, можем ввести требование непосредственности доступа в уточненное понятие осведомленности. Согласно этому уточненному понятию, неактуальные мысли не могут трактоваться как часть содержания осведомленности, тогда как актуальные мысли — могут. Соответственно, мы должны ожидать, что актуальные мысли будут связаны с опытными переживаниями, даже если неактуальные мысли не связаны с ними. Именно это мы и обнаруживаем. Моя неактуальная мысль о том, что Клинтон — президент, не оказывает влияния на мою феноменологию, но актуальная мысль об этом будет связана с опытным переживанием. Чтобы убедиться в этом, отметим, что мысль о президентстве Клинтона переживается *определенным* образом; если бы я сейчас не думал об этом, характер моего наличного существования был бы несколько иным².

Таким образом, представляется правдоподобным, что при надлежащем определении осведомленности оказывается, что сознание все-

гда сопровождается осведомленностью, и наоборот. Соответствующий тип осведомленности можно уточнять и дальше; я сделаю это в дальнейшем, в частности, при рассмотрении некоторых интересных случаев. Но даже на этом грубом уровне мы можем увидеть, что данное отношение удачно фокусирует нас в плане понимания когерентности сознания и познания.

Пока мы имели дело с гипотезой: там, где есть сознание, имеется осведомленность, и там, где есть (надлежащая) осведомленность, имеется сознание. Корреляцию между ними можно еще больше детализировать. В частности, различные *структурные* черты сознания напрямую соответствуют структурным чертам, репрезентированным при осведомленности.

Сознательный опыт индивида обычно не является некоей гомогенной каплей; у него есть детализированная внутренняя структура. Мое зрительное поле, к примеру, обладает вполне определенной геометрией. Вот тут — большое красное пятно, неподалеку — маленькое желтое, а между ними что-то белое; тут есть полоски, квадраты, треугольники и т. д. В трехмерном опыте мне даны такие, к примеру, формы, как куб, у меня есть опытное представление о нахождении одной вещи за другой и переживание других проявлений геометрии глубины. Мое зрительное поле содержит громадное множество деталей, сложенных во всеохватную структуру.

Существенно то, что все эти детали когнитивно репрезентированы внутри того, что можно трактовать как структуру осведомленности. Так, величина и форма различных пятен репрезентирована в моей визуальной системе: не исключено, что в виде достаточно точной топографической карты, но даже если и нет, то мы знаем, что они как-то репрезентированы. Так и должно быть — о чем свидетельствует тот факт, что соответствующая информация доступна для осуществления контроля за поведением. Это же верно и для перцептивной репрезентации полосок, кубических очертаний и т. д. Каждая из этих структурных деталей доступна для когнитивной системы и может быть использована для контроля за поведением, а потому и репрезентируется в качестве содержания осведомленности.

В принципе, тот, кто в совершенстве знал бы мои когнитивные процессы, мог бы выявить все эти структурные детали. Геометрия зрительного поля может быть выявлена путем анализа информации, которая предоставляется в распоряжение визуальной системой для использования последующими контрольными процессами; сам факт того, что каждая из этих деталей может быть отражена в поведенческих возможностях субъекта — субъект, к примеру, мог бы фиксировать раз-

**Принцип
структурной
когерентности**

личные структурные детали движениями руки или комментировать их в вербальных отчетах — означает, что эта информация должна где-то присутствовать. Разумеется, детали такого анализа были бы весьма запутанными, и они далеко выходят за пределы возможностей современных методов, но мы все же знаем о существовании соответствующей информации. И таким образом мы можем убедиться, что структура сознания отражается в структуре осведомленности.

Сказанное верно и для *имплицитной* структуры феноменального поля, в частности, отношения между цветами. Даже если в какой-то момент я вижу только один цвет, я *мог бы* видеть множество цветов, находящихся в некотором структурном отношении к данному цвету. Один цвет очень похож на другой и совершенно отличен от третьего. Два цвета могут казаться дополняющими друг друга; одна группа цветов может казаться «теплой», другая — «холодной». При более внимательном анализе выясняется, что наши феноменальные цвета располагаются в трехмерной структуре, упорядоченной по осям красное-зеленое, желтое-синее и белое-черное (выбор осей до известной степени произволен, но их всегда будет три). И оказывается, что эта трехмерная феноменальная структура отражается в трехмерной структуре информации о цвете, обрабатываемой нашими перцептивными системами. Конечно, это предсказуемо — мы знаем, что соответствующая информация доступна для контроля за поведением — но интересно наблюдать, как детализируется эта структура в современных исследованиях визуальной системы (см. Hardin 1988 для обсуждения). Мы могли бы сказать, что в данном случае *структура различий* нашего сознательного опыта (пространство различий между возможными опытными переживаниями) отражается в структуре различий осведомленности: многообразию переживаний цвета и отношений между ними соответствует многообразие представлений цвета и соответствующих отношений между ними.

Подобные имплицитные структуры, а также их аналогичные соответствия имплицитным структурам на уровне обработки информации мы можем обнаружить и в других феноменальных областях. К примеру, феноменологическая структура музыкального аккорда должна отражаться в презентирующей ее структуре, чтобы о ней можно было отчитаться и чтобы она могла оказывать влияние на другие контрольные процессы. Это же справедливо и для имплицитной структуры вкусовых переживаний. Подобные соответствия очень часто обнаруживаются в эмпирических исследованиях интересующих нас процессов; но даже и без таких исследований — поразмыслив над тем фактом, что эти структурные детали могут быть использованы в целях контроля за поведением — можно понять, что какое-то соответствие здесь должно

существовать. В общем, такие рассуждения ведут нас к заключению, что любая детализированная структура, которую можно было бы обнаружить в феноменальном поле, будет отражаться в структурах, представленных в актах осведомленности.

Существует множество и более конкретных характеристик опыта, которые также находят свое отражение в осведомленности. Наиболее очевидной из них является *интенсивность* опытных переживаний. Ясно, что эта интенсивность оказывает влияние на последующие процессы, так что она должна как-то репрезентироваться в структуре осведомленности. И впрямь, правдоподобным выглядит предположение, что интенсивность переживания напрямую соответствует той силе, с которой лежащая в его основании репрезентация стремиться играть контролирующую роль, задействуя ресурсы более поздних процессов (подумаем о различии интенсивной и слабой боли или всепоглощающей эмоции и фонового настроения). Другой такой характеристикой является *разрешение* переживаний, как, к примеру, в случае различия высокого разрешения в центре зрительного поля и низкого разрешения на его периферии. Вполне можно было бы ожидать, что оно будет отражаться в разрешении базовых репрезентаций, и мы обнаруживаем, что дело обстоит именно таким образом.

В общем, даже если опытные переживания в каком-то смысле «невыразимы», отношения между ними не таковы; мы не встречаем трудностей при обсуждении этих отношений — неважно, отношения ли это сходства и различия, геометрические отношения, отношения интенсивности и т. д. Как указывал Шлик (Schlick 1938), даже если *содержание* (внутреннее качество) опыта и не кажется напрямую передаваемым, его *форма* кажется таковой: я могу охарактеризовать отношение между переживанием красного и переживанием зеленого, пусть и не само красное и зеленое³. Поэтому мы должны ожидать, что эти отношения будут когнитивно репрезентированы, что мы, собственно говоря, и обнаруживаем. Сходства и различия опытных переживаний соответствуют сходствам и различиям, репрезентированным при осведомленности; геометрия опыта соответствует геометрии осведомленности и т. д. Если мы уточним понятие осведомленности в вышеуказанном ключе и получим, что состояния осведомленности всегда сопровождаются опытными состояниями, то вероятным окажется и наличие обратного структурного соответствия: структура, репрезентированная при осведомленности, отражается в структуре опыта.

Так что наряду с общим принципом — там, где есть сознание, имеется и осведомленность, и наоборот — мы располагаем и более конкретным принципом: структура сознания отражается структурой осведомленности, а структура осведомленности отражается структурой

сознания. Я буду называть его *принципом структурной когерентности*⁴. Это важное и систематическое отношение между феноменологией и психологией, в конечном счете сводимое к отношению между феноменологией и базовыми физическими процессами. Мы еще увидим его многообразную пользу.

3. Еще о понятии осведомленности

Одним из самых интересных философских проектов в исследовании сознания является проект такого уточнения понятия осведомленности, при котором оно становится более совершенным психологическим коррелятом сознания. Согласно изначальной definции, осведомленность лишь несовершенно соответствует сознанию, однако ее понятие можно уточнить так, чтобы справиться с проблемными случаями. В конечном счете, мы хотели бы найти такую характеристику психологического состояния, при которой была бы правдоподобной его повсеместная корреляция с сознательным опытом, по крайней мере, в известных нам случаях.

Первоначально я определил осведомленность как состояние, в котором некоторая информация доступна для верbalного отчета и намеренного контроля за поведением. Соображения, связанные с пропозициональной осведомленностью при отсутствии опыта, навели на мысль об изменении этого понятия путем добавления требования *прямого доступа*. Возможны и другие модификации. Самая явная из них сопряжена с тем, что доступность для верbalного отчета не является строго необходимой для сознательного опыта, на что указывают соображения, касающиеся опытных переживаний, имеющихся у млекопитающих, хотя при наличии языка она и может быть хорошим эвристическим критерием. Естественным будет изменить definцию осведомленности в таком ключе: это *прямая доступность для глобального контроля*. То есть субъект осведомлен о какой-то информации, если эта информация напрямую доступна для осуществления широкого спектра поведенческих процессов. Это позволяет говорить о возможности опытных переживаний у животных, отличных от человека, и хорошо сочетается с критерием отчета. В случаях, где об информации можно отчитаться, она, как правило, доступна для глобального контроля (в частности, для намеренной координации широкого спектра поведенческих актов). Обратный вывод не всегда верен (как яствует из случаев животного поведения), но, по крайней мере, если речь идет о субъектах, способных отчитаться о своем поведении, доступность информации для глобального контроля обычно означает доступность для отчета.

Конечно, этот уточняющий проект не может выходить за указанные пределы, так как у нас нет устройства для измерения опыта, которое

позволило бы эмпирически подтвердить и уточнить эти гипотезы. Тем не менее в нашем собственном случае мы хорошо представляем, в каких состояниях мы располагаем опытными переживаниями, в каких — нет, и анализ того, что происходит в таких случаях, обычно позволяет нам давать функциональные характеристики подобных состояний. Так что размышление об отношении между опытом и функцией в известных нам случаях оказывается существенным подспорьем для нас. Мы могли бы также попробовать эмпирически уточнить эти гипотезы путем экспериментов от первого лица. К примеру, мы могли бы вызвать у себя какое-то функциональное состояние и посмотреть, какие у нас будут переживания. При небольшом содействии принципов гомогенности и достоверности мы можем делать выводы и на основе исследований аналогичных ситуаций у других.

Определенную роль может играть и эмпирическое рассмотрение более необычных случаев, к примеру, вопроса о том, какие переживания скорее всего имеются у субъектов, страдающих от тех или иных патологий, или у животных. Мы, разумеется, никогда не можем быть полностью уверенными в том, какие именно переживания имеются в этих случаях, но некоторые выводы о них оказываются гораздо правдоподобнее других. По сути, подобные случаи позволяют сфокусировать наши размышления и помочь воображению выделить правдоподобные принципы, касающиеся связи между опытом и функцией. Эти принципы, в конечном счете, могут быть основаны на неэмпирическом анализе, но фокусировка на эмпирических случаях хотя бы привязывает такие размышления к реальному миру.

Так, размышления о приписывании переживаний млекопитающим хорошо сочетается с уточненным критерием, предложенным мной выше. Мы, как правило, готовы приписать перцептивное переживание стимула млекопитающим в тех случаях, когда направленность поведения может быть поставлена в зависимость от этого стимула, особенно если это проявляется в широком спектре поведенческих действий. Если мы обнаруживаем, что информация о стимуле может проявляться лишь в какой-то одной, относительно несущественной поведенческой реакции, то мы могли бы предположить, что эта информация полностью бессознательна. Но по мере того, как ее доступность для использования расширяется, предположение о ее переживаемости становится более вероятным. Так что когерентность сознания и этого понятия осведомленности совместима как с данными от первого лица, так и с естественными мыслями о животных.

Имеется и множество других проблемных случаев, представляющих интерес для анализа. Одним из примеров может быть *слепое зрение* (описанное в Weiskrantz 1986). Это дефект, возникающий при таком

повреждении участков коры головного мозга, отвечающих за зрение, когда нарушаются обычные пути обработки визуальной информации, хотя ограниченным образом подобная информация, как кажется, все же обрабатывается. Субъекты со слепым зрением не могут ничего увидеть в определенных участках их зрительного поля, или, по крайней мере, так они говорят. Если поместить красную или зеленую лампочку в их «слепую зону», они скажут, что ничего не видят. Но если заставить их сделать выбор относительно того, что находится в этой области — к примеру, красная там или зеленая лампочка — то они будут угадывать гораздо чаще, чем ошибаться. Каким-то образом они «видят» предметы этой области, реально не видя их.

Слепое зрение иногда трактуется в качестве примера разведения сознания и связанной с ним функциональной роли. В конце концов, при слепом зрении есть различие, категоризация и даже своего рода вербальный отчет, но, как кажется, отсутствует сознательный опыт. Если бы здесь действительно имело место разведение функциональной роли и опыта, то этот случай явным образом представлял бы собой проблему для принципа когерентности. К счастью, вывод о том, что перед нами пример осведомленности при отсутствии сознания, необоснован. Начать с того, что вовсе не очевидно, что в этих случаях вообще нет опыта; быть может, тут имеются слабые переживания, необычным способом соотнесенные с верbalным отчетом. Но еще более важно то, что это далеко не типичный случай *осведомленности*. Функциональные роли, исполняемые здесь, сильно отличаются от играемых в обычном случае — ведь именно из-за этого различия в функциональных ролях мы и замечаем какой-то дефект⁵.

В частности, субъекты со слепым зрением, по-видимому, лишены обычного доступа к имеющейся информации. Их доступ имеет до странности непрямой характер, о чем свидетельствует тот факт, что она не является непосредственно доступной для вербального отчета и намеренного контроля за поведением. Данная информация доступна гораздо меньшему числу контрольных процессов, чем обычная перцептивная информация она может стать доступной для других процессов, но лишь путем использования необычных методов, вроде побуждения или вынуждения выбора. Поэтому нельзя считать эту информацию напрямую доступной для глобального контроля, и субъекты не являются в полной мере осведомленными о ней в релевантном смысле слова. Отсутствие опыта напрямую соответствует отсутствию осведомленности. Не исключена и возможность того, что субъекты со слепым зрением имеют слабые переживания — и в этом случае можно было бы также принять, что у них имеется слабый уровень осведомленности, при надлежащей трактовке стандартов непосредственности и глобальности. Описание этой ситуации несколько неопределенно, учитывая,

что у нас нет доступа к фактическим данным, но в любом случае оно стыкуется с когерентностью сознания и осведомленности.

В общем, подобные случаи не могут служить свидетельством *против* существования связи между функциональной организацией и сознательным опытом, так как наши выводы о наличии или отсутствии сознания делаются в таких ситуациях именно на функциональных основаниях. В частности, свидетельства о наличии необычных состояний сознания в этих патологических случаях, как правило, базируются исключительно на свидетельствах, говорящих о необычных состояниях осведомленности. Это означает, что подобные случаи не могут нанести ущерб принципу когерентности; они могут лишь подкреплять и уточнять последний.

Не так все гладко с переживаниями, имеющимися во время сна. Во время сна у нас, вероятно, имеются переживания (хотя см. Dennett 1978b), но при этом отсутствует возможность отчитаться о них, равно как и какая-либо роль в контроле за действиями, так как действия полностью отсутствуют. Тем не менее эти случаи можно было бы правдоподобно проанализировать в терминах *доступности* для глобального контроля; дело лишь в том, что сами эти контрольные процессы по большей части выключены. Не исключено, что сама информация размещена так, как она размещается при обычном ее использовании в целях контроля; это предположение подтверждается доступностью наличных содержаний сновидений в полусонном состоянии. Если это так, то мы по-прежнему можем высказывать контрафактический тезис: если бы способности отчитываться и контролировать были активированы (то есть если бы моторные участки коры нормально функционировали), то данная информация могла бы играть какую-то роль. Впрочем, здесь нужен более тщательный анализ, а также эмпирические исследования того, что именно происходит во время сна.

Ряд интересных случаев упоминается Блоком (Block 1995) в ходе его подробного обсуждения различия феноменального сознания и «сознания доступа». Согласно Блоку, состояние «доступно-сознательно», если его содержание готово к использованию в качестве посылки при размышлении, для рационального контроля действия и речи. Так что сознание доступа приблизительно соответствует моей первоначальной дефиниции осведомленности, хотя в моей дефиниции меньше акцентируется рациональность. Блок рассматривает ряд случаев, в которых две этих разновидности сознания могли бы оказываться оторванными друг от друга. Полезно посмотреть, как принцип когерентности мог бы справляться с ними.

Говоря о возможности сознания доступа без феноменального сознания, Блок апеллирует лишь к концептуально возможным случаям,

таким, как зомби; ясно, что эти неактуальные ситуации не могут представлять угрозу для принципа когерентности. Он упоминает слепое зрение, но отмечает, что оно лишь в слабом смысле позволяет говорить о сознании доступа. Он также обсуждает «суперслепое зрение», отличающееся от слепого зрения тем, что натренированность субъекта позволяет ему иметь гораздо лучший доступ к информации из слепой области. Нельзя сомневаться, что это случаи *представимости* оторванности осведомленности от сознания, но сам же Блок замечает, что у нас нет оснований верить в их реальность. Интересно, что он сообщает, что в эмпирически зафиксированных ситуациях, ближе всего соответствующих указанным случаям (речь идет об обезьяне, описанной в (Humphrey 1992), и пациенте, о котором говорится в (Weiskrantz 1992); см. также (Cowey and Stoerig 1992)), есть основание допускать реальное присутствие феноменального сознания.

Что же касается существования феноменального сознания без сознания доступа, то здесь Блок упоминает несколько реальных случаев. Один из них — это ситуация, в которой субъект неожиданно осознает тот факт, что уже какое-то время фоном в его ушах громко визжала дрель. Блок предполагает, что еще до осознания этого обстоятельства, субъект феноменально осознавал звук дрели, хотя и не осознавал его в плане наличия сознания доступа. Используя, однако, мое объяснение осведомленности, кажется уместным сказать, что субъект все это время был осведомлен о дрели. Правдоподобно допустить, что соответствующая информация о дрели была *доступна* все это время; доступ к ней просто не был *востребован*. Поэтому если диспозиционально определить сознание доступа или осведомленность, то этот случай не будет составлять проблемы для принципа когерентности. Блок также упоминает случай, когда три тройных ряда букв на короткое время демонстрируются субъекту (Sperling 1960). Если испытуемых просят назвать буквы верхнего ряда, они могут сделать это, но в таком случае не могут назвать буквы из других рядов; то же самое с другими рядами. Блок доказывает, что субъект феноменально осознает все девять букв, но имеет сознательный доступ только к трем из них. Но и здесь вероятным кажется допущение, согласно которому информация о всех девяти буквах была изначально доступна; дело лишь в том, что востребован был доступ только к информации о трех буквах, и сам этот процесс отрицательно повлиял на возможность востребования доступа к другой информации. Так что этот случай тоже совместим с принципом когерентности, если мы даем диспозициональное объяснение осведомленности.

Можно было бы рассмотреть и множество других случаев. Здесь я ограничился перечислением некоторых из них и кратким иллюстративным анализом для общей характеристики интересного философского проекта. При более тщательном анализе можно было бы попы-

таться наложить более строгие ограничения на то, какой именно *тип* доступности сопровождает сознательный опыт, и о каком именно типе глобального контроля надлежит при этом говорить. Объяснение осведомленности в терминах прямой доступности для глобального контроля — не более, чем начальная точка на данном пути. Это богатая тема для последующего анализа.

Очерченный мной проект может рассматриваться в русле поиска некоего функционалистского объяснения сознания. Это не *редуктивное функционалистское объяснение* —

Отношение к функционалистским теориям сознания*

здесь не утверждается, что для сознания или его объяснения требуется лишь какая-то функциональная роль. Скорее это *нередуктивное объяснение*, указывающее на функциональные критерии возникновения сознания. Тем не менее в определенном смысле оно играет на том же поле, что и редуктивные функционалистские объяснения; наряду с более амбициозными метафизическими тезисами они тоже содержат указание на функциональные критерии появления сознания. Интересно отвлечься от метафизических различий и сравнить различные объяснения исключительно в терминах их функциональных критериев.

Так, идея связи сознания с прямой доступностью для глобального контроля напоминает трактовку Деннетом (Dennett 1993b) сознания как *церебральной селебрити*: «Сознание есть церебральная селебрити — не больше и не меньше. Осознаны те процессы, которые сохраняют и монополизируют ресурсы достаточно долго для оказания типичных и „симптоматичных“ воздействий на память, контроль за поведением и т. д.» (с. 929).

Если оставить в стороне то обстоятельство, что Деннет считает это концептуальной истиной, то его взгляды окажутся весьма близки моим. Главное различие состоит в том, что в моей концепции сознание связывается с *потенциальной* церебральной селебрити. Здесь не требуется, чтобы осознаваемое действительно играло роль в глобальном контроле — нужно лишь, чтобы оно было доступно для такой роли. Это, как кажется, лучше гармонирует с характеристиками опытных переживаний. К примеру, периферия нашего визуального поля переживается нами, но большую часть времени такие периферийные переживания не играют какой-либо важной роли в глобальном контроле; они просто доступны для этого при возникновении соответствующей надобности. Многие слышимые нами шумы могут не оставлять значительных следов в памяти, не оказывать существенного влияния на поведение и т. п., хотя информация о них и *могла бы* оказывать такое влияние. Не исключено, конечно, что Деннет просто употребляет термин «сознание» в более сильном смысле, когда выходит, что мы не осознаем

ни периферию, ни шумы (в конце концов, он подозрительно относится к самой идее опыта), но сравнение все же любопытно.

Другое функционалистское объяснение можно найти у Розенталя (Rosenthal 1996), утверждающего, что состояние будет осознанным, когда оно является объектом мысли более высокого уровня. В используемой мной терминологии это означает, что первопорядковое состояние составляет содержание сознания только тогда когда о нем высказывается суждение второго порядка. Это значительно более сильное утверждение, чем то, которое делаю я — аналогично случаю с Деннетом. Кажется, что у нас нет серьезных оснований полагать, что мы формируем второпорядковые суждения о всех наших опытных переживаниях, включая переживания всех деталей визуального поля, фоновых шумов и т. д. Розенталь считает, что сами второпорядковые суждения обычно бессознательны, и поэтому мы не замечаем их присутствия, но против них, по-видимому, свидетельствуют даже рассуждения с позиции третьего лица. Все эти второпорядковые суждения представляются совершенно необязательными для устройства нашей когнитивной системы. Можно было бы ожидать от некоей системы наличия у нее способности формировать подобные суждения при необходимости, что и происходит в случае наших наиболее заметных переживаний, но система, выносящая второпорядковые суждения о всех деталях визуального поля выглядела бы совершенно избыточной (рис. 6.1).

Согласно концепции Розенталя, сознательные состояния — это такие состояния, которые мы осознаем. Это могло бы иметь видимость правдоподобия, но, на мой взгляд, это так лишь в слабом смысле (о котором шла речь в предыдущей главе) знакомства со всеми нашими переживаниями⁶. Совершенно не очевидно, что большинство наших опытных переживаний является объектом наших мыслей. Предположение о том, что каждая деталь опыта связана с двумя отдельными когнитивными состояниями — с первопорядковым и второпорядковым суждением, создает хаотичную картину ментального. Трудно понять, зачем эволюции могло бы понадобиться повсеместно встраивать эти второпорядковые суждения, когда ее целям в равной степени могла бы служить простая доступность для глобального контроля. Концепцию Розенталя лучше трактовать в качестве объяснения интроспективного сознания, хотя он подает ее как объяснение сознания в аспекте «каково это — быть тем-то и тем-то».

Полезно разделить функциональные объяснения сознания на *первопорядковые* и *второпорядковые*. Согласно второпорядковым объяснениям (к числу которых относится розенталевская концепция высокого уровня мышления, а также концепция высокоуровневого восприятия Лайкана (Lycan 1995) и других), ключевое значение для сознания имеет присутствие какого-то второпорядкового когнитивного состояния.



Рис. 6.1. Zippy the Pinhead о высокоуровневых теориях сознания.
(Перепечатано со специального разрешения King Features Syndicate.)

- (1) — Я наблюдаю себя, наблюдающего себя, пока я рисую себя, наблюдающего себя, наблюдающего себя! — А что делаешь ты, Зиппи?
- (2) — Я наблюдаю себя, наблюдающего, как я наблюдаю тебя, наблюдающего, как ты рисуешь себя, наблюдающего себя. — Говори о своем самосознании!!
- (3) — Я делаю наблюдение о само наблюдении, пока я рисую себя и двух других Я, наблюдающих процесс наблюдения!! — Кто заказывал ананас с ветчиной??

Первопорядковые теории говорят о необходимости лишь первопорядкового когнитивного состояния, с определенными ограничениями играемой им роли. Второпорядковые теории могут хорошо объяснять интроспекцию или рефлексивное сознание, но первопорядковые теории, как представляется, более плотно сочленены с сознательным опытом⁷.

Разумеется, не все первопорядковые когнитивные состояния соответствуют сознательным опытным переживаниям; некоторые первопорядковые суждения о мире могут вообще не соответствовать опыту. В подобных теориях, таким образом, должен быть какой-то дополнительный компонент, который позволит выделить тот релевантный класс первопорядковых состояний. Самый очевидный способ сделать это предполагает ограничение роли таких состояний. Именно это я и сделал, предположив, что релевантными первопорядковыми суждениями являются именно те суждения, которые напрямую доступны для глобального контроля. Сходные ограничения, как мы увидим, вводятся и в других первопорядковых объяснениях. Можно было бы попробовать доказать, что каузальность высокоуровневого мышления есть не что иное, как вариация на тему подобных ограничений, но проблема в том, что оно кажется слишком сильным.

Интересной промежуточной идеей является идея о том, что сознательное состояние соответствует первопорядковому суждению, наделенному потенциалом вызывать второпорядковое суждение о нем⁸.

Такое допущение позволяет избежать хаотичности предыдущего варианта и имеет видимость правдоподобия. В самом деле, оно не так уж и отличается от представления о том, что сознательное состояние соответствует первопорядковому суждению, доступному для глобального контроля: доступность для глобального контроля и доступность для второпорядкового суждения, как можно предположить, будут по большей части пересекаться. Однако у этой концепции могут возникнуть трудности при рассмотрении таких систем, как дети или животные, которые, по-видимому, не лишены переживаний, но не обладают способностью к второпорядковым суждениям; кажется, что здесь требуется большая концептуальная утонченность, чем та, что может быть необходима для наличия переживаний. Если это действительно так, то характеристика в терминах доступности для глобального контроля является более удачной.

Скорее всего, любые первопорядковые функционалистские концепции будут использовать ограничения, отсылающие к чему-то вроде доступности. Это яствует из рассмотрения существующих редуктивных объяснений. К примеру, Кирк (Kirk 1992) предполагает, что перцептивное сознание требует того, чтобы входящая информация была «презентована» главным процессам, ответственным за вынесение решений в системе, а в другом месте говорит (Kirk 1994), что для этого требуется «непосредственно активная» информация. Дретске (Dretske 1995) допускает нечто подобное, рассуждая о том, что опыт — это информация, представленная системе, сходные мотивы можно найти и у Тая (Tye 1995), согласно которому информация должна быть надлежащим образом «подготовлена» для когнитивной обработки. Эти идеи, вероятно, без особого труда можно согласовать друг с другом. Все они, как кажется, хотят выразить нечто сходное.

В любом случае, интересно то, что сторонник нередукционистского подхода к сознанию не обязан смотреть на споры, возникающие между различными первопорядковыми и второпорядковыми концепциями, как на внутренние войны обреченных теорий. Хотя эти концепции не могут объяснить сознание, они все же являются вполне достойными кандидатами на роль теорий когнитивного базиса сознания, и одни из них оказываются более успешными, чем другие. Даже дуалист свойств может допускать наличие в них элемента истины и признавать значимость их различий.

Первопорядковые суждения и перво- порядковые регистраций*

Я嘗試了 доказать, что опытные состояния напрямую соответствуют фундирующими их когнитивным состояниям, названным мной первопорядковыми суждениями. Но, как я отмечал выше и как подчеркивал Дретске (Dretske 1995), именование их *суждениями* может вводить в заблуждение. На-

помним, что изначально суждения были определены так, что они мало чем отличались от убеждений (при допущении исключения каких-либо феноменальных элементов). Но хотя оправданным выглядит предположение о наличии репрезентативного состояния, соответствующего каждой из деталей переживаемого визуального поля, наличие у субъекта убеждений о всех этих деталях не является очевидным. Могло бы, к примеру, показаться, что у меня нет никаких убеждений о данностях, находящихся на периферии моего визуального поля, по крайней мере до тех пор, пока я не обращаю на них внимание. Тем не менее даже при отсутствии убеждений здесь есть какое-то когнитивное состояние, несущее релевантную информацию, так как она по крайней мере доступна.

Мы могли бы просто провозгласить, что термин «суждение» используется нами в более широком смысле, покрывая не только развернутые убеждения, но и когнитивные состояния указанного типа. В конце концов, правдоподобным было бы предположить, что репрезентации периферии визуального поля могут рассматриваться в качестве «микросуждений» или в качестве имплицитных суждений, формируемых в ходе процессов, которые протекают внутри когнитивных систем, даже если они и не являются суждениями, выносимыми субъектом в целом. Но, похоже, лучшим решением было бы избежать путаницы в этом вопросе и ввести более широкий термин для репрезентативных состояний, не являющихся с необходимостью суждениями. Для этого я буду использовать термин «регистрации». К примеру, носителями когнитивных содержаний перцептивных состояний будут не первопорядковые суждения, а первопорядковые регистрации. Первопорядковая регистрация может и не быть состоянием, одобряемым субъектом, хотя она все же является содержательным состоянием, доступным для субъекта и играющим какую-то роль в когнитивной системе.

Первопорядковые регистрации иногда могут даже вступать в противоречие с первопорядковыми суждениями. Наглядным примером сказанного являются оптические иллюзии: субъект мог бы знать, что два объекта имеют одинаковую величину, но в восприятии они могут все равно казаться разными. Другой пример приводит Дретске: вы показываете семь пальцев, и я вижу все семь. Но у меня нет времени подсчитать их, и я ошибочно решаю, что их восемь. Соответственно, я *формирую суждение о восьми пальцах*, хотя мой феноменальный опыт свидетельствует о семи. Данное суждение поэтому не в точности параллельно феноменологии. Тем не менее где-то в перцептивной системе репрезентирована информация о семи пальцах, и она доступна для последующего использования в каких-то системах. Именно эту более раннюю репрезентацию я именую первопорядковой регистрацией. Мы можем представлять такие регистрации в качестве непосредственного

порождения перцептивных и интроспективных процессов, предшествующего их рациональной интеграции в единое целое⁹.

Таким образом, содержательная сторона осведомленности будет, строго говоря, конституироваться не первопорядковыми суждениями, а первопорядковыми регистрациями. А именно, она будет приблизительно покрываться теми первопорядковыми регистрациями, которые непосредственно доступны для использования в целях глобального контроля. При такой дефиниции содержательная сторона осведомленности напрямую соответствует содержательной стороне сознания. Разумеется, некоторые первопорядковые регистрации, связанные, к примеру, с подсознательными восприятиями, не будут входить в состав содержательной стороны осведомленности. Так же как с суждениями, мы можем говорить о первопорядковых *феноменальных* регистрациях для отличия регистраций, соответствующих опытных переживаниям, от тех, которые не соответствуют им. Но меня будут интересовать только регистрации первого класса, так что обычно я буду рассуждать просто о «первопорядковых регистрациях», подразумевая указанное уточнение, но не эксплицируя его.

(Репрезентативное содержание первопорядковых регистраций, похоже, лучше всего трактовать в качестве *неконцептуального* содержания, так же как и параллельное ему содержание опыта. В примечании в конце книги я обсуждаю эту и некоторые другие проблемы, связанные с содержательной стороной осведомленности и содержательной стороной опыта¹⁰.)

4. Объяснительная роль принципов когерентности

Очерченные мной принципы когерентности — не просто какой-то набор метафизических фигур. Они могут играть ключевую роль в ходе эмпирической работы с сознательным опытом. Любое эмпирическое исследование сознания нуждается в качестве необходимого стартового условия в некоем предварительном рассуждении, чтобы получить возможность делать выводы о сознательном опыте на основе физических данных. Принципы когерентности задают эту необходимую точку опоры. Опираясь на них, мы обретаем методологический фундамент для эмпирического исследования сознательного опыта в целом ряде областей. Немало работ такого рода уже проводится; принципы когерентности просто выводят на свет те допущения, которые неявно предполагаются в ходе таких изысканий.

Есть, по крайней мере, три главных проекта, в которых эти принципы могли бы сыграть объяснительную роль. Во-первых, принцип

структурной когерентности может содействовать реализации проекта использования фактов о физических процессах для оказания помощи при объяснении структуры специфических видов опыта. Во-вторых, когерентность сознания и осведомленности работает в качестве своего рода эпистемического рычага, позволяя исследователям делать выводы об опыте, опираясь на данные, полученные с позиции третьего лица. И, в-третьих, когерентность сознания и осведомленности может служить фоновым принципом при поисках физических коррелятов сознания. Я рассмотрю их по очереди.

Первый из этих проектов ярче всего может быть проиллюстрирован современной практикой. Использование эмпирической работы с нейробиологическими и когнитивными процессами для прояснения структурных черт опыта является самым обычным делом. Я, к примеру, уже обсуждал то обстоятельство, что исследование процессов, лежащих в основании цветного зрения, может значительно содействовать объяснению структуры феноменального цветового пространства. Аналогичным образом, исследование топографических карт визуальных участков коры головного мозга помогает прояснить структуру феноменального визуального поля, а исследование процессов, протекающих в слуховой коре, помогает нам понять многие структурные аспекты слуховых переживаний (к примеру, отношения высот звуков и моменты, связанные с направлением). Нечто подобное справедливо и для многих других феноменальных областей.

Кто-то мог бы выразить удивление: как информация о физических процессах могла бы использоваться для прояснения характеристик опыта — учитывая то, что я говорил о невозможности редуктивного объяснения? Принцип структурной когерентности позволяет нам понять, что происходит. В сущности, этот принцип используется как *фоновое допущение* для наведения моста от характеристик физических процессов к характеристикам опыта. Если мы исходим из когерентности структуры сознания и структуры осведомленности, то для объяснения какого-то специфического аспекта первого нам нужно лишь объяснить соответствующий аспект второго. Оставшаяся работа продлевается этим соединительным принципом.

Если, к примеру, речь идет о цвете, то информация о физических процессах дает нам редуктивную картину структуры осведомленности, объясняя релевантные сходства и различия визуальных стимулов, которые обрабатываются колористической системой и делаются доступными для последующего использования теми или иными системами. После того, как мы получили в свое распоряжение эту картину релевантной структуры осведомленности о цвете, принцип когерентности говорит нам, что эта структура будет отражаться в структуре

опытного переживания цвета. Поэтому, если мы исходим из принципа когерентности, то функциональное объяснение визуальных процессов оказывается косвенным объяснением структуры феноменального цветового пространства. Тот же метод можно применять и для объяснения многих других характеристик опыта.

Некоторые авторы были до такой степени впечатлены когерентностью структуры сознания и познания, что высказывали предположение о том, что для физического объяснения сознания больше ничего и не нужно. Ван Гулик (van Gulick 1993), к примеру, отмечает тот факт, что структура нашего цветового пространства напрямую соответствует структуре, репрезентированной при визуальной обработке информации, и высказывает предположение, что это закрывает «провал в объяснении», предоставляя функциональное объяснение ощущения цвета. Кларк (Clark 1993) посвящает целую книгу данной стратегии, доказывая, что чувственные качества могут быть полностью объяснены рассмотрением отношений сходства и различия, существующих внутри квазитативных пространств.

Если сказанное мной выше верно, эти тезисы слишком сильны. Во-первых, данный метод не объясняет *внутреннюю* природу переживания цвета, о чем свидетельствует возможность сохраняющей структуру инверсии спектра. В лучшем случае, он объясняет структуру отношений *между* переживаниями такого рода или между частями сложного переживания; так что для полного объяснения сознания требуется нечто большее. Во-вторых — и это еще более важно — никакое объяснение структуры осведомленности не объясняет, почему она вообще сопровождается опытом, именно потому, что оно не может объяснить, почему принцип структурной когерентности вообще работает. Приняв этот принцип как фоновое допущение, мы уже вышли за пределы *редуктивного объяснения*: данный принцип просто допускает существование сознания и никак не объясняет его. Это разновидность *нередуктивного объяснения* — мы исходим из существования сознания и пытаемся объяснить какие-то из его свойств.

Внутри этих границ принцип структурной когерентности представляет нам исключительно полезное объяснительное отношение между физическим и феноменальным. Если мы хотим объяснить какую-то структуру, явленную нам в области феноменального, скажем, переживаемые нами соотношения музыкальных аккордов, то мы можем исследовать функциональную организацию соответствующей психологической области, пользуясь достижениями когнитивной науки и нейронауки для редуктивного объяснения структуры осведомленности в этой области. Действуя подобным образом, мы даем объяснение структуре феноменального с учетом того вклада, который вносит принцип структурной когерентности. Из-за того, что мы обращаемся к этому

принципу, данные действия не позволяют объяснить само сознание, но все же позволяют объяснить многие характерные черты, присущие конкретной феноменальной области.

При такой трактовке принцип структурной когерентности может служить основой проекта, называемого Криком и Кохом¹¹ «естественной историей квалиа». Даже если нейронаука не может объяснить существование опыта, она может объяснить очень многие факты о нем. Нейронаука может косвенным образом объяснять отношения сходства и различия между опытными переживаниями; геометрию опытных пространств, таких как пространство вкуса и цвета; детальную структуру опытных полей, таких как визуальное поле; восприятие локализации переживаний внутри подобных полей; интенсивность переживаний; длительность переживаний; ассоциации переживаний и многое другое. Как говорят Крик и Кох, нейронаука может объяснить все те характеристики переживаний, которые могут быть объективно передаваемы в процессе коммуникации. Сама сообщаемость этих характеристик означает, что они отражаются в физических характеристиках данной системы, а также и осведомленности. Структурная когерентность сознания и осведомленности — неявный или явный фундамент, на котором базируются подобные объяснения. (Особое значение он обретает в психофизике — я обсуждаю этот вопрос в примечании¹².)

Используя эти методы, мы могли бы даже до какой-то степени понять, каково это — быть летучей мышью! Функциональная организация позволяет нам узнать многое о том, к какого рода информации имеет доступ летучая мышь — какие различия она может осуществлять, как она категоризирует вещи, какие свойства определяют ее перцептивное поле и т. д. — а также о том, каким образом она использует эту информацию. В конечном счете мы должны оказаться в состоянии нарисовать детальную картину структуры осведомленности, присущей когнитивной системе летучей мыши. Используя принцип структурной когерентности, мы затем сможем составить хорошее представление о структуре ее переживаний. Мы не будем знать *всего* относительно того, каково это — быть летучей мышью, — к примеру, у нас будет отсутствовать ясное представление о внутренней природе этих переживаний — но мы будем знать не так уж и мало. Интересная статья Экинса (Akins 1993) о ментальной жизни летучих мышей может рассматриваться в качестве вклада в реализацию этого проекта.

Книга Чейни и Сейфарта (Cheney and Seyfarth 1990) «Как обезьяны видят мир» тоже является ответом на вопрос, сходный с вопросом о летучих мышах, перенося нас в ментальную жизнь другого биологического вида. На деле на протяжении всей этой книги в качестве фонового допущения используется принцип структурной когерентности.

Здесь объясняются определенные функциональные процессы и вытекающая из них структура осведомленности — а затем нас приглашают сделать выводы о соответствующей структуре опыта. Разумеется, по понятным причинам это не позволяет устраниТЬ подлинную причину нагелевской обесспокоенности, но тем не менее нельзя не признать ее поразительным достижением. Нет нужды вспоминать о предельной таинственности сознания всякий раз, когда мы хотим объяснить конкретную феноменальную область.

Принципы когерентности как эпистеми- ческие рычаги

Эмпирические исследователи в нейронауке, психологии, этологии и родственных им областях подчас склонны высказывать утверждения о наличии в системе сознательного опыта. Хотя сознание чаще всего оставляется в этих областях где-то на обочине, имеется немало работ, в которых на основе эмпирических результатов делаются выводы о сознательном опыте. Как это возможно, учитывая трудности прямого наблюдения опыта? Если наблюдать можно лишь физические процессы, что вообще можно рассматривать в качестве оправдания для хоть каких-то выводов о нем?

Ответ должен состоять в том, что всякий раз, когда на основании эмпирических результатов делаются выводы об опыте, они делаются благодаря соединительному принципу, связующему физические процессы с опытом. Соединительный принцип будет давать *критерий* присутствия сознания в системе, критерий, применимый на физическом уровне. Такой принцип будет действовать подобно *эпистемическому рычагу*, позволяющему переходить от знания о физических процессах к знанию об опыте. Сам этот эпистемический рычаг не может стать предметом экспериментальной проверки, по крайней мере с позиции третьего лица; скорее он работает в качестве своего рода априорного фонового допущения. Эти допущения не всегда эксплицируются, но только благодаря им такого рода исследования могут хоть как-то выходить к сознательному опыту.

Соединительные принципы настолько важны здесь, что стоит по-рассуждать о них поподробнее. В определенном смысле всякий, кто использует какой-то из подобных принципов — то есть всякий, кто заключает об опыте, отталкиваясь от внешних наблюдений — «философствует», так как соединительные принципы сами по себе не являются экспериментальными выводами. Такие принципы должны опираться на выкладки с позиции первого лица и на общие принципы правдоподобия. На деле они предшествуют любым экспериментальным результатам, так как они сами и говорят нам о том, как интерпретировать те результаты. Разумеется, в любом экспериментальном начинании задействованы априорные допущения, но здесь они играют необычайно

значимую роль. Важно поэтому приложить все усилия для обоснования этих допущений путем тщательного анализа. Мой проект в этой главе можно трактовать в том числе и в таком ключе.

Соединительный принцип, который был рекомендован мной, говорит о когерентности сознания и осведомленности: если некая система осведомлена о какой-то информации, в том смысле, что данная информация непосредственно доступна для глобального контроля, то эта информация осознается. Подозреваю, что если кто-то провел бы кропотливое исследование соединительных принципов, используемых эмпирическими учеными и теми, кто интерпретирует результаты эмпирических изысканий, то он увидел бы, что почти все подобные принципы оказались бы совместимыми с этим принципом, а на деле и производными от него. Самый обычный соединительный принцип, разумеется, связан с использованием способности давать отчет в качестве критерия наличия опыта: по крайней мере, если речь идет о системах, которые пользуются речью, принято считать, что информация осознана тогда, когда о ней можно отчитаться. Способность давать отчет есть разновидность осведомленности — если об информации можно отчитаться, она всегда доступна для контроля — так что этот критерий очевидно гармонирует с принципом когерентности, хотя он и более ограничен, если говорить об области его возможного применения.

Иногда используются и другие критерии; порой исследователи желают высказаться о переживаниях животных, которые не пользуются речью, или людей, механизмы отчета у которых не функционируют надлежащим образом. В таких случаях лучшим признаком опыта обычно считается заметное воздействие какой-то информации на контроль за поведением. К примеру, Логотетис и Шолл (*Logothetis and Schall 1998*) пишут, что в своем исследовании они изолировали «нейрональные корреляты субъективного визуального восприятия» у обезьян. При этом допускается, что у обезьяны есть перцептивный опыт движущегося объекта в ее окружении, когда она может предсказуемо двигать глазами или нажимать на рычаг в ответ на это движение. Это тоже прекрасно сочетается с критерием осведомленности, или прямой доступности для глобального контроля.

Кто-то может счесть обращение к доэкспериментальным соединительным принципам в экспериментальной науке тем, что нарушает ее чистоту; и впрямь, необходимость использования подобных эпистемических рычагов могла быть одной из причин того, что такие науки так часто держались в стороне от сознания. Тем не менее это та лодка, в которой мы плывем, и лучше делать выводы на основании этих принципов, чем вовсе не делать их. Стоит, однако, артикулировать эти принципы и обосновать их путем тщательного анализа, а не держать

их в подполье. Действуя подобным образом, можно прояснить природу тех рассуждений, которые ведут к эмпирическим заключениям о сознательном опыте.

Физические корреляты сознания

Каковы нейронные и информационные *корреляты* сознания? Это один из центральных вопросов о сознании в эмпирических исследованиях. Выдвигались разные эмпирические гипотезы. К примеру, Крик и Кох (Crick and Koch 1990) высказали предположение, что нейронными коррелятами опыта являются определенные колебания в коре головного мозга частотой 40 Гц. Баарса (Baars 1998) можно понять так, будто он считает, что информационным базисом опыта является глобальное рабочее пространство, а содержательная сторона опыта прямо соответствует содержанию этого рабочего пространства. Фарах (Farah 1994) доказывает, что сознание связано с «высококачественными» репрезентациями в мозге. Либет (Libet 1993) выдвигает нейронную теорию «tempоральной пролонгации», в которой утверждается, что сознание связано с пролонгированными в достаточной степени нейрональными процессами, с минимальной длительностью около 500 миллисекунд. Высказывалось и множество других сходных предположений.

Когерентность сознания и осведомленности естественным образом позволяет придать осмысленность многим подобным исследованиям. Примечательно, что каждый из этих претендентов сам по себе является достойным кандидатом на роль фактора, содействующего осведомленности — прямой доступности для глобального контроля. Колебания, о которых говорят Крик и Кох, принимаются во внимание благодаря той роли, которую они могут играть в объединении информации и ее размещении в рабочую память; и, разумеется, рабочая память есть не что иное, как система, посредством которой данные становятся доступными для контроля. Темпорально пролонгированная нейронная активность у Либета может иметь отношение к делу именно потому, что подобная активность оказывает масштабное влияние на когнитивную систему, нужное для осведомленности. Это же справедливо и по отношению к «высококачественным репрезентациям», о которых говорит Фарах; можно попробовать показать, что «низкокачественные репрезентации» могли бы не справляться с тем, чтобы оказывать надлежащее всеохватное воздействие на когнитивное функционирование. Но наиболее очевидна эта связь в случае глобального рабочего пространства Баарса: оно предлагается для рассмотрения именно в силу его роли в опосредовании глобального доступа и контроля.

Дефляционистская интерпретация этого обстоятельства предполагала бы, что упомянутые исследователи просто *имеют в виду* осведомленность, говоря о сознании, так что эти совпадения не так уж инте-

ресны. Думаю, однако, что из контекста ясно, что многие из них — по крайней мере Крик, Кох, Фарах, а вероятно, и Либет — говорят о сознании в полном феноменальном смысле слова и пытаются выделить его физические корреляты. Из замечания свидетельствуют о том, что все они признали бы *концептуальное* различие сознания и осведомленности в моей трактовке.

Более интересная интерпретация состоит в том, чтобы признать, что эти исследователи говорят о сознании в феноменальном смысле и обратить внимание на то, что все их предложения совместимы с всеохватным соединительным принципом когерентности сознания и осведомленности. В самом деле, эти гипотезы могут быть *производными* от принципа когерентности и релевантных эмпирических данных. Если мы признаем принцип когерентности в качестве фонового допущения, то мы признаем, что опыт непосредственно связан с прямой доступностью для глобального контроля. Если эмпирические данные указывают на то, что у каких-то биологических видов (таких, как *Homo sapiens*) колебания в 40 Гц содействуют глобальной доступности, то у нас есть основание считать, что эти колебания являются коррелятами опыта у данных биологических видов. Если полученные данные указывают на то, что глобальной доступности содействует темпорально пролонгированная активность, то у нас есть основание считать, что подобная активность является коррелятом опыта. И т. д.

Разумеется, верной могла бы оказаться не только одна из упомянутых гипотез. Не исключено, что в различных случаях глобальной доступности содействуют как колебания, так и темпорально пролонгированная активность. Или они вместе могли бы играть какие-то роли на различных стадиях процесса доступа/контроля. Быть может, колебания обслуживают высококачественные презентации в глобальном рабочем пространстве. Это эмпирические вопросы. Впрочем, эти гипотезы могли бы оказаться и ложными. Возможно, выяснится, что те самые колебания не играют какой-либо особой роли в глобальном контроле, а вовлечены лишь в какие-то периферийные процессы. Не исключено, что они оказывают лишь очень небольшое влияние на последующие процессы и на поведение.

Примечательно, что, если у нас было бы основание считать, что колебания оторваны от осведомленности, у нас было бы также основание считать, что они оторваны и от опыта. Если бы оказалось, к примеру, что эти колебания не имеют особого отношения к способности давать отчет и к осведомленности, то это подорвало бы гипотезу о корреляции. В конце концов, у нас нет независимых свидетельств в пользу этой гипотезы: все наши свидетельства опираются на допущение связи со способностью давать отчет и с осведомленностью. Поскольку

у нас нет «опытометра», мы всегда должны опираться на косвенные критерии такого рода, и лучшими из них кажутся критерии способности давать отчет и осведомленности. Из этого следует, что мы можем обладать эмпирическим свидетельством в пользу существования связи между процессом N и сознанием, лишь если у нас есть свидетельство существования связи между N и осведомленностью.

Это подсказывает ясную методологию отыскания физических коррелятов опыта. Доэкспериментальные соображения наводят на мысль, что главным информационным коррелятом сознания является осведомленность или глобальная доступность. Эмпирические данные показывают, что физические состояния, содействующие осведомленности у данного биологического вида, — это состояния определенного типа N . Сочетание априорного соединительного принципа и эмпирических данных позволяют предположить, что N есть физический коррелят сознания у этого вида.

(Интересно, что Деннет (Dennett 1993b) предлагает почти такую же методологию. Он указывает, что конкретные идеи о базисе сознания, такие, как идея колебаний, правдоподобны именно потому, что релевантные процессы содействуют появлению «церебральной селебрити». Я могу согласиться почти со всем, за исключением того, что я выдвигаю требование только *потенциальной* церебральной селебрити, и того, что, на мой взгляд, связь между сознанием и церебральной селебрити — номический принцип, а не концептуальная истина. В самом деле, *всякий*, кто занимается эмпирическим исследованием сознания, будет нуждаться в неэмпирическом соединительном принципе для интерпретации физических результатов в терминах сознательного опыта. Для редукциониста он будет концептуальной истиной, для дуалиста свойств — номическим принципом, основанным на выкладках от первого лица и правдоподобном анализе. Однако многие из высказываемых мной тезисов будут верны в любом случае.)

Естественно допустить, что *главная* корреляция между физическими процессами и опытом — это когерентность сознания и осведомленности. Опыт *непосредственно* порождается не колебаниями, темпорально пролонгированной активностью или высококачественными репрезентациями, а процессом прямой доступности для глобального контроля. Любое более конкретное физическое состояние будет считаться коррелятом лишь в той мере, в какой оно содействует глобальной доступности; поэтому более конкретные корреляции производны от той всеохватной корреляции.

Таких коррелятов может быть множество. Различные типы физических процессов могут, к примеру, обслуживать доступность в различных модальностях. На разных этапах пути обработки информации

тоже могут быть различные корреляты; даже если взять только визуальный опыт, то в зрительной коре может быть один тип коррелята, в дальнейших путях — другой. Конечно, нет никакой гарантии, что даже в рамках конкретной модальности будет задействован какой-либо простой коррелят. Не исключено, что процессы, обслуживающие опыт в зрительной коре, не будут допускать однозначной характеристики; могло бы оказаться так, что их можно было бы сгруппировать *только* по их функциональной роли (то есть опираясь на тот факт, что они обслуживают осведомленность). Но мы можем хотя бы надеяться, что для них можно было бы отыскать более определенную характеристику: если не для перцептивных участков коры, то, быть может, для каких-то последующих сегментов информационного пути.

Мы могли бы также находить корреляты на более высоких, нежели нейронный, уровнях. Создавая надлежащую когнитивную или вычислительную модель, мы могли бы отыскать какой-то способ характеристики в терминах обработки информации тех сущностей, которые ответственны за осведомленность и которые поэтому лежат в основе опыта. Разумеется, простейшая из подобных характеристик будет тавтологичной — процессы, обслуживающие осведомленность, это процессы, ответственные за глобальный доступ, контроль и вербальные отчеты. Но содержательная когнитивная модель могла бы предоставлять нам менее тавтологичные характеристики. Не исключено, к примеру, что мы могли бы охарактеризовать в относительно локальных терминах тот тип информации, который по факту играет значительную глобальную роль при данном общем устройстве системы. В качестве иллюстрации можно привести идею о том, что содержания сознания соответствует данным «селекторных вводных» определенным системам действия (Shallice 1972). По замыслу этой модели, селекторные вводные определяют, какая из систем действия будет «доминировать», то есть играть какую-то роль в глобальном контроле. Если это так, то селекторные вводные содействуют осведомленности и вполне могут оказаться коррелятами сознательного опыта.

Другие информационные корреляты распределяются между тавтологичными и нетавтологичными. Примером может быть глобальное рабочее пространство Баарса. Мы могли бы *определить* глобальное рабочее пространство как своего рода «виртуальную» область, в точности соответствующую широко распространяемым данным; в таком случае то, что заполняет это рабочее пространство, почти по дефиниции оказывается тем, относительно чего имеется осведомленность. Но модель Баарса более эмпирична: он предполагает, что глобальное рабочее пространство есть единая унифицированная система (которую мы можем локализовать по крайней мере в информационных терминах), в которой интегрируется и распространяется информация.

Это предположение могло бы оказаться неверным, так что в нем есть эмпирические моменты, которые могли бы подкрепляться или опровергаться эмпирическими исследованиями. Однако его характеристика рабочего пространства все же достаточно близка характеристике осведомленности — что и объясняет легкий привкус априорности, присущий его предложению; порой кажется, что чуть ли не все эмпирические данные можно было бы согласовать с подобной моделью. (Разумеется, Баарс высказывает многое более конкретных тезисов о процессах в рабочем пространстве, и эти его тезисы в значительной степени эмпиричны.) Аромат тавтологичности можно уловить и в идеи «высококачественной репрезентации» Фарах, хотя это зависит от того, как определять последнюю: если для «высококачественной репрезентации» требуется всего лишь способность играть значительную глобальную роль, то аромат априорности силен, но если определить ее в терминах путей формирования такой репрезентации, он гораздо более слаб.

Даже на когнитивном уровне нет особых оснований верить в существование какого-то одного особого механизма, лежащего в основе опыта. Шактер (Schacter 1989) полагает, что такой механизм, что-то вроде модуля, возможен, но это совсем не обязательно. Дело *могло бы* обстоять так, что глобальному контролю всегда содействовал бы какой-то центральный механизм (вроде глобального рабочего пространства Баарса), но кажется не менее вероятной ситуация, когда надлежащая доступность обеспечивалась бы в разное время множеством разнообразных процессов, даже если говорить об одном биологическом виде или одном-единственном субъекте.

Относительно физических коррелятов порой высказываются и более сильные тезисы, чем те, о которых я говорил. К примеру, если мы обнаружим, что колебания в 40 Гц составляют основу опыта в известных нам случаях, то не могли бы мы предположить, что колебания в 40 Гц — это *предельная основа опыта*? Возможно, эти колебания порождают опыт, даже если они не связаны с осведомленностью, и, соответственно, не может ли быть так, что исправно функционирующие системы без соответствующих колебаний могли бы оказываться системами, лишенными опыта? Подобное заключение было бы, однако, неправомерным. Колебания в 40 Гц были сочтены релевантными *в силу* их связи с осведомленностью; у нас нет оснований считать, что, когда они не играют эту роль, в них есть что-то особенное. И уж точно нет оснований считать, что колебания в 40 Гц в пробирке должны порождать опытные переживания, подобные тем, что есть у меня! И даже в промежуточных случаях, когда речь идет о животных или о ситуациях, когда применяется анестезия, выводы о наличии опыта на основе имеющихся колебаний такого рода были бы рискованными,

за исключением ситуаций, когда наличие этих колебаний давало бы нам основание считать, что там имелась бы какая-то осведомленность.

В общем, мы можем быть уверены, что эти эмпирические методы не позволят нам отыскать такие всеобщие психофизические принципы, которые окажутся несовместимыми с теми принципами, с которых мы начали. Они могут содействовать обнаружению более *конкретных* принципов, приложимых к данному биологическому виду, но их можно будет получить путем прямого применения априорных соединительных принципов. Поскольку эти априорные принципы несут на себе все бремя выводов об опыте на основе физических данных, невозможна ситуация, при которой эти данные подкрепляли бы вывод, противоречащий указанным принципам.

Поэтому мы не должны ожидать от поиска нейронных коррелятов сознания отыскания пути к священному граалю универсальной теории. Мы могли бы ожидать от него ценной помощи в понимании конкретных случаев сознания, в частности человеческого сознания; скажем, более полное представление о процессах, лежащих в основе осведомленности, наверняка поможет нам уяснить структуру и динамику сознания. Но доэкспериментальные принципы когерентности всегда будут играть центральную роль в соединении физических процессов и сознательного опыта.

5. Когерентность как психофизический закон

До сих пор я по большей части рассматривал когерентность в контексте относительно привычных случаев, связанных с людьми и другими биологическими системами. Естественно, однако, предположить, что эти принципы когерентности могут иметь статус всеобщих законов. Если сознание всегда сопровождается осведомленностью, и наоборот, у меня самого и всех людей, то возникает подозрение, что мы имеем дело с каким-то систематическим явлением. В этих привычных случаях, несомненно, имеет место законоподобная корреляция. Поэтому мы можем выдвинуть гипотезу о том, что данная когерентность представляет собой закон природы: в любой системе сознание будет сопровождаться осведомленностью, и наоборот. Это же верно и для развернутого принципа структурной когерентности. Примечательная корреляция структуры сознания и структуры осведомленности кажется слишком уж специфической, чтобы быть чем-то случайным. Естественно поэтому вывести более фундаментальный закон: в любой системе, в любой точке пространства-времени структура сознания будет отражать структуру осведомленности и отражаться в ней.

Подобные законы вносили бы существенный вклад в теорию сознания. Все, что мы пока знаем, — так это то, что сознание каким-то

образом возникает из физического. Однако мы не знаем, благодаря каким физическим свойствам это происходит; то есть мы не знаем, какие свойства задействованы на физической стороне соединения. При учете законов когерентности мы получаем частичный ответ: сознание возникает благодаря функциональной организации, связанной с осведомленностью. При использовании принципа структурной когерентности мы можем даже достаточно конкретно уяснить компоненты супервентностного отношения: дело обстоит не только так, что сознание возникает из осведомленности, но и структура сознания определяется структурой осведомленности.

Конечно, этот закон скорее всего не будет *фундаментальным* психофизическим законом. Фундаментальные законы сочленяют более фундаментальные свойства — или по крайней мере определенные более четко, чем такие высокоровневые конструкции, как «осведомленность». Но не все законы являются фундаментальными. Может даже оказаться, что принципы когерентности — это вообще не *строгие* законы; они могут допускать какие-то маргинальные исключения, особенно если учесть недостаточно определенный характер понятия осведомленности. Но даже если законы не являются ни фундаментальными, ни строгими, они тем не менее накладывают жесткие ограничения на любые фундаментальные психофизические законы. Новая теория сознания, из которой не вытекают принципы когерентности, будет обречена на проблемы. И наоборот, если предлагаемый фундаментальный психофизический закон прост, хорошо мотивирован и имеет своим следствием принципы когерентности, то это может служить хорошим основанием для того, чтобы принять такой закон.

Но каковы же основания признавать принципы когерентности законами? В пользу такого решения главным образом свидетельствуют корреляции в хорошо известных нам случаях — в конечном счете у меня самого, в моем собственном случае. Видимые корреляции между осведомленностью и сознанием у меня самого настолько детализированы и приметны, что здесь должно быть нечто большее, чем случайная регулярность. В их основе должен лежать какой-то закон. Вопрос лишь в том, *какой* закон? Из этого закона должно вытекать, что в моем собственном случае осведомленность всегда будет сопровождаться сознанием, и наоборот, и, кроме того, что их структуры должны соответствовать друг другу. Выдвинутые мной принципы когерентности удовлетворяют этим условиям. Но мог бы использоваться для этого и какой-нибудь другой принцип?

Крайне вероятным выглядит утверждение, что некая осведомленность *необходима* для сознания. Все известные мне случаи сознания совершенно точно сопровождаются осведомленностью. Кажется, что у нас нет серьезных оснований считать, что какие-то случаи сознания

не сопровождаются функциональными процессами. Если они и существуют, у нас нет никаких свидетельств на этот счет, даже косвенных, и мы в принципе не могли бы иметь их. Поэтому из соображений экономии разумно допустить, что везде, где есть сознание, имеется и осведомленность. Если мы ошибаемся в этом — скажем, если стационарный электрон живет богатой сознательной жизнью Пруста — то мы наверняка никогда не узнаем об этом.

Вопрос о *достаточности* осведомленности более сложен. Если признать необходимость осведомленности, то любой кандидат на роль базового закона будет иметь такой вид: «Осведомленность плюс нечто порождает сознание». По крайней мере, любой базовый закон должен *имплицировать* принцип такого вида для объяснения регулярностей в моем собственном случае. И остается, таким образом, спросить: что же это за дополнительное нечто, — или тут вообще не нужно ничего дополнительного?

Назовем этот гипотетический дополнительный ингредиент *X-фактором*. Либо я обладаю сознанием только благодаря осведомленности, либо я обладаю сознанием благодаря осведомленности и X-фактору. X-фактор вполне мог оказаться любым свойством, лишь бы я обладал им в настоящий момент, а лучше бы на протяжении всей моей жизни. Быть может, он связан с национальностью, и осведомленность порождает сознание только у австралийцев. Быть может, он связан с местонахождением, и осведомленность порождает сознание только в радиусе ста миллионов миль от звезды. Или, быть может, он связан с *идентичностью*, и осведомленность порождает сознание только у Дэвида Чалмерса.

Все эти законы были бы совместимы с имеющимися у меня свидетельствами и объясняли бы данную корреляцию — но почему же все они кажутся столь нелепыми? Причина в том, что в каждом из этих случаев X-фактор представляется чем-то совершенно *произвольным*. Нет оснований считать, что сознание должно зависеть от подобных вещей; они выглядят как излишества, не имеющие никакого отношения к делу. Дело обстоит не так, как если бы X-фактор играл какую-то роль в объяснении какого-либо из феноменов, связанных с сознанием. Осведомленность по крайней мере могла бы объяснить наши феноменальные суждения, тесно сопряженные с сознанием, так что здесь есть основание допускать какую-то связь. Каждый же из X-факторов кажется, по контрасту, возникающим из ниоткуда. Почему мир оказывался бы таким, что осведомленность порождала бы сознание у одной личности и только у нее? Такой способ существования мира был бы странным, произвольным способом существования.

Сказанное применимо и к более «правдоподобным» X-факторам, которые могли бы предлагаться в качестве таковых не в шутку, а всерьез.

Естественным претендентом на роль подобного X-фактора является клеточная биология или даже человеческая нейрофизиология. Некоторые и в самом деле допускали, что сознание ограничено существами с надлежащей биологической организацией. Аналогичным образом некоторые предполагали, что сознание возникает из функциональной организации только тогда, когда эта организация не реализована «гомункульным» способом, как в примере с Китайской нацией. Но такие X-факторы в равной степени произвольны. Они лишь усложняют законы без какой-либо дополнительной компенсации. Почему мир должен быть устроен таким образом, что осведомленность порождает сознание только у существ с особой биологией или у таких, у которых нет внутренних гомункулов? Содержа посторонние ингредиенты, эти гипотезы выглядят причудливо.

Почему кто-то вообще мог бы поверить в X-фактор? Думаю, что вера него возникает по естественной, но обманчивой причине. Существует базовая интуиция, говорящая о том, что сознание есть нечто выходящее за пределы функциональной организации. Эту интуицию я, разумеется, разделяю — сознание есть дополнительный факт, и никакая функциональная организация не может быть логически достаточной для него. Существует также естественная тенденция верить в то, что все имеет физический характер и что сознание можно так или иначе объяснить физически. Это двойное давление вызывает естественную реакцию: мы должны добавить что-то еще, и эта добавка должна быть физической. Человеческая биология оказывается естественным кандидатом на роль такого дополнительного ингредиента. И можно было бы подумать, что так мы закрываем провал между функциональной организацией и человеческой биологией.

Но это совершенно неверный путь. Добавление в картину биологии не оказывает никакой помощи в решении первоначальной проблемы. Провал столь же велик, как и прежде: сознание кажется выходящим и за пределы биологии. Как я доказывал раньше, *никакие* физические факты не достаточны для объяснения сознания. X-фактор не может выполнить для нас никакой работы; мы ищем не там, где надо, решения нашей проблемы. Проблема возникла при допущении материализма. Как только мы признаем, что материализм ложен, становится ясно, что поиск физического X-фактора не имеет отношения к делу; вместо этого мы должны искать «Y-фактор», нечто *дополнительное* по отношению к физическим фактам, которое поможет объяснить сознание. Мы находим такой Y-фактор, провозглашая нередуцируемые психофизические законы. После введения этих законов в нашу модель мы сохраняем интуицию сознания как дополнительного факта и устранием упомянутую проблему.

Желание отыскать физический X-фактор — отголосок попытки съесть материалистический пирог и при этом закусить сознанием. Если мы признаем, что сознание есть дополнительный нефизический факт и что существуют независимые психофизические законы, X-фактор окажется совершенно излишним. Требовать независимой психофизической связи и X-фактора — значит требовать двойной награды там, где нужна только одна.

Таким образом, для X-фактора не остается объяснительной роли, которую он мог бы сыграть в теории сознания, и он только усложняет картину. Любой подобный фактор лишь без нужды усложняет фундаментальные законы. В сравнении с простотой картины, когда осведомленность порождает сознание, мир, в котором сознание зависит от обособленного X-фактора, выглядит весьма неразумно устроенным. С тем же основанием можно было бы дополнить законы Ньютона условием, гласящим, что любое действие сопряжено с равным противодействием, за исключением тех случаев, когда взаимодействующие объекты сделаны из золота. Принципы простоты диктуют нам рассматривать в качестве наилучшей гипотезы ту, которая не требует X-фактора и которая указывает на то, что осведомленность порождает сознание без каких-либо оговорок.

Некоторые по-прежнему будут сомневаться в справедливости сделанного мной функционалистского вывода, даже если речь в данном случае идет о дуалистической версии функционализма. Аргумент от X-факторов действительно несколько зыбок, и он в значительной степени опирается на допущения, определяемые соображениями простоты. В следующей главе я приведу более конкретные аргументы в пользу того же самого вывода, используя мысленные эксперименты для обоснования того, что функциональная копия сознательного существа будет наделена в точности таким же видом сознательного опыта. Пока же я отмечу, что эти выкладки по крайней мере *prima facie* позволяют уверенно заявить, что указанный тип функционализма кажется верным решением.

Здесь стоит на мгновение остановиться и охватить взглядом общую эпистемологическую схему. По сути, тут идет речь о заключении к лучшему объяснению. Мы констатируем примечательные регулярные связи между сознанием и осведомленностью в нашем собственном случае и допускаем существование простейшего из возможных базовых законов. Такие же по типу рассуждения применяются и при формулировке физических теорий, а также при критике скептических гипотез о причинности и внешнем мире. Во всех этих случаях делается базовое допущение о том, что мир является простым и рационально устроенным местом. Без этого допущения возникает полный хаос. С ним же все становится на свои места.

Представляется также, что данный подход позволяет нам получить наилучшее из возможных для нас решений проблемы других сознаний. Мы замечаем регулярные связи между опытом и физическими или функциональными состояниями в нашем собственном случае, допускаем простые и гомогенные базовые законы для их объяснения и используем эти законы для вывода о том, что у других тоже существует сознание. В своем убеждении о наличии сознания у других мы, возможно, неявно опираемся на это рассуждение, — а может и нет, но в любом случае оно, как кажется, предоставляет нам рациональное обоснование наших убеждений.

Интересно поразмышлять о том, какие выводы следуют из наших принципов когерентности относительно существования сознания за пределами человеческого рода и, в частности, у гораздо более простых организмов. Вопрос этот неясен, так как наше понятие осведомленности четко определимо лишь для комплексных случаев, сходных с человеческими. Кажется оправданным сказать, что собака осведомлена о чем-то и даже что мышь осведомлена о чем-то (быть может, они не осведомлены о самих себе, но это уже другая история). К примеру, кажется, что мы вправе говорить, что собака осведомлена о пожарном гидранте в базовом смысле термина «осведомленность». Контрольные системы собаки совершенно точно имеют доступ к информации об этом гидранте и могут использовать ее для надлежащего контроля за поведением. В соответствии с принципом когерентности выглядит вероятным, что гидрант присутствует в *опыте* собаки, примерно так, как мир присутствует в нашем визуальном опыте. Это согласуется со здравым смыслом; и в данном случае я лишь делаю это умозаключение здравого смысла чуть более развернутым.

То же самое, по-видимому, верно для мышей и даже для мух. Мухи обладают каким-то перцептивным доступом к окружению, и получаемые ими перцептивные данные, похоже, проникают в их когнитивные системы и становятся доступными для регулирования поведения. Кажется разумным допустить, что это не что иное, как осведомленность, и что, по принципу когерентности, здесь имеется некий сопровождающий ее опыт. Но где-то на этом уровне вопрос становится более запутанным. Заманчиво опустить когерентность и дальше вниз по шкале обработки информации; но рано или поздно понятие «осведомленности» иссякнет и не сможет выполнять объяснительной работы из-за своей неопределенности. Пока я не буду больше рассуждать на эту тему, но в дальнейшем еще вернусь к ней.

Глава 7

Отсутствующие квалиа, блекнувшие квалиа, скачущие квалиа

1. Принцип организационной инвариантности

Если сознание возникает из физического, то благодаря какого рода физическим свойствам оно возникает? Судя по всему, речь идет о свойствах, которые могут быть реализованы в мозге, но не очевидно, какие именно эти свойства. Одни полагали, что это биохимические свойства; другие — что квантовые; многие высказывали неуверенность на этот счет. Естественно предположить, что сознание возникает вследствие определенной *функциональной организации* мозга. Согласно этой позиции, химический, да и квантовый субстрат мозга не имеют отношения к порождению сознания. Важна абстрактная каузальная организация мозга, которая могла бы реализовываться на множестве различных физических субстратов.

Функциональную организацию лучше всего трактовать в качестве *абстрактной схемы каузального взаимодействия* различных частей системы, а также, возможно, этих частей и внешних данных на входе и выходе. Функциональная организация определяется конкретизацией (1) количества абстрактных компонентов, (2) множества различных возможных состояний для каждого из компонентов и (3) системы отношений зависимости, уточняющих, как состояние каждого из компонентов зависит от предыдущих состояний всех компонентов и от данных на входе в систему, и как данные на выходе системы зависят от предшествующих им состояний ее компонентов. В остальном, то есть помимо конкретизации количества компонентов и состояний и отношений зависимости между ними, природа этих компонентов и состояний остается неспецифицированной.

Физическая система *реализует* данную функциональную организацию, если эта система может быть разделена на надлежащее мно-

жество физических компонентов, каждый из которых наделен надлежащим множеством состояний — таких, что отношения каузальной зависимости между компонентами данной системы и данными на входе и выходе в точности отражают отношения зависимости, представленные в спецификации той функциональной организации. (Более формальное объяснение в подобном ключе можно найти в (Chalmers 1994a, 1994b), оно также суммировано в главе 9 этой книги, но пока хватит и этой неформальной трактовки.)

Наличная функциональная организация может быть реализована разными физическими системами. К примеру, организация, реализованная мозгом на нейронном уровне, могла бы быть в принципе реализована кремниевой системой. Описание функциональной организации мозга абстрагируется от физической природы задействованных частей, и от того, как именно реализуются каузальные соединения. Важно лишь наличие соответствующих частей и отношений зависимости между их состояниями.

Функциональная организация физической системы предстает на множестве разных уровней — в зависимости от того, насколько детально мы индивидуализируем ее части и разделяем состояния этих частей. На грубом уровне, к примеру, два полушария мозга вполне могут рассматриваться как реализующие простую двухкомпонентную организацию, если мы возьмем их надлежащие взаимозависимые состояния. Как правило, однако, полезнее рассматривать когнитивные системы на более детальном уровне. Если нас интересует познание, то мы, скорее всего, сосредоточимся на уровне, достаточно детальном для того, чтобы определять связанные с мозгом поведенческие способности, причем поведение при этом индивидуализируется с необходимым уровнем точности. Слишком грубый уровень организации (вроде указанной выше двухкомпонентной организации) не позволит определять поведенческие способности, так как механизмы, порождающие поведение, не попадут в ее описание; простая система могла бы иметь ту же самую организацию при ином поведении. Но на более детальном уровне — возможно, нейронном — функциональная организация будет определять поведенческие способности. Даже если бы наши нейроны были заменены кремниевыми чипами, то, если бы эти чипы были наделены состояниями с теми же самыми схемами каузальных взаимодействий, что и у нейронов, эта система порождала бы точно такое же поведение.

В последующем речь всегда будет идти о таком уровне детализации функциональной организации системы, которого будет достаточно для определения поведенческих способностей. Назовем такую организацию *высокодетализированной функциональной организацией*. В иллюстративных целях я обычно буду ссылаться на нейронный уровень

организации мозга, хотя для этого мог бы оказаться достаточным и более высокий уровень; не исключено также, что потребовался бы более низкий уровень. В любом случае, аргументы имеют общее значение. Для целей последующего рассмотрения мы должны также допустить, что для совпадения функциональной организации систем они должны находиться в соответствующих состояниях в надлежащее время; хотя мой спящий двойник в широком смысле слова мог бы считаться совпадающим по своей функциональной организации со мной, его нельзя будет счесть таковым в строгом смысле, который будет требоваться в дальнейшем. Если две системы одинаковы по своей функциональной организации в этом строгом смысле, то я буду говорить, что они являются *функциональными изоморфами*.

Мой тезис в том, что сознательный опыт порождается высокодетализированной функциональной организацией. Если более конкретно, то я буду обосновывать *принцип организационной инвариантности*, согласно которому, если имеется система, обладающая сознательными переживаниями, то любая система с точно такой же высокодетализированной функциональной организацией будет обладать качественно тождественными переживаниями. В соответствии с этим принципом, сознание — это организационный инвариант: свойство, остающееся неизменным у всех функциональных изоморфов данной системы. Не имеет значения, реализована ли эта организация в кремниевых чипах, народонаселении Китая, пивных банках или теннисных шариках. Если эта функциональная организация такова, какой она должна быть, это определит сознательный опыт.

Это утверждение часто ассоциировалось с редуктивным функционалистским взглядом на сознание, в частности, с воззрением, что для наличия сознания нужно лишь исполнение надлежащей функциональной роли. Принцип инвариантности естественным образом вытекал бы из этого воззрения, но он может быть истинным и независимо от последнего. Так же, как можно верить в то, что сознание порождается физической системой, но не является физическим состоянием, можно верить в то, что сознание порождается функциональной организацией, но не является функциональным состоянием. Отстаиваемая мной позиция именно такова — мы могли бы назвать ее *нередуктивным функционализмом*. Ее можно рассматривать в качестве объединения функционализма и дуализма свойства.

В дальнейшем я не буду уделять особого внимания нередуктивным аспектам моей позиции, так как главной моей задачей будет доказательство принципа инвариантности. Мои аргументы могли бы использоваться даже редуктивными функционалистами. Хотя эти аргументы и не дают полностью редуктивных выводов, они тем не менее могут

рассматриваться как доводы, поддерживающие эту позицию в противовес другим редуктивным воззрениям, таким как концепция, согласно которой сознание отождествляется с каким-то биохимическим свойством. Конечно, я считаю, что все редуктивные взгляды в конечном счете неверны, но последующее обсуждение в значительной степени не зависит от этого момента.

По сути, я уже приводил довод в пользу одной из разновидностей принципа инвариантности, когда выдвигал аргумент от «Х-фактора» в главе 6. Но в этой главе я буду использовать мысленные эксперименты, позволяющие обосновывать этот принцип гораздо более прямым путем.

Отсутствующие квалиа и инвертированные квалиа

Принцип инвариантности далеко не общепризнан. Многие авторы как

дуалистической, так и материалистической направленности пытались доказать, что он неверен. Многие считали, что для того чтобы какая-то система была сознательной, она должна обладать надлежащей биохимической организацией; и если так, то металлический робот или кремниевый компьютер никогда не могли бы иметь опытные переживания, невзирая на их каузальную организацию. Другие признавали, что робот или компьютер могли бы обладать сознанием при надлежащей организации, но полагали, что их переживания могли бы существенно отличаться от наших.

Соответственно этим двум позициям обычно выдвигалось два типа аргументов против принципа инвариантности. Первый из них связан с аргументами от *отсутствующих квалиа*. В них описывается какая-нибудь особенно причудливая реализация данной функциональной организации, в системе настолько странной, что естественно предположить *отсутствие* в ней качеств (квалиа) сознательного опыта. Широко известен пример, приводимый Блоком (Block 1978), где речь идет о реализации присущей нам организации народом какой-то страны (как в главе 3). И утверждается, что *такое* наверняка не могло бы порождать сознательный опыт. Значит, сознание не может порождаться функциональной организацией.

Второй класс аргументов включает аргументы от *инвертированных квалиа* или *инвертированного спектра*. Согласно этим аргументам, если бы наша функциональная организация была реализована на ином физическом субстрате, система по-прежнему могла бы иметь опытные переживания, но иного рода. Там, где мы переживаем красное, она могла бы переживать синее и т. д. Зачастую эти аргументы разворачиваются в рамках сложных сценариев, связанных с хирургией мозга, когда в один прекрасный день мы просыпаемся и видим синее, а не красное, хотя наша функциональная организация и не претерпевает каких-либо изменений.

Многие из тех, кто доказывал возможность отсутствующих и инвертированных квалиа, доказывали лишь логическую возможность; этого вполне достаточно для опровержения редуктивной формы функционализма. Да и я сам использовал подобные аргументы в главе 3. Контрагументы в данной главе не нацелены против защитников такой позиции. Спор идет здесь о более слабой разновидности функционализма, которая не затрагивается вопросами о логической возможности.

Центральный вопрос этой главы — это вопрос о том, являются ли отсутствующие или инвертированные квалиа *естественно* (или *эмпирически*) возможными. Логически возможно, что если выпустить тарелку из рук в вакууме или на поверхности планеты, то она полетит вверх, но это тем не менее естественно невозможно. Законы природы запрещают это. Так же и установление логической возможности отсутствующих квалиа и инвертированных квалиа далеко не означает установление их естественной возможности. Принцип инвариантности утверждает, что функциональная организация детерминирует сознательный опыт посредством некоего законосообразного связующего звена в актуальном мире; и логическая возможность не имеет отношения к этому вопросу. Когда в этой главе я говорю просто о «возможности», речь идет о естественной возможности.

В последующем я обсужу аргументы, выдвигавшиеся в защиту естественной возможности отсутствующих и инвертированных квалиа, а затем предложу детальные аргументы *против* таких возможностей. Эти аргументы нельзя будет развернуть без мысленных экспериментов. Против возможности отсутствующих квалиа я выдвину мысленный эксперимент, связанный с *блекнущими квалиа*. Против возможности инвертированных квалиа я выдвину мысленный эксперимент, связанный со *скачущими квалиа*.

Как обычно, указанные аргументы от мысленных экспериментов будут лишь правдоподобными аргументами, но, думаю, они весьма сильны. Для утверждения естественной возможности отсутствующих и инвертированных квалиа перед лицом этих мысленных экспериментов потребуется принятие ряда неправдоподобных тезисов о природе сознательного опыта и, в частности, об отношении между сознанием и познанием. При определенных естественных допущениях касательно этого отношения, принцип инвариантности оказывается наиболее правдоподобной гипотезой.

Быть может, полезно было бы взглянуть на роль этих мысленных экспериментов по аналогии с ролью, которую играл мысленный эксперимент «кошка Шредингера» в интерпретации квантовой механики. Мысленный эксперимент Шредингера не выносит окончательный вердикт в пользу той или иной интерпретации, но он выявляет различные

правдоподобные и неправдоподобные моменты в этих интерпретациях, и любая интерпретация в конечном счете должна принимать это во внимание. Так и любая теория сознания должна в итоге что-то делать со сценариями блекнущих и скачущих квалиа, и одни из них справляются с ними лучше, чем другие. Так будут выявлены достоинства и недостатки различных теорий.

2. Отсутствующие квалиа

Позитивные аргументы в пользу естественной возможности отсутствующих квалиа были не столь многочисленны, как аргументы в пользу инвертированных квалиа, но они все же выдвигались. Наиболее подробное изложение этих аргументов можно найти у Блока (Block 1978).

Эти аргументы почти всегда одинаковы по форме. Они состоят в демонстрации реализации нашей функциональной организации в какой-то необычной среде, сочетающейся с обращением к интуиции. Говорится, к примеру, что организация нашего мозга могла бы быть симулирована народом Китая или даже отражена в боливийской экономике. Если мы сделаем так, что каждый человек в Китае будет симулировать нейрон (для этого нам надо было бы увеличить население Китая в десять или сто раз, но это не имеет значения), и если мы снабдим этих людей радиопередатчиками для симуляции синаптических связей, то данная функциональная организация будет воспроизведена. Но, как утверждается в этом аргументе, эта причудливая система наверняка не будет обладать *сознанием*!

Этот аргумент имеет определенную интуитивную силу. Многие люди склонны считать, что подобная система просто не может быть тем, что могло бы обладать сознательным опытом. Такое «групповое сознание»казалось бы скорее темой для научно-фантастического рассказа, чем объектом, который мог бы реально существовать. Но, кроме интуитивной силы, здесь ничего нет. И это уж точно далеко не решающий аргумент. Многие указывали¹, что, хотя порождение подобной системой сознания и может быть интуитивно неправдоподобным, столь же интуитивно неправдоподобно и то, что мозг должен был бы порождать сознание! Кто мог бы подумать, что этот кусок серой материи мог бы оказаться тем, что смогло бы порождать яркие субъективные переживания? Но он делает это. Разумеется, это не показывает, что народ мог бы порождать сознание, но является сильным противовесом интуитивному доводу, что такое невозможно.

Конечно, в подобной системе мы не увидели бы какого-либо сознательного опыта. Но здесь нет ничего нового; мы не видим сознательного опыта вообще ни у кого. Могло бы показаться, что в подобной

системе нет «места» для сознательного опыта, но, опять-таки, это кажется справедливым и по отношению к мозгу. Наконец, мы могли бы объяснить функционирование данной системы без обращения к сознательному опыту, но это опять же известно нам и в стандартном случае. После того, как мы прочувствуем реальный смысл отсутствия логической супервентности, мысль о том, что население страны могло бы порождать сознательный опыт, будет удивлять нас не больше, чем о том, что мозг мог бы делать это².

Некоторые высказывали возражения против принципа инвариантности на том основании, что данная функциональная организация могла возникнуть случайно, в боливийской экономике или даже в ведре (Hinckfuss, цитируется в (Lycan 1987, с. 32)). Но это могло бы случиться лишь в случае самого что ни на есть невероятного совпадения³. Система должна была бы иметь более миллиарда частей, каждая из которых должна была бы обладать набором собственных состояний (скажем, по десять у каждой). Между этими состояниями должна была бы иметься громадная и запутанная система надлежащих каузальных соединений, такая, что при наличии паттерна *вот такого* состояния появится паттерн такого-то состояния, а при наличии *вот такого* — паттерн вот такого, и т. д. Для реализации той функциональной организации эти кондиционалы не могут быть простыми регулярностями (когда данный паттерн состояния по факту следует за другим паттерном состояния в данном случае); они должны быть прочными связями, сохраняющимися в контрфактических ситуациях, такими, что данный паттерн состояния следовал бы за тем, когда бы тот ни появился⁴.

Нетрудно убедиться, что для реализации такой системой надлежащей функциональной организации, если мы разбиваем ее на миллиард частей, потребовалось бы 10^{10^9} подобных кондиционалов. Шанс того, что эти кондиционалы были бы реализованы произвольной системой при условии принятого нами количества ее частей и состояний, был бы порядка 1 из $(10^{10^9})^{10^{10^9}}$ (на деле гораздо меньше, так как требование прочности каждого кондиционала еще более снижает шанс его реализации)⁵. Даже при учете произвольного деления системы на части реализация такой организации произвольной системой — да и вообще любой системой, не сформированной крайне далеким от произвольности механизмом естественного отбора, — является предельно маловероятной.

Осознав силу ограничений структуры системы спецификацией функциональной организации, мы уже не будем считать столь маловероятным то, что даже население Китая могло бы при надлежащей организации оказываться субстратом сознательного опыта. Если, вообразив это население, ускорить его примерно в миллион раз и сжать

его до размера головы, то получится что-то очень похожее на мозг, за исключением того, что вместо нейронов мозга у нас будут гомункулы — маленькие люди. И кажется, нет серьезных оснований считать, что нейроны непременно будут справляться с задачей фундирования опыта лучше, чем гомункулы.

Разумеется, мы, как указывает Блок, *знаем*, что нейроны могут справиться с этой задачей, тогда как о гомункулах нам ничего не известно. Вопрос, тем самым, остается открытым. Важно, однако, то, что подобные аргументы могут служить лишь очень слабым свидетельством в пользу того, что отсутствующие квалиа являются естественно невозможными. Для решения этого спора нужен более убедительный аргумент. *Не исключено*, что правильно было бы сказать, как и делает Блок, что наши интуиции перекладывают бремя доказательства на тех, кто считает, что квалиа организационно инвариантны, хотя для меня это и не очевидно. Так или иначе, в дальнейшем я взвалю на себя эту ношу.

В пользу естественной возможности отсутствующих квалиа иногда выдвигается и не связанный с теми рассуждениями аргумент, отталкивающийся от феномена слепого зрения. Утверждается, что пациенты со слепым зрением в существенных моментах функционально сходны с нами — они способны к различению, к вербальным отчетам и т. д., — но при этом у них отсутствует визуальный опыт. Значит, функциональная организация зрительных процессов не определяет наличия или отсутствия опыта.

В главе 6 мы, однако, видели, что процессы обработки информации у нормальных субъектов и у субъектов со слепым зрением существенно *различны*. У этих субъектов нет обычного прямого доступа к визуальной информации. Если соответствующая информация вообще доступна, доступ к ней имеет непрямой характер, и к данной информации уж точно нет обычного доступа для использования ее при контроле за поведением. В самом деле, именно из-за этого различия в организации их обработки информации, проявляющегося в их поведении, мы и замечаем нечто необычное, что ведет нас к допущению отсутствия опыта. Так что эти случаи не могут свидетельствовать против принципа инвариантности.

3. Блекнущие квалиа

Мой позитивный аргумент против возможности отсутствующих квалиа будет основан на мысленном эксперименте, предполагающем постепенную замену частей мозга, к примеру, кремниевыми чипами. Подобные мысленные эксперименты нередко использовались для ответа на аргументы об отсутствующих квалиа в устной традиции ис-

следований искусственного интеллекта, а иногда и в печати. Сценарий постепенной замены обсуждал Пилишин (Pylyshyn 1980), хотя он и не связывал с ним систематической аргументации. В чем-то сходные с моим аргументы выдвигались Сэвиттом (Savitt 1982) и Када (Cuda 1985), хотя они развивают их иначе и делают несколько иные выводы из этого сценария⁶.

Этот аргумент «блекнущих квалиа» не будет моим сильнейшим и главнейшим аргументом против возможности отсутствующих квалиа; эту роль играет аргумент «скачущих квалиа», о котором пойдет речь в пятом параграфе и который является также аргументом против возможности инвертированных квалиа. Тем не менее аргумент от блекнущих квалиа сам по себе силен и может серьезно мотивировать нас ко второму, более сильному аргументу, равно как и послужить хорошим фоном для последнего.

Этот аргумент имеет вид *reductio ad absurdum*. Допустим, что отсутствующие квалиа естественно возможны. В таком случае могла бы существовать система с такой же функциональной организацией, как и у какой-нибудь сознающей системы (скажем, у меня), но при этом абсолютно лишенная сознательного опыта. Не забывая об общем характере наших рассуждений, предположим, что это происходит потому, что данная система состоит не из нейронов, а из кремниевых чипов. Позже я покажу, как этот аргумент можно распространить на другие виды изоморфов. Назовем этого функционального изоморфа «Роботом». Каузальные схемы когнитивной системы Робота не отличаются от моих, но он обладает сознанием зомби.

В такой ситуации мы можем сконструировать ряд переходных случаев между мной и Роботом, таких, что каждый шаг будет представлять собой лишь очень небольшое изменение при полном сохранении функциональной организации на протяжении всего ряда. Мы можем вообразить, к примеру, замену какого-то количества моих нейронов кремниевыми чипами. В первом из подобных случаев будет заменен только один нейрон. На его место встанет кремниевый чип, выполняющий точно такую же локальную функцию, что и этот нейрон. Там, где последний соединен с другими нейронами, соединяется с ними и указанный чип. Там, где состояние этого нейрона чувствительно к электрическим стимулам и химическим сигналам, чувствителен к ним и кремниевый чип. Мы могли бы вообразить, что он снабжен маленькими датчиками, принимающими электрические сигналы и химические ионы и преобразующими их в цифровые сигналы, распространяющиеся по всему чипу. Там, где нейрон продуцирует электрические и химические реакции, их продуцирует и чип (мы можем вообразить, что он снабжен маленькими эффекторами, порождающими электрические и химические реакции в зависимости от внутреннего

состояния этого чипа). Важно то, что внутренние состояния этого чипа таковы, что стимульно-реакционная функция чипа ничем не отличается от соответствующей функции нейрона. То, как чип достигает этого результата, не имеет значения — возможно, он использует таблицу соответствия, увязывающую каждый стимул с надлежащей реакцией, или производит вычисления, симулирующие нейронные процессы — главное, чтобы он обеспечивал верные зависимости между данными на входе и данными на выходе. Если локальные взаимодействия адекватны, то интересующая нас замена на изменит общего функционирования системы.

Во втором случае мы заменяем кремниевыми чипами два нейрона. Легче всего предположить, что речь идет о смежных нейронах. При этом, после замены обоих нейронов, мы можем расстаться с неуклюжими датчиками и эффекторами, опосредующими соединение двух чипов. Мы можем заменить их любыми звенями, лишь бы они были чувствительны к внутреннему состоянию первого чипа и надлежащим образом влияли на внутреннее состояние второго чипа (эти соединения могут быть, разумеется, и двунаправленными). Здесь мы добиваемся того, чтобы данное соединение не отличалось от соответствующего соединения у Робота; возможно, оно будет состоять в прохождении какого-то электронного сигнала.

Дальнейшие случаи конструируются по очевидной схеме. В каждом последующем случае кремниевыми чипами заменяется все большая группа смежных нейронов. Внутри этой группы чипов биохимический субстрат полностью исчезает. Биохимические механизмы присутствуют только в остальных частях системы и в сочленениях между чипами на границах этой группы и смежными с ними нейронами. В конце концов все нейроны системы оказываются заменены на чипы, и биохимические механизмы перестают играть какую-либо существенную роль. (Я отвлекаюсь здесь от детального обсуждения того, играют ли, к примеру, глиальные клетки какую-то значимую роль; если да, то они будут компонентами надлежащей функциональной организации и тоже будут заменены.)

Можно вообразить, что эта внутренняя система повсюду сочленена с телом, чувствительна к телесным стимулам и надлежащим образом, через датчики и эффекторы, продуцирует моторные реакции. Каждая из систем в этом ряду функционально изоморфна мне в детализации, позволяющей ей иметь мои поведенческие диспозиции. Но хотя система на одном конце спектра — это я, на другом конце мы видим, по сути, копию Робота.

Чтобы довершить картину, представим, что, в качестве первой системы, я наделен богатым сознательным опытом. Скажем, я нахожусь

на баскетбольном матче в толпе орующих фанатов; я вижу яркие разноцветные одежды, ощущаю приятный аромат каких-то легких закусок, возможно, чувствую пульсирующую головную боль и т. п. Сосредоточимся на ощущениях ярко-красного и желтого, которые возникают у меня, когда я смотрю на форму игроков. Финальная система, Робот, находится в такой же ситуации, обрабатывает такие же данные на входе и продуцирует сходное поведение, но, по предположению, вообще ничего не переживает.

Между мной и Роботом будет множество промежуточных инстанций. Вопрос: *каково это — находиться в них?* Что переживают такие системы, если они вообще что-то переживают? Как меняется сознательный опыт по мере нашего смещения вдоль указанного спектра? Можно предположить, что в самых первых случаях их переживания во многом похожи на мои, а в самых последних у них нет переживаний или есть лишь очень незначительные переживания, но как обстоит дело с промежуточными случаями?

Учитывая, что система на другом конце спектра (Робот) лишена сознания, кажется, что по ходу дела должно происходить одно из двух. Либо (1) сознание постепенно блекнет перед тем, как в итоге исчезнуть, либо (2) где-то в этом ряду сознание неожиданно выключается, хотя в предыдущем случае был богатый сознательный опыт. Назовем первую возможность *блекнущими квалиа*, вторую — *внезапно исчезающими квалиа*.

Нетрудно исключить вариант с внезапно исчезающими квалиа. Согласно этой гипотезе, замена одного-единственного нейрона (при том, что все остальное остается неизменным) могла бы быть ответственна за исчезновение всего поля сознательного опыта. Это кажется крайне неправдоподобным, если не абсолютно фантастичным. Если бы такое было возможно, то в законах природы были бы грубые разрывы, отличные от того, что мы обнаруживаем во всех других случаях. Любой конкретный момент внезапного исчезновения квалиа (50 процентов нейронов? 25 процентов?) был бы совершенно произвольным. Можно даже представить, что наш мысленный эксперимент разворачивается на более детальном уровне, внутри нейрона, и тогда в конечном счете окажется, что исчезновение всего поля опытных переживаний будет вызвано заменой нескольких молекул (если нет, то мы возвращаемся к сценарию с блекнущими квалиа). Как всегда в подобных случаях, эта гипотеза не может быть опровергнута, но ее изначальная правдоподобность невелика.

(Можно было бы возразить, что в нелинейной динамике бывают ситуации, когда одна величина чувствительно зависит от другой и большие изменения первой могут быть вызваны незначительными изменениями второй. Однако в этих ситуациях зависимость все же имеет

непрерывный характер, так что здесь будут промежуточные случаи, в которых зависимая величина принимает промежуточные значения; поэтому данная аналогия ведет к блекнущим квалиа. И в любом случае подобная чувствительная зависимость обычно возникает как результат совокупных воздействий множества более фундаментальных градуальных зависимостей. Во всех известных до сих пор фундаментальных законах зависимость одной величины от другой непрерывной величины подобным образом непрерывна, и из непрерывности никоим образом нельзя образовать ее отсутствие. Внезапно исчезающие квалиа, в отличие от нелинейной динамики, требовали бы существования грубых разрывов в фундаментальных законах.)

Если внезапно исчезающие квалиа исключены, то мы остаемся с блекнущими квалиа. Чтобы разобраться в этом сценарии, рассмотрим систему, находящуюся на середине спектра между мной и Роботом, в которой сознание уже серьезно деградировало, но еще не полностью исчезло. Назовем эту систему «Джо». Каково это — быть Джо? Джо, разумеется, функционально изоморфен мне. Он *говорит* о своих переживаниях все то, что я говорю о своих. На баскетбольном матче он не удерживается от восклицаний по поводу ослепительно яркой красно-желтой формы баскетболистов.

Согласно допущению, однако, Джо вообще лишен переживаний ярких красных и желтых цветов. Быть может, вместо них он ощущает что-то розоватое и тускло-коричневое. Или самые бледные оттенки красного и желтого. Или его ощущения затемняются почти до черноты. Можно представить много вариантов постепенного превращения ощущений красного в полное отсутствие ощущений, а еще большее число таких вариантов мы, быть может, даже не можем себе представить. Но можно предположить, что в каждом из них перед своим уничтожением ощущения должны утрачивать яркость (в ином случае мы останемся с проблемой внезапно исчезающих квалиа). Сходным образом можно предположить, что на каком-то этапе тонкие различия моих переживаний должны исчезать из опыта промежуточной системы; признавая, что все различия моих переживаний сохраняются вплоть до момента их одновременного исчезновения, мы остались бы с еще одной версией внезапно исчезающих квалиа.

Для конкретности вообразим, что Джо видит что-то розоватое там, где я вижу ярко-красное, причем множество различий оттенков моих переживаний больше не представлено в оттенках его переживаний. Громкий шум для меня, для Джо, возможно, лишь далекий рокот. Но не все так плохо для Джо: я чувствую спазмы в голове, а он всего лишь покалывание.

Существенный момент состоит здесь в том, что Джо систематически *ошибается* относительно всего, что он переживает. Конечно, он

говорит, что он ощущает что-то ярко-красное и ярко-желтое, но в действительности он ощущает лишь розоватое⁷. Если вы спросите его, он заявит, что различает в своем опыте всевозможные тонкие оттенки красного, но на самом деле многие из них совершенно неразличимы в его переживаниях. Он может даже жаловаться на шум, хотя в реальности в его слуховом опыте нет ничего подобного. Хуже того, при функциональном конструировании убеждений, Джо будет даже *верить* в то, что у него имеются все те комплексные переживания, которые у него в действительности отсутствуют. Одним словом, Джо до крайности оторван от своего сознательного опыта, и он не способен войти в какой-либо контакт с последним.

Это кажется совершенно неправдоподобным. Здесь мы видим существование с функционирующими рациональными механизмами и реально обладающее *сознанием*, но абсолютно неверно трактующее свои собственные сознательные переживания. Быть может, если говорить о предельном случае, когда все темно внутри, предположение о том, что подобная система могла бы столь ошибаться в своих заявлениях и суждениях, и могло бы оказаться небезосновательным — в конце концов, в известном смысле в этой системе вообще не было бы того, кто мог бы ошибаться. Но в промежуточном случае это гораздо менее правдоподобно. Во всех известных нам случаях сознательные существа, как правило, способны формировать точные суждения о своих переживаниях, когда их ничего не отвлекает и если они разумны. Если здравое, разумное существо, лишенное какой-либо функциональной патологии, до такой степени систематически оторвано от своих переживаний, то это означает, что сознание очень слабо связано с познанием. У нас нет серьезных оснований полагать, что сознание — настолько аномальный феномен, и есть весомое основание считать иначе.

Конечно, блекнущие квалиа *логически* возможны. В описании системы, настолько ошибающейся относительно своих переживаний, нет противоречия⁸. Но логическая возможность и естественная возможность — разные вещи. У нас нет оснований считать, что подобный случай можно встретить на практике, и самые серьезные основания считать иначе. Одним из наиболее бросающихся в глаза фактов о сознании кажется тот факт, что когда сознательное существо с надлежащей детализацией его понятий переживает что-то, оно способно сформировать суждения о своих переживаниях. Не исключено, что в некоторых необычных случаях рациональные процессы системы оказываются существенно нарушены, и это приводит к сбоям в механизмах суждения, но это не тот случай. Процессы *протекают* у Джо так же, как и у меня — согласно предположению, он является моим функциональным изоморфом. Дело лишь в том, что, как оказывается, он совершенно неверно трактует собственные переживания.

Разумеется, в повседневной жизни можно отыскать немало различных примеров блекнущих квалиа. Подумаем, к примеру, о том, что происходит, когда кто-то погружается в сон; или представим как мы смещаемся назад по эволюционной цепи от людей к трилобитам. В каждом из таких случаев по мере продвижения вдоль спектра сознательный опыт постепенно блекнет. Но в каждом из этих случаев этот феномен сопровождается соответствующими функциональными изменениями. Когда я впадаю в сонное состояние, я не верю в то, что я абсолютно не хочу спать и что-то ярко переживаю (если только я не заснул — и тогда я вполне вероятно что-то ярко переживаю). Бедность цветового опыта собаки сопряжена с соответствующим дефектом ее различающих визуальных механизмов. Эти случаи очень непохожи на тот, который мы рассматриваем, так как в последнем переживания блекнут, а функционирование остается неизменным. Механизмы Джо по-прежнему могут отмечать тонкие различия длины световых волн, и он, конечно, утверждает, что подобные различия отражаются в его опыте, но мы должны верить, что это совершенно не так.

Серл (Searle 1992) обсуждает сходный мысленный эксперимент и рассматривает возможность следующего развития событий:

По мере того, как кремний постепенно имплантируется в ваш ухудшающийся мозг, вы находитите, что область вашего сознательного опыта сужается, но это не оказывает действие на ваше внешнее поведение. И вы находитите, к своему полнейшему изумлению, что вы и в самом деле теряете контроль над своим внешним поведением. Вы, к примеру, обнаруживаете, что когда врачи проверяют ваше зрение, вы слышите, как они говорят: «Мы показываем красный объект напротив вас; скажите нам, пожалуйста, что вы видите». Вам хочется закричать: «Я же ничего не вижу. Я совершенно ослеп». Однако вы слышите, как ваш голос говорит так, будто он совершенно не поддается вашему контролю: «Перед собой я вижу красный объект». И если мы до предела доведем этот эксперимент, то получим еще более удручающий результат, чем в прошлый раз. Мы воображаем, что ваш сознательный опыт медленно сужается до ничего, в то время как внешне наблюдаемое поведение остается тем же самым (с. 66–67; пер. А. Ф. Грязнова).

Здесь Серл признает возможность блекнущих квалиа, но допускает, что подобная система может и не ошибаться относительно ее убеждений по поводу ее опыта. Система могла бы обладать истинными убеждениями о своем опыте; дело лишь в том, что они не могли бы влиять на ее поведение.

Думаю, однако, что такую возможность можно исключить. В данной системе попросту нет места для формирования каких-либо новых убеждений. Если не брать очень сильный дуализм, то подобное различие убеждений должно отражаться в функционировании системы — *может*, и не в поведении, но хотя бы в каком-то процессе. Но эта

система идентична изначальной системе (мне) на надлежащем уровне детализации. В ней просто нет места для таких новых убеждений, как «я ничего не вижу», новых желаний вроде желания закричать и других новых когнитивных состояний, таких как изумление. Ничто в этой физической системе не может соответствовать этому изумлению. Для него нет места в нейронах, которые, в конце концов, идентичны группе нейронов, фундирующей обычные убеждения; и Серл уж точно не хочет сказать, что кремневый суррогат сам по себе фундирует новые убеждения! Если только между нейронами и кремнием нет какого-то поразительного, магического взаимодействия — которое к тому же не проявляется в других процессах, так как организация в целом остается неизменной — такие новые убеждения не возникнут.

Не меняющий организацию переход от нейронов к кремнию по-просту не производит изменений, достаточных для того, чтобы вызвать столь заметную трансформацию в содержании и структуре когнитивных состояний. Переключение с опыта красного на опыт синего — это одно, но совсем другого порядка был бы переход от убеждения «Замечательный баскетбольный матч» к убеждению «О нет! Меня, кажется, занесло на дурацкий фильм ужасов». Если бы столь сильное изменение когнитивных данных не отражалось в изменениях функциональной организации, познание было бы совершенно оторвано от выполнения внутренних функций подобно бестелесному картезианскому духу. Если содержательные аспекты когнитивных состояний вообще были бы супервентными на физических состояниях, они могли бы оказываться таковыми только в силу самых что ни на есть произвольных и капризных правил (если это сделано из нейронов, то «Приятные цвета!», если из кремния, то «Увы!»).

Таким образом, возможность блекнущих квалиа предполагает либо причудливое отношение между содержанием убеждений и физическими состояниями, либо возможность существ, колоссально заблуждающихся по поводу своих собственных сознательных переживаний, несмотря на то, что они являются в полной мере разумными. Обе эти гипотезы гораздо менее правдоподобны, чем гипотеза о том, что разумные сознательные существа, как правило, точны в своих суждениях о собственных переживаниях. Гораздо более правдоподобной поэтому представляется та гипотеза, согласно которой при замене нейронов квалиа вообще не блекнут. Переживания системы, подобной Джо, на практике будут не беднее, чем у меня. И, если так, то наше изначальное допущение было неверным, и наш исходный изоморф, Робот, наделен сознательными переживаниями.

Этот аргумент можно напрямую перенести и на других функциональных изоморфов. Если говорить о ситуации, когда население страны

реализует мою организацию, то мы можем сконструировать сходный спектр промежуточных случаев между моим кремниевым изоморфом и этим населением. Вначале мы можем постепенно раздвинуть кремниевую систему до размеров нескольких квадратных миль. Мы также замедлим ее, чтобы можно было справиться с частотой получающих сигналы чипов. После этого мы сделаем так, что люди по одному будут занимать место чипов, и убеждаться, что они выдают надлежащие сигналы в ответ на поступающие данные. В итоге мы получим ситуацию, в которой все население будет организовано так же, как были организованы мои нейроны; не исключено, что оно даже будет контролировать радиоуправляемое тело. На каждом этапе система будет функционально изоморфна моей, и к ней будет применим в точности такой же аргумент. Либо сознательный опыт будет сохранен, либо он поблекнет, либо внезапно исчезнет. Две последние возможности столь же неправдоподобны, как и раньше. Мы можем заключить, что система, созданная населением, будет фундировать сознательные переживания в точно такой же степени, как и мозг.

То же самое можно проделать с любой функционально изоморфной системой, включая те, что отличаются по форме, размеру, скорости, физическому составу и т. п. Во всех случаях результат будет тем же. Если подобная система не обладает сознанием, то существует такая промежуточная изоморфная система, которая обладает сознанием, имеет поблекшие переживания и совершенно неверно трактует последние. Если мы не готовы допускать такой большой разрыв между сознанием и познанием, то получается, что изначальная система все же должна была обладать сознанием.

Если отсутствующие квалиа возможны, то возможны и блекнущие квалиа. Но выше я доказывал, что блекнущие квалиа почти наверняка невозможны. Значит, отсутствующие квалиа почти наверняка невозможны.

Теперь я рассмотрю различные возражения на этот аргумент.

Возражение 1:

Заменить нейроны на практике невозможно

Людей практического склада могла бы мало впечатлить эта методология мысленных экспериментов. Они могли бы возразить, что замена нейронов на кремниевые чипы — тема для научно-фантастических историй, а не то, что имеет отношение к реальности. В частности, они могли бы возразить, что подобная замена была бы невозможна на практике, так что любые выводы на ее основе не отражали бы реалий этой ситуации.

Если предполагается, что речь идет лишь о *технологической* невозможности, то это не такая уж большая проблема. Нас интересует,

какие переживания были бы у таких систем, *если бы* они существовали, и неважно, можем ли мы в действительности создать их. Хотя естественная невозможность могла бы иметь отношение к делу. Быть может, кремний просто не может выполнять те функции в мозге, которые выполняет нейрон, и поэтому никакой кремниевый чип не мог бы справиться с поставленной задачей. Неясно, можно ли обосновать это возражение; мы уже располагаем протезами рук и ног, скоро появятся и протезы глаз, и тогда почему не могут существовать протезы нейронов? В любом случае, даже если бы кремниевый функциональный изоморф оказался невозможным (быть может, из-за невычислимости нейронной функции?), это не затронуло бы аргумент в пользу принципа инвариантности. Принцип инвариантности говорит лишь о том, что *если* существует функциональный изоморф сознательной системы, *то* он будет иметь такие же осознанные переживания. Если кремниевые изоморфы невозможны, оценочные рассуждения о кремниевых системах попросту нерелевантны в данном случае.

Оппонент мог бы попробовать сосредоточиться на проблемах кремниево-нейронного *интерфейса*, когда дело могло бы обстоять так, что чисто нейронные системы и чисто кремниевые системы были бы вполне возможными, но промежуточные системы оказывались бы под вопросом. Быть может, в крошечных чипах просто не хватало бы места для датчиков и эффекторов? В конце концов, эффекторам надо было где-то хранить химические вещества, чтобы их можно было бы высвобождать при необходимости. Но ведь нам нужен очень маленький резервуар; для аргумента требуется лишь изоморфизм на протяжении нескольких секунд! И мы всегда могли бы провести мысленный эксперимент, предположив расширение этой системы. В любом случае, трудно понять, как такого рода моменты могли бы подкрепить глубокие и основательные возражения против принципа инвариантности. Скорее всего, отыщутся *какие-то* системы, которые можно будет путем постепенной замены превращать друг в друга; будет ли критик доказывать, что принцип инвариантности значим только для них? Если да, то это выглядит совершенно произвольным; если нет, то должно существовать какое-то более глубокое возражение.

Это возражение отталкивается от того обстоятельства, что в некоторых случаях субъекты действительно совершенно неверно трактуют свой опыт. К примеру, при отрицании слепоты люди убеждены, что имеют визуальные переживания, хотя они, похоже, отсутствуют. В подобных случаях, однако, мы не имеем дела с полностью разумными системами. В системах с нарушенными механизмами формирования убеждений

Возражение 2:
Некоторые системы
совершенно неверно
трактуют свой опыт

возможно все. Такие системы могли бы считать себя Наполеоном или верить в то, что Луна розовая. А вот мой «поблекший» изоморф Джо совершенно разумен, и его когнитивные механизмы работают не хуже моих. В разговоре он кажется абсолютно вменяемым. Мы не можем указать на какие-то необычно слабые выводные связи между его убеждениями или на какой-либо систематический психиатрический дефект, приводящий к ошибочным рассуждениям. Джо — в высшей степени осмысленная, разумная личность, не выказывающая никаких симптомов конфабуляции, характерных для тех, кто отрицает слепоту. Эти случаи, таким образом, не похожи друг на друга. Правдоподобно не то, что никакая система не может кардинально заблуждаться относительно своих переживаний, а то, что подобным образом не может заблуждаться рациональная система с нормально функционирующими когнитивными механизмами. Нет сомнений, что Джо — это рациональная система, механизмы которой работают не хуже, чем у меня, и поэтому это не затрагивает наш аргумент.

Возражение 3:

Аргументы с соритами подозрительны

Некоторые возражают, что этот аргумент имеет форму сорита или «скользкой горки» и что такие аргументы обычно сомнительны. С помощью соритов можно «показать», что даже крупица песка — это куча; в конце концов, миллион песчинок образует кучу, а если мы уберем одну песчинку из кучи, мы все равно останемся с кучей. Эта реакция, однако, основана на поверхностном прочтении данного аргумента. Аргументы, в которых используются сориты, обычно достигают своих целей, игнорируя тот факт, что та или иная кажущаяся дилемма в действительности представляет собой континуум: между кучами и не-кучами, к примеру, имеется множество неопределенных промежуточных случаев. Мой же аргумент недвусмысленно признает возможность континуума, но показывает, что промежуточные случаи невозможны по независимым причинам.

Аргумент был бы соритом, если бы он имел такой вид: я обладаю сознанием; если заменить один нейрон сознательной системы на кремниевый чип она по-прежнему будет обладать сознанием; поэтому система, состоящая из одного кремния, будет обладать сознанием. Но он имеет иную форму. Верно, что аргумент против внезапно исчезающих квалиа основан на невозможности внезапного перехода, но важно то, что он нацелен на демонстрацию невозможности внезапных больших переходов, от богатого сознательного опыта к его полному отсутствию. Это неправдоподобно по причинам, совершенно не связанным с выкладками соритов⁹.

Четвертое возражение указывает на то, что этот аргумент доказывает слишком много. Если он устанавливает принцип организационной инвариантности, то сходный аргумент установил бы принцип *поведенческой* инвариантности. Для достижения этой цели мы сконструировали бы континуум случаев, простирающийся от меня до любой поведенчески эквивалентной со мной системы. И сходное рассуждение привело бы нас к выводу, что такая система должна была бы обладать сознанием. Похоже, однако, что некоторые из подобных систем, к примеру гигантская таблица соответствий, содержащая данные о реакциях на любые вводимые паттерны, которую описывает Блок (Block 1981), лишены сознания. Поэтому в рассматриваемом нами аргументе должен быть какой-то пробел.

Это возражение не проходит по двум причинам. Во-первых, мой аргумент отчасти опирается на тот факт, что функционально изоморфная система будет обладать той же когнитивной структурой, что и моя, и, в частности, будет формировать те же суждения. Именно это привело нас к выводу, что блеклая система Джо должна кардинально ошибаться в своих суждениях. Аналогичный тезис неверен по отношению к поведенчески эквивалентным системам. Идеальный актер не обязан формировать те же суждения, что и я. Так же и таблица соответствий и промежуточные системы. Они будут работать, используя совершенно другие механизмы.

Во-вторых, далеко не очевидно, как можно было бы малыми шагами переходить от меня к произвольному поведенческому изоморфу при полном сохранении поведенческой эквивалентности. Как это можно было бы, к примеру, сделать с таблицей соответствий? Быть может, это можно было бы осуществить резкими большими шагами, но этого будет недостаточно для аргумента: если между смежными системами надо делать большие шаги, то внезапно исчезающие квалиа уже не кажутся столь невероятными. В случае функциональных изоморфов можно было естественным образом продвигаться очень малыми шагами, но для поведенческих изоморфов подобного естественного метода не существует. Кажется поэтому маловероятным, что подобный аргумент имеет хоть какие-то перспективы.

В общем, думаю, что единственной осмысленной реакцией на этот аргумент со стороны оппонента организационной инвариантности было бы смириться с возможностью блекнущих квалиа и признать вытекающую из нее возможность того, что разумная сознательная система могла бы кардинально заблуждаться относительно своего опыта, или,

Возражение 4:

Сходные аргументы могли бы приводить к выводу о поведенческой инвариантности

быть может, смириться с другой возможностью и допустить внезапно исчезающие квалиа и связанные с этим резкие разрывы. Эта позиция малопривлекательна, так как она предполагает разрыв между сознанием и познанием, альтернатива которого выглядит гораздо более правдоподобно; но, в отличие от других возражений, она не является очевидно неверной. Впрочем, аргумент от скачущих квалиа в параграфе 5 предоставит дополнительные свидетельства против возможности отсутствующих квалиа, так что противникам принципа инвариантности не стоит рассчитывать на спокойную жизнь.

Мимоходом отмечу, что, используя похожий аргумент, можно было бы показать, что системы, функциональная организация которых *сходна* (но не идентична) с организацией сознательной системы, будут обладать сознательным опытом. Принцип инвариантности сам по себе совместим с солипсистским тезисом о том, что опыт порождается моей организацией и только ей. Но так же, как мы воображали постепенное изменение моего физического состава, мы можем вообразить постепенное изменение моей организации, при котором мои убеждения о моем собственном опыте останутся по большей части неизменными, я останусь разумной системой и т. д. По соображениям, сходным с теми, что приводились выше, кажется очень вероятным, что сознательный опыт сохранится при таком переходе.

4. Инвертированные квалиа

Аргумент от блекнущих квалиа указывает на то, что мои функциональные изоморфы будут обладать сознательным опытом, но он не позволяет заключить, что изоморфы будут обладать сознательным опытом *того же* типа. Иными словами, функциональная организация определяет наличие или отсутствие сознательного опыта, но она могла бы не определять природу этого опыта. Для заключения о том, что функциональная организация определяет природу опыта, нам предстоит показать невозможность функциональных изоморфов, обладающих *инвертированными* квалиа.

Идея инвертированных квалиа известна большинству из нас. Мало кто в какой-то момент не задумывался, не могло ли быть так, что то, что кажется одному красным, выглядит для другого как синее, и наоборот. Это одна из тех философских головоломок, где вначале даже неясно, имеет ли соответствующая идея вообще хоть какой-то смысл и которые могут вызывать затруднения даже после продолжительных размышлений.

Похоже, что впервые возможность инвертированных квалиа была рассмотрена Джоном Локком в его «Опыте о человеческом разумении»:

Впрочем, идея голубого, имеющаяся у одного человека, может отличаться от этой идеи у другого. В наших простых идеях не было бы ничего от ложности и в том случае, если бы вследствие различного строения наших органов было бы так определено, что *один и тот же предмет* в одно и то же время *производил бы в умах нескольких людей различные идеи*; например, если бы идея, вызванная фиалкой в уме одного человека при помощи его глаз, была тождественна с идеей, вызванной в уме другого ноготками, и *наоборот*. Ведь этого никогда нельзя было бы узнать, потому что ум одного человека не может перейти в тело другого, чтобы воспринять, какие представления вызываются с помощью органов последнего; и потому не перепутались бы ни идеи, ни имена и ни в тех, ни в других не было бы никакой ложности. В самом деле, если все вещи, имеющие строение фиалки, будут постоянно вызывать в ком-нибудь идею, которую он назовет «голубое», а все вещи, имеющие строение ноготков, будут постоянно вызывать идею, которую он также постоянно будет называть «желтое», то, каковы бы ни были эти представления в его уме, он будет в состоянии так же правильно различать по ним вещи для своих надобностей и понимать и обозначать эти различия, отмеченные именами «голубое» и «желтое», как если бы эти представления или идеи в его уме, полученные от этих двух цветков, были совершенно тождественны с идеями в умах других людей (кн. 2, гл. 32, пар. 15; пер. А. Н. Савина).

Здесь Локка интересуют инвертированные квалиа скорее у систем с сходным поведением, чем у точных функциональных изоморфов. Кажется несомненным, что тут фиксируется концептуальная возможность. Нас интересует вопрос, идет ли при этом речь и об эмпирической возможности.

Даже те люди, которые считают себя материалистами, нередко полагали, что функциональные изоморфы могли бы иметь иные осознанные переживания. Так, нередко признается естественно возможным, что мой функциональный изоморф с иным физическим составом мог бы ощущать синее там, где я ощущаю красное, или что-то в этом роде. Это и есть гипотеза инвертированных квалиа. Если она верна, то, хотя наличие сознательного опыта могло бы зависеть только от функциональной организации, характеристики опытных переживаний зависели бы от физиологического состава или какого-то другого нефункционального фактора¹⁰.

Ранее мы видели, что эту позицию трудно согласовать с материализмом. Если наличие у функционального изоморфа инвертированных квалиа является естественно возможным, то оно логически возможно. Значит, в равной степени логически возможно и наличие инвертированных квалиа у физического изоморфа, так как концептуальная связь между нейронами и конкретным видом квалиа существует не в большей степени, чем концептуальная связь между кремнием и конкретным видом квалиа. Из этого следует, что характер конкретных опытных переживаний представляет собой некий дополнительный

факт, выходящий за пределы физических фактов, и что материализм должен быть ложен (если только не принять вариант «сильной метафизической необходимости»). В дальнейшем, однако, я абстрагируюсь от этого обстоятельства. Мое обсуждение не будет зависеть от истинности материализма или дуализма.

Возможность инвертированных квалиа, или «инвертированного спектра», как ее порой называют, иногда подвергается сомнению по верификационистским основаниям, состоящим в том, что мы никогда не могли бы узнать о том, что происходит что-то иное, так что здесь могло бы и не быть реального различия (напр., Schlick 1932). Очевидно, что я не принимаю эти доводы: одного лишь факта невозможности сказать, какие именно квалиа переживает система, недостаточно для заключения, что этот вопрос не имеет фактического характера, так как природа квалиа концептуально не связана с поведением; поэтому здесь я не буду принимать во внимание подобные возражения. Как уже отмечалось в главе 3, эту гипотезу иногда критикуют также и по тем соображениям, что наше цветовое пространство ассиметрично, так что никакая инверсия не могла бы установить надлежащие соответствия (напр., Hardin 1987; Harrison 1967, 1973). Некоторые из прежних моих ответов применимы и здесь, хотя теперь нас интересует естественная возможность; так, мы по-прежнему можем апеллировать к возможности существа с ассиметричным цветовым пространством и спрашивать о возможности их инвертированных функциональных изоморфов. В любом случае, я буду игнорировать эту трудность, допуская в целях аргумента, что наше цветовое пространство симметрично, и доказывая, что инвертированные квалиа все равно невозможны.

Обсуждение инвертированных квалиа чревато путаницей. Когда я говорю «переживание синего», имею ли я в виду (1) то, что *субъект* называет переживанием «синего», (2) переживание, вызываемое *синим* объектом, или же (3) то, что я называю переживанием «синего»? Я выбираю третий вариант. В ходе всего обсуждения под «переживанием синего» будет пониматься такое переживание, которое я называю «синим», которое обычно имеется у меня, когда я смотрю на синие объекты — небо, море и т. д. При таком словоупотреблении представимо, что другие (или даже я сам в будущем) могли бы иметь такие переживания синего, которые были бы вызваны желтыми объектами, объектами, называемыми ими «красными» и т. п.

Аргументы в пользу инвертированных квалиа, как и в случае с отсутствующими квалиа, зачастую состоят просто в тезисе о представимости, хотя очевидно, что такие тезисы оставляют открытым вопрос о естественной возможности. Выдвигалась и пара аргументов в пользу естественной возможности разного рода инвертированных квалиа,

но ни один из этих аргументов не опасен для принципа организационной инвариантности.

Первый из этих аргументов позволяет сделать вывод о возможности инвертированных квалиа при неизменности *поведения* (см. Gert 1965; Lycan 1973 и Wittgenstein 1968). Вначале мы отмечаем, что квалиа могли бы инвертироваться в субъекте, возможно, путем переброски соединений между сетчаткой и центральными участками моего мозга, пока я сплю. После пробуждения я скажу, что небо неожиданно стало казаться красным, трава — желтой и т. д., и у нас будут все основания считать, что мои квалиа оказались инвертированными. После этого мы рассматриваем возможную ситуацию, когда мозг какого-то человека был переделан подобным образом при рождении. Квалиа такого человека могли бы быть систематически инвертированными по отношению к норме, но, разумеется, он приучился бы называть небо голубым, траву — зеленою и т. д., так что подобное инвертирование могло бы никогда не отразиться на его поведении¹¹.

Это, однако, не позволяет заключить о возможности инвертированных квалиа при фиксированной функциональной организации. Чтобы убедиться в этом, достаточно обратить внимание на то обстоятельство, что при переброске моих соединений демон серьезно *изменил* мою функциональную организацию. Аналогичным образом изменилась и функциональная организация субъекта, у которого подобные изменения были произведены при рождении, так что он не будет моей функциональной копией. Он может разделять со мной какие-то из моих функциональных свойств на грубом уровне, но он совершенно точно не будет иметь одинаковую со мной высокодетализированную функциональную организацию. Значит, такие случаи не угрожают принципу организационной инвариантности¹².

Родственный аргумент, выдвинутый Блоком (Block 1990), отсылает к «Инвертированной Земле», где небо желто, трава красна и т. п.¹³ Мы можем предположить, что я схвачен и перенесен на Инвертированную Землю, но при этом на меня надеты инвертирующие цвета контактные линзы, так что все кажется мне нормальным. Блок использует этот сценарий для выдвижения аргументов против репрезентационалистского взгляда на квалиа, приравнивающего, к примеру, переживание синего к перцептивному состоянию, связанному с синими объектами (через какое-то время, проведенное на Инвертированной Земле, переживания синего будут связаны у такого человека с *желтыми* объектами). Он также использует его против функционализма такого рода, при котором переживания синего отождествляются с состояниями, порождаемыми синими объектами.

Этот случай опять-таки никак не может опровергнуть принцип организационной инвариантности. В конце концов, когда на Инвер-

тированной Земле я смотрю на желтое небо через мои инвертирующие линзы, моя внутренняя функциональная организация не будет отличаться от той, которая имеется у меня, когда я смотрю на синее небо на Земле, не будет различий и в моем опыте, в точном соответствии с предсказаниями этого принципа. В лучшем случае этот мысленный эксперимент позволяет оторвать переживания от свойств нашего окружения и от «широких» функциональных свойств, включающих *окружение*; но я говорю только о внутренней функциональной организации. Данный мысленный эксперимент не указывает случаи наличия различных переживаний у двух внутренне изоморфных систем, так что он не затрагивает принципа инвариантности¹⁴.

5. Скачущие квалиа

Кто-то мог бы подумать, что можно напрямую адаптировать аргумент блекнущих квалиа для получения аргумента против возможности инвертированных квалиа. К сожалению, это не проходит. Представим, как мог бы разворачиваться подобный аргумент. Мы начинаем с меня, ощущающего что-то красное, и инвертированной системы, ощущающей синее. Путем постепенной замены мы конструируем ряд случаев, в каждом из которых имеется какой-то промежуточный цвет. Но в этой картине нет ничего неверного! Промежуточные системы — это попросту случаи мягкого инвертирования квалиа и не более проблемны, чем предельный случай.

Конечно, может быть не очевидным, что именно ощущают эти промежуточные системы. Возможно, никакой из цветов нашего обычного цветового пространства не может подходить на эту роль, не вступая в противоречие со схемами категоризации и различиями, присущими этой системе. Не исключено, однако, что они ощущают совершенно новые цвета, недоступные для моего ощущения, но тем не менее составляющие континuum от красного до синего. Это было бы странным, но не таким уж невероятным. Важно то, что здесь полностью отсутствовала бы та проблема, которая была характерна для случая с блекнущими квалиа. Эти системы *не* заблуждались бы систематически относительно характеристик своего опыта. Там, где они говорили бы о переживании различий, они так и могли бы переживать различия; там, где они заявляли бы о интенсивных переживаниях, они имели бы интенсивные переживания, и т. д. Конечно, цвета, именуемые ими «красным», были бы отличны от того, что я называю красным, но в этом нет никакой проблемы; это происходит и при обычной инверсии: Важно, что, в отличие от случая с блекнущими квалиа, структурные черты переживаний этих систем остаются полностью неизменными.

Тем не менее можно отыскать сходный и хороший аргумент против возможности инвертированных квалиа¹⁵. Предположим опять-таки для целей *reductio*, что инвертированные квалиа эмпирически возможны. В таком случае могут существовать две функционально изоморфные системы, находящиеся в одинаковом функциональном состоянии, но имеющие различные переживания. Примем для наглядности, что эти системы — я, ощущающий красное, и мой кремниевый изоморф, ощущающий что-то синее (здесь есть деталь, связанная с генерализацией, и позже я обсажу ее).

Как и прежде, сконструируем ряд промежуточных случаев, связывающих меня и моего изоморфа. Но тут аргумент будет развиваться иначе. Нас не должно беспокоить, *каким образом* изменяются переживания по мере продвижения по этому ряду. Быть может, они претерпевают резкие изменения, совершают громадные скачки, хотя, конечно, эти изменения скорее всего имеют постепенный характер. Важно лишь, чтобы в этом ряду было два пункта *A* и *B*, таких, что: (1) между *A* и *B* заменено не более одной десятой части мозга и (2) переживания в *A* и *B* существенно разнятся. Чтобы убедиться в том, что дело должно обстоять именно так, нужно лишь посмотреть на те пункты, где заменено десять, двадцать и т. д. вплоть до девяноста процентов мозга. Переживания красного и синего отличаются друг от друга в степени, достаточной для того, чтобы какие-то из соседствующих пар здесь *должны* были заключать в себе заметное различие (то есть это различие должно быть достаточным для того, чтобы оно было заметным, если бы эти пары ощущались одним и тем же субъектом); невозможно перейти от красного к синему десятью незаметными прыжками.

Конечно, между различными переживаниями могут существовать незаметные различия. Если чуть-чуть изменить оттенок красного, то я не смогу заметить это различие. Можно было бы предположить, что это происходит потому, что различие здесь хотя и существует в мире, но отсутствует в переживаниях; но если бы дело ограничивалось только этим, то можно было бы повторить подобные изменения тысячу раз, показав в итоге, что красное и синее производят одни и те же ощущения, что нелепо. Поэтому могут существовать *какие-то* незаметные различия переживаний. Можно наблюдать этот феномен, глядя на широкую полосу краски с небольшими вариациями оттенков цвета; порой оказывается крайне непросто сказать, отличаются или совпадают ощущения ее различных частей. Важно, однако, что незаметные отличия очень *малы*. Десять подобных прыжков в лучшем случае могли бы перенести нас от одного оттенка красного к другому, незначительно отличающемуся от него оттенку того же цвета. (Это создает небольшую проблему для генерализации моего аргумента, и в дальнейшем я еще вернусь к обсуждению этого вопроса.)

Итак, между красной и синей системами должны существовать две системы, отличающиеся не более чем на десять процентов по своему внутреннему составу, но имеющие существенно разные ощущения. Ради наглядности представим, что эти системы — это я и Билл. Там, где я ощущаю красное, Билл имеет несколько иное ощущение. Мы вполне можем предположить, что Билл видит синее; не исключено и то, что его ощущение будет больше похоже на мое, чем в случае синего, но это не имеет значения для аргумента. Две этих системы отличаются еще и тем, что на месте нейронов какого-то небольшого участка моего мозга в мозге Билла находятся кремниевые чипы. Эта замена нейронной цепи на кремниевую — единственное физическое отличие между мной и Биллом.

Ключевым шагом в этом мысленном эксперименте будет изготовление точно такой же, как у Билла, кремниевой цепи и размещение ее в моей голове в качестве *дублирующего контура*. Эта резервная цепь будет функционально изоморфна цепи, уже имеющейся в моей голове. Мы снарядим этот контур датчиками и эффекторами, чтобы он мог взаимодействовать с другими частями моего мозга, но мы не будем напрямую подключать его. Вместо этого мы установим *переключатель*, дающий возможность прямого переключения между нейронной и кремниевой цепями. При активации переключателя нейронная цепь выключается, а кремниевая включается. Можно вообразить, что переключатель контролирует те участки интерфейса, в которых соответствующие цепи оказывают воздействие на другие участки мозга. При активации переключателя нейронные соединения убираются и их место занимают эффекторы кремниевой цепи. (Можно представить, что датчики обеих цепей остаются присоединенными все время, чтобы состояние обеих цепей менялось надлежащим образом, но в любое время только одна из цепей вовлечена в протекающие в мозге процессы. Мы могли бы также придумать сходный эксперимент, в котором отключались бы как датчики, так и эффекторы, для гарантии полной изоляции резервной цепи от остальной системы. Это привело бы к изменению некоторых деталей, но мораль осталась бы неизменной.)

Непосредственно после активации переключателя обработка информации, ранее производившаяся нейронной цепью, начинает производиться кремниевой цепью. Контролирующий поток внутри системы оказывается перенаправленным. Однако моя функциональная организация никак не отличается от той, которая была бы у меня, если бы мы не активировали переключатель. Два эти случая отличаются, по сути, лишь физическим составом одной из цепей системы. Еще одно отличие связано с физическим составом другой, «бездельничающей» цепи, но оно нерелевантно для функциональной организации,

так как оно не играет никакой роли в воздействии на другие компоненты системы и в координации поведения.

Что происходит с моими ощущениями, когда мы активируем переключатель? Перед установкой той цепи я ощущал красное. После ее установки, но перед активацией переключателя я, судя по всему, по-прежнему буду ощущать красное, так как эти случаи будут различаться лишь добавлением цепи, никак не вовлеченнной в протекающие в мозге процессы; влияние на эти процессы было бы не большим, если бы я съел ее. После активации переключателя, однако, я мало чем отличаюсь от Билла. Единственным различием между Биллом и мной в данный момент является то, что у меня имеется каузально нерелевантная нейронная цепь, выпадающая из системы (много было бы даже представить, что она уничтожается после активации переключателя). Билл же, по предположению, ощущал что-то синее. Значит, после переключения у меня тоже будет ощущение синего.

Случится, таким образом, то, что мои ощущения поменяются «на моих глазах». Там, где я ощущал красное, теперь я буду ощущать синее. Совершенно внезапно яблоко на моем столе *посинеет*. Можно даже вообразить, как мы будем несколько раз дергать переключатель туда-сюда, и тогда ощущения красного и синего будут «скакать» перед моим взором.

На первый взгляд могло бы показаться, что тут нет ничего нелепого — и сама картина до странности привлекательна — но здесь происходит что-то необычное. Мои ощущения переключаются с красного на синее, но я не замечаю каких-либо изменений. Даже после того, как мы много раз дернем переключатель и мои квалиа заскачат туда-сюда, я так и буду продолжать свои занятия, не замечая ничего необычного. По предположению, моя функциональная организация остается полностью нормальной. В частности, после активации переключателя она меняется во времени точно так, как она менялась бы и без этой активации. В моих поведенческих диспозициях не обнаруживается никакого особого различия. У меня не возникнет вдруг желание сказать: «Хм! Происходит что-то странное!» Здесь нет места для того чтобы встрепенуться, воскликнуть, даже для того, чтобы отвлечь свое внимание. Любая необычайная реакция предполагала бы функциональное различие двух цепей, что расходилось бы с их изначально предположенным изоморфизмом. Согласно замыслу, моя когнитивная организация остается такой же, что и обычно, и, в частности, не отличается от той, какая была бы без активации переключателя.

Очевидно, что при любом функциональном конструировании убеждений получается, что я не смогу обретать каких-либо новых убеждений при активации переключателя. Но даже если оспаривать функциональное объяснение, кажется крайне неправдоподобным, что простая

замена нейронной цепи на кремниевую цепь при сохранении общей организации могла бы оказаться ответственной за добавление столь заметного нового убеждения, как «мои квалиа только что перевернулись». Как и в случае с блекнущими квалиа, здесь просто нет места для подобного изменения, если только оно не происходит в смежном картезианском бестелесном духе.

Мы, таким образом, опять пришли к *reductio ad absurdum*. Кажется совершенно неправдоподобным предположить, что мои ощущения могли бы меняться столь существенным образом при том, что я не обращал бы на это пристального внимания и не был способен заметить это изменение. Это опять-таки означало бы наличие радикального разрыва между сознанием и познанием. Если подобные вещи могли бы случаться, то психология и феноменология были бы радикально нескончордированными — в гораздо большей степени, чем даже в сценарии с блекнущими квалиа.

Этот сценарий со «скачущими квалиа» может быть логически возможным (хотя данный случай столь экстремален, что он представляется *всего лишь* логически возможным), но это не означает его правдоподобия как эмпирической возможности — в этом смысле он не более правдоподобен, чем то, что мир был создан пять минут назад. В качестве эмпирической гипотезы кажется гораздо более вероятным то, что при значительном изменении ощущений субъект, при условии его вменяемости и внимательности, должен быть в состоянии заметить это изменение. Если нет, то сознание и познание связаны между собой лишь самыми тонкими нитями.

В самом деле, если мы должны были бы допустить естественную возможность скачущих квалиа, то это привело бы нас к тревожной мысли: такое могло бы *действительно* существовать и происходить постоянно. Физиологические свойства наших функциональных механизмов все время меняются. Функциональные свойства этих механизмов весьма устойчивы; можно было бы ожидать, что эта устойчивость обеспечивалась бы эволюцией. Но не существует адаптивных причин константности нефункциональных свойств. С течением времени несомненно происходят изменения низкоуровневых молекулярных свойств. Такие свойства, как положение, атомарный состав и т. п. могут меняться при сохранении функциональной роли, и такие изменения почти наверняка постоянно происходят.

Если мы допустим, что квалиа зависят не только от функциональной организации, но и от деталей ее реализации, то может оказаться, что *наши* квалиа действительно все время скачут перед нашим взором. Не видно убедительного основания считать, что изменение от нейронов к кремнию должно иметь последствия, а изменение в нейронной реализации — нет¹⁶; грань можно *обоснованно* провести только

на функциональном уровне¹⁷. Причина, по которой мы сомневаемся, что подобные скачки происходят у нас, состоит в том, что мы придерживаемся следующего принципа: при значительном изменении наших переживаний мы можем заметить это изменение. Если бы мы были вынуждены согласиться с возможностью скачущих квалиа в исходном случае, мы должны были бы отказаться от этого принципа, и им уже нельзя было бы воспользоваться в качестве оборонительного средства против скептицизма даже в более привычных ситуациях.

Нельзя исключить, что мы и в самом деле могли бы осуществить подобный эксперимент. Конечно, нам пришлось бы преодолеть громадные практические трудности, но — по крайней мере, в принципе — кто-нибудь мог бы установить такую цепь у меня и я мог бы увидеть, что произойдет и рассказать об этом остальным. Впрочем, разумеется, нет смысла ставить этот эксперимент, так как мы знаем, каков будет его результат. Я скажу, что мои опытные переживания все время оставались неизменными, что я ощущал один и тот же оттенок красного и не заметил ничего странного. Я проникнусь еще большей, чем до этого, убежденностью, что квалиа детерминируются функциональной организацией. Конечно, это не будет *доказательством*, но с этим свидетельством будет непросто спорить всерьез.

Я делаю вывод, что наиболее правдоподобной гипотезой явным образом является та, согласно которой замена нейронов при сохранении функциональной организации оставит квалиа неизменными. Проблемы сценария со скачущими квалиа могут быть связаны с изначальным допущением того, что функционально изоморфная кремниевая система могла бы ощущать синее там, где я ощущаю красное. И самой разумной реакцией в таком случае будет отказ от этого допущения и, соответственно, заключение, что опыт целиком определяется функциональной организацией.

Следует отметить, что этот мысленный эксперимент столь же успешно работает против возможности отсутствующих квалиа, как и против возможности инвертированных квалиа. Мы просто берем две точки на пути к отсутствующим квалиа, между которыми имеется значительное различие переживаний, и тем же способом устанавливаем резервную цепь. Как и прежде, если отсутствующие квалиа возможны, то активация переключателя приведет к осцилляции квалиа перед моим взором — сначала они будут яркими, потом тусклыми, и наоборот, а я ничего не замечаю. Опять-таки, гораздо более вероятно то, что такие незамеченные скачки невозможны, а это значит, что невозможны и отсутствующие квалиа.

Сам я считаю этот аргумент против отсутствующих квалиа даже более убедительным, чем тот, который был выдвинут в параграфе 3,

хотя оба они могут сыграть какую-то роль. Оппонент мог бы просто закусить удила и признать возможность блекнущих квалиа, но признание скачущих квалиа связано с трудностями, на порядок более серьезными, чем те. Серьезное отличие связано с самой непосредственностью переключения, как и с тем фактом, что феномен, который не может заметить субъект, является очень динамичным и бросающимся в глаза. Блекнущие квалиа означали бы, что какие-то системы оторваны от их сознательного опыта, но скачущие квалиа указывали бы на наличие еще более странного разрыва.

Структура аргумента от скачущих квалиа такова, что он оставляет место для большего числа лазеек, чем аргумент от блекнущих квалиа. Похоже, однако, что ни одна из этих лазеек не может быть использована для получения оппонентом каких-то дивидендов. Ниже я рассмотрю эти лазейки, а также обсужу еще одно возражение против этого аргумента. Здесь можно было бы опять выдвинуть и возражения, выдвигавшиеся в связи со случаем с блекнущими квалиа, но ответы на них более или менее совпадали бы с теми, что были даны ранее; так что я не буду повторять их.

Возражение 1:

Лазейки, связанные со скоростью и историей

Выдвинутый мной аргумент естественным образом может быть расширен снейронно-кремниевого случая на многие другие примеры функциональных изоморфов,

но есть и пара исключений, связанных со скоростью и историей. Если изоморф гораздо быстрее или медленнее, чем исходная система, то мы не можем просто подставить цепь из одной системы в другую и ожидать, что все будет нормально функционировать. Поэтому мой аргумент не исключает возможности того, что изменение в скорости, оставляющее функциональную организацию неизменной, могло бы приводить к инвертированию квалиа. Сходная лазейка имеется и для физических изоморфов, отличающихся только своей *историей*: быть может, если я родился в Южном Хэмпшире, я буду ощущать зеленое, тогда как мой физический двойник, родившийся в Северном Хэмпшире, — красное. История не может варьироваться в сценарии со скачущими квалиа (в сценарии с блекнущими квалиа — может), и поэтому данный аргумент не затрагивает гипотезу, в соответствии с которой квалиа супервентны на прошлом.

Но ведь ни одна из этих гипотез не была столь уже правдоподобной. Разумно предположить, что история должна влиять на наши квалиа путем воздействия на наше физическое устройство, но зависимость от истории, о которой идет речь выше, была бы гораздо более сильной: здесь, по сути, было бы нечто вроде «нелокального» воздействия отдаленной истории на имеющиеся в настоящий момент квалиа,

не опосредованного какими-либо компонентами физического устройства или пространственно-временным окружением. Что же касается скорости, то допущение того, что изменение скорости инвертировало бы квалиа, хотя их не могло бы инвертировать ничто другое, выглядело бы совершенно произвольным. Эти гипотезы непротиворечивы, и поэтому оппонент *мог бы* принять их, но нет серьезных оснований поступать подобным образом. После того, как мы установили, что все другие не меняющие организацию изменения сохраняют квалиа, идея о том, что их изменения могли бы быть связаны только со скоростью или историей, кажется малопривлекательной.

Другая небольшая лазейка связана с тем, что данный аргумент не исключает возможности очень мягких спектральных инверсий. К примеру, между темно-красным и еще чуть более темно-красным может быть девять промежуточных оттенков, таких, что нельзя различить ни один из двух смежных оттенков. В подобном случае сценарий скачущих квалиа не составляет проблемы; если система не замечает каких-либо различий при активации переключателя, то именно этого нам и следовало ожидать.

Разумеется, в одной десятой — как характеристике различия между двумя смежными системами — нет ничего такого уж особого. Но нельзя делать эту величину слишком большой. Если бы мы довели ее до половины, мы столкнулись бы с проблемой личного тождества: вполне можно было бы предположить, что при активации переключателя мы создаем новую личность, и в том, что она не замечала бы изменения, не было бы никакой проблемы. Быть может, мы могли бы доходить до 20 или 25 процентов, не создавая таких трудностей, но это опять-таки допускало бы возможность очень мягких инверсий, таких, которые могли бы быть образованы четырьмя или пятью незаметными изменениями. Мы, впрочем, можем уменьшить воздействие этого затруднения, указав на крайнюю маловероятность равной зависимости опыта от всех участков мозга. Если переживания цвета главным образом зависят, скажем, от небольшого участка зрительной коры, то мы одним махом могли бы осуществлять любую инверсию квалиа, меняя лишь небольшой фрагмент системы, и аргумент прошел бы даже против мягчайшей из заметных инверсий квалиа.

В любом случае, лазейки здесь были бы неопасными. В лучшем случае мы оставляем возможность крайне незначительной недоопределенности переживаний организацией. Подобная недоопределенность могла бы казаться до такой степени незначительной, что не представляла бы интереса, но в любом случае можно отметить, что она ведет

Возражение 2: А что же мягкие инверсии?

к малопривлекательной позиции. Основательным кажется предположение, что переживания должны быть либо повсеместно инвертируемыми, либо вообще не инвертируемыми, и почему мир должен быть таким, что в нем допустимо лишь незначительное инвертирование? Это выглядело бы совершенно произвольным. Мы не можем исключить данную гипотезу, но последняя не обладает большой изначальной правдоподобностью.

**Возражение 3:
Остающиеся без внимания квалиа**

Аналогичным образом данный аргумент оставляет лазейку для предположения, что инвертируемыми могли бы быть *остающиеся без внимания* квалиа. К примеру, если мы не обращаем внимания на периферию нашего визуального поля, то здесь могла бы происходить инверсия квалиа, а мы не замечали бы ее. В самом деле, недавние эксперименты (Rensink, O'Regan, and Clark 1995) показывают, что можно изменить довольно важные характеристики изображения, которое демонстрируется субъекту, при том, что он ничего не заметит, если только он не сконцентрирован на этих характеристиках (обычно в этих экспериментах между демонстрацией двух изображений имеется небольшой промежуток времени, так что эта ситуация не вполне совпадает со сценарием скачущих квалиа, хотя и близка ему). Так что эти аргументы оставляют открытой возможность *остающихся без внимания* инвертированных квалиа.

Ничто, однако, в подобных выкладках не указывает на возможность инвертирования *не остающихся без внимания* квалиа. Так что использование этой лазейки привело бы к малопривлекательной позиции, в соответствии с которой квалиа организационно инвариантны, когда они находятся в фокусе достаточного внимания, но зависят от других черт в ином случае. (Вероятно, инвертированное ощущение зеленого на периферии опять отскочит к красному, когда на него будет обращено внимание?) Возможность сделать эту позицию теоретически приемлемой кажется крайне маловероятной. Как и в случае с другими лазейками, она ведет только к концепциям, изначально лишенным какого-либо значимого правдоподобия.

**Возражение 4:
Двойное переключение**

Еще одно возражение таково¹⁸. Мы можем представить сходный эксперимент, в котором мы меняем местами соединения, идущие от тех мест, куда поступают данные о красном и синем, в центральные участки мозга таким образом, что синие стимулы играют ту роль, которую ранее играли красные стимулы, и в котором для компенсации мы также систематически меняем местами *исходящие* из центральной

части соединения. Когда входящие данные о синем становятся причиной того, что центральная часть оказывается в состоянии, ранее связывавшемся с красным, исходящие из центральной части в другие участки мозга соединения перебрасываются так, что остальные части мозга функционируют именно так, как они функционировали бы в случае отсутствия каких-либо изменений. При этом мои ощущения почти наверняка поменяются с красного на синий, но мои поведенческие диспозиции останутся полностью неизменными. В этом случае повторные переключения точно приведут к скачущим квалиа. Не выходит ли в конце концов, что в скачущих квалиа нет несообразности?

Во-первых, надо отметить, что такая переделка была бы гораздо более масштабной задачей, чем все те, о которых шла речь выше. Центральная часть мозга будет воздействовать на весь мозг в самых разных местах. Каждое из этих соединений надо будет перебросить, причем, что существенно, во всех этих случаях нельзя будет ограничиться простой переброской. Мы не можем просто переключить «красные исходящие» на «синие исходящие», как это мы могли бы сделать с входящими; исходящие из центральной системы данные могут презентировать такие разные вещи, как извлеченные воспоминания, двигательные команды и т. д., и здесь дело не исчерпывается простым различием «полярности» красных и синих исходящих. Для определения надлежащих «синих исходящих», вероятно, потребовалось бы симуляция всех процессов обработки информации в центральной зоне при наличном изначальном состоянии и данных на входе, чтобы увидеть, что из этого выйдет. Если так, то каузальную работу будет производить эта симулирующая система, а не сама центральная зона, и это приведет к утрате значимости данного сценария.

Во-вторых, даже если бы оказалось, что можно как-нибудь легко перебросить исходящие соединения, отметим, что при этом сохранятся только поведенческие диспозиции, а не функциональная организация. Как это могло бы переживаться? Думаю, что в данном случае я заметил бы переключение и попытался действовать в соответствии с ним, но ощущал бы себя так, как если бы какой-то кукловод воспрепятствовал осуществлению моих замыслов. В отличие от разбиравшегося ранее случая, здесь имелось бы *место* для возникновения таких дополнительных убеждений и других когнитивных состояний; они подкреплялись бы различными состояниями центральной зоны. И, быть может, при появлении обратной связи и индикации данными на входе в центральные участки того, что моторные движения совершенно не совпадали с планируемыми, состояние центральной зоны, как можно вообразить, подверглось бы серьезному возмущению. На деле это возвращает нас к первому возражению, так как казалось бы почти

невозможным систематически компенсировать эти эффекты обратной связи. Так или иначе, но существенное различие в функциональной организации говорит об отсутствии аналогии этих случаев.

6. Нередуктивный функционализм

Подводим итоги: мы установили, что если отсутствующие квалиа возможны, то блекнущие квалиа тоже возможны; если инвертированные квалиа возможны, то скачущие квалиа возможны; и если отсутствующие квалиа возможны, то скачущие квалиа возможны. Но возможность блекнущих квалиа маловероятна, а скачущих квалиа — крайне маловероятна. Поэтому крайне маловероятна возможность отсутствующих и инвертированных квалиа. Из этого следует, что у нас есть серьезное основание считать истинным принцип организационной инвариантности, равно как и то, что функциональная организация полностью определяет сознательный опыт.

Следует отметить, что эти аргументы не устанавливают функционализм в самом сильном его смысле, так как в лучшем случае они показывают эмпирическую (или естественную) невозможность отсутствующих и инвертированных квалиа. Эти доводы нельзя расширить до аргументов в пользу логической невозможности отсутствующих и инвертированных квалиа — чего могли бы желать некоторые функционалисты — по двум причинам. Во-первых, как блекнущие, так и скачущие квалиа, несмотря на неправдоподобность, кажутся непротиворечивыми гипотезами. Кто-то мог бы оспорить логическую возможность этих гипотез, утверждая, к примеру, что возможность заметить отличия между квалиа является чем-то конститutивным для последних. Эта концептуальная интуиция спорна, но в любом случае есть и вторая причина, по которой эти аргументы не могут установить логическую детерминацию опыта функциональной организацией.

Чтобы увидеть эту вторую причину, заметим, что данные аргументы используют в качестве эмпирической посылки ряд фактов о распределении функциональной организации в физических системах: о том, что я обладаю сознательными переживаниями определенного рода или что ими обладают такие-то биологические системы. Если бы мы установили логическую невозможность блекнущих и скачущих квалиа, то это могло бы устанавливать логическую необходимость следующего *кондиционала*: если система с высокодетализированной функциональной организацией F имеет сознательный опыт определенного вида, то любая система с организацией F имеет такие же опытные переживания. Но мы не можем установить логическую необходимость заключения этого кондиционала без установления логической необходимости его

посылки, а эта посылка сама по себе эмпирична. Для установления логической детерминации опыта функциональной организацией мы должны были бы вначале убедиться в логической супервентности опыта на физическом, что, как я доказывал, невозможно. Даже если бы мы могли установить логическую супервентность на физическом, это скорее всего было бы сделано путем функциональной дефиниции, но при наличии такой дефиниции можно было бы заключить к логической невозможности отсутствующих и инвертированных квалиа без необходимости прибегать для этого к каким-то причудливым аргументам. Так что аргументы от блекнущих и скачущих квалиа в любом случае малополезны для доказательства логической или метафизической невозможности отсутствующих или инвертированных квалиа.

Данные аргументы поэтому не могут устанавливать сильную разновидность функционализма, где функциональная организация оказывается *конститутивной* для сознательного опыта; но они приносят успех попыткам утвердить более слабую его форму, которую я назвал *нередуктивным функционализмом*, согласно которому функциональная организация с естественной необходимостью оказывается достаточной для сознательного опыта. В соответствии с этим воззрением, сознательный опыт детерминируется функциональной организацией, но не обязан сводиться к ней.

В любом случае, вывод все еще силен. Принцип инвариантности говорит нам, что сознательными, в общем, могут быть когнитивные системы, реализованные на самых разных носителях. В частности, этот вывод оказывает серьезную поддержку амбициям исследователей в области искусственного интеллекта, о чем мы еще поговорим в главе 9. Если нередуктивный функционализм верен, то нередуцируемость сознания не является помехой для создания сознательного вычислительного устройства.

Важнее всего то, что мы продвинулись в деле ограничения принципов, благодаря которым сознание оказывается чем-то естественно супервентным на физическом. Мы сузили круг релевантных свойств в супервентностной базе до *организационных* свойств. В определенном смысле мы можем сказать, что сознание супервентно не только на физическом, но и на организационном. Тут нужно тщательно подбирать слова, так как всякая система реализует множество видов функциональной организации, но мы можем утверждать следующее: для всякой физической системы, порождающей сознательный опыт, существует такая функциональная организация F , реализованная этой системой, что естественно необходимым является то, что любая система, реализующая F , будет иметь идентичные сознательные переживания. Для отыскания надлежащей F мы должны углубиться в детали настолько,

чтобы зафиксировать такие когнитивные состояния, как суждения. Это, в свою очередь, может быть достигнуто требованием детализации F , достаточной для фиксации механизмов, ответственных за производование поведения, и фиксации поведенческих диспозиций. Это все, что требуют аргументы от блекнущих и скачущих квалиа, и поэтому все, что нужно для организационной инвариантности.

Законом, таким образом, является то, что для определенных функциональных организаций F реализация F будет сопровождаться сознательным опытом определенного рода. Это не значит, что речь здесь идет о *фундаментальном* законе. Было бы странным, если бы универсум содержал фундаментальные законы, связывающие комплексные функциональные организации с сознательными опытными переживаниями. Скорее можно ожидать, что он будет следствием более простых, более фундаментальных психофизических законов. Пока же принцип организационной инвариантности работает как сильный ограничитель финальной теории.

Глава 8

Сознание и информация: спекулятивные рассуждения

1. На пути к фундаментальной теории

Наши результаты пока состояли в том, что мы выделили ряд связующих звеньев между сознанием и физическими процессами, заслуживающих именования психофизическими законами. Одно из таких звеньев — принцип когерентности, сопрягающий сознание с осведомленностью, или глобальной доступностью. Второе — более конкретный принцип структурной когерентности, связующий структуру сознания и структуру осведомленности. Третье — принцип организационной инвариантности. Эти принципы могут быть компонентами окончательной теории сознания. Они позволяют нам использовать физические факты для предсказания и даже объяснения некоторых фактов о сознательном опыте. И они уж точно *накладывают ограничения* на облик финальной теории сознания: если такая теория несовместима с этими законами, она едва ли верна. Но картина не может ограничиваться этим. Сами по себе эти принципы не составляют окончательной теории или даже хоть какого-то подобия последней.

Проблема в том, что ни один из этих принципов не является подходящим кандидатом на роль *фундаментальных* законов в теории сознания. Все они выражают регулярности на достаточно высоком уровне. Скажем, понятие осведомленности (или глобальной доступности) — это высокоуровневое понятие с весьма размытыми границами; и очень маловероятно, чтобы это понятие использовалось в каком-то фундаментальном законе. Принцип организационной инвариантности, возможно, менее размыт, но и он выражает регулярность на далеком от фундаментальности уровне. Другая проблема: эти принципы оставляют крайне недоопределенной природу психофизической связи. Самые разные вопросы о ней остаются без ответа. К примеру, какого рода организация порождает сознательный опыт? Насколько она может быть простой для сохранении опыта? И как можем мы предсказывать конкретный характер опыта (не только его структуру), исходя из его

физической основы? Мы хотели бы, чтобы исчерпывающая теория сознания отвечала на них, но указанные принципы не помогают в этом.

Для окончательной теории нам нужен набор психофизических законов, аналогичных фундаментальным законам физики. Эти фундаментальные (или *базовые*) законы будут выражены на уровне, связующем базовые свойства опыта с элементарными характеристиками физического мира. Они должны быть точными законами и в своей совокупности не должны оставлять место для недоопределенности. При сочетании с физическими фактами о системе они должны позволять нам в полной мере предсказывать феноменальные факты об этой системе. Кроме того, подобно тому, как из базовых законов физики вытекают все высокоуровневые физические законы и регулярности (по крайней мере в сочетании с ограничительными условиями), из базовых законов сознания должны вытекать и объясняться различные небазовые законы, такие как принцип когерентности и принцип организационной инвариантности. При наличии набора фундаментальных физических и психофизических законов мы в определенном смысле сможем уяснить для себя базовое устройство универсума.

Это высокая планка, и в ближайшее время мы не возьмем эту высоту. Но мы по крайней мере можем двигаться в этом направлении. Принципы организационной инвариантности и структурной когерентности уже накладывают серьезные ограничения на облик фундаментальной теории, и существует не так уж много кандидатов на роль базовых конструктов, которые могли бы быть фундаментальными ингредиентами такой теории. В этой главе я изложу некоторые идеи относительно фундаментальной теории. Я не предложу здесь полноценной теории с исчерпывающим набором базовых законов, но высажу предположения о том, какие конструкты должны использоваться в этих законах, и о том, какая общая форма могла бы быть у последних. Это можно было бы рассматривать в качестве *прототеории*, скелета, на который можно было бы наращивать теорию.

Идеи этой главы гораздо более эскизы и спекулятивны, чем те, что высказываются в других частях этой книги, и вопросы, которые они вызывают, не уступают по своему числу ответам, которые они дают. И именно эти идеи скорее всего могут оказаться полностью неверными. Выдвигая эти разрозненные идеи, я стремлюсь не к тому, чтобы создать модель, которая сможет выдержать тщательную философскую проверку; нет, я всего лишь хочу, чтобы в нашем распоряжении оказались хоть какие-то идеи. Мы должны начать думать о фундаментальных теориях сознания, и, быть может, что-то из высказанного здесь пригодится и в дальнейшем.

2. Аспекты информации

Базовым понятием, с которым я буду работать в этой главе, является понятие *информации*. В наши дни можно столкнуться с множеством различных понятий информации, так что первое, что надо сделать, заводя речь об информации, — это прояснить, о чем именно мы будем говорить. Интересующее меня понятие информации во многом сходно с тем, которое обсуждал Шеннон (Shannon 1948). Я адаптирую и разовью эту идею. Моя трактовка останется относительно неформальной, заключая в себе лишь столько формализма, сколько требуется для уяснения наиболее важных аспектов данного понятия, релевантных для наших целей. В этом параграфе будет технический материал, но в других его будет меньше.

Шенна не интересовало семантическое понятие информации, согласно которому информация — это всегда информация *о* чем-то. Он сосредоточился скорее на формальном или синтаксическом понятии информации, где ключевым оказывается понятие о состоянии, выбранном из множества возможностей. Самым фундаментальным видом информации является *бит*, отражающий выбор между двумя возможностями: информацию, как утверждается, несет единичный бит (0 или 1), выбранный в пространстве, имеющем два состояния. В более сложном случае аналогичным образом несет информацию «сообщение», такое как «0110010101», выбранное из пространства возможных бинарных сообщений. Согласно Шенону, важна не какая-либо *интерпретация* этих состояний, а *специфика состояния* в пространстве различных возможностей.

Мы можем формализовать эту идею, обратившись к понятию *информационного пространства*. Информационное пространство — это абстрактное пространство, образованное множеством состояний, которые я буду называть *информационными состояниями*, и базовой структурой *отношений различия* между этими состояниями. Простейшее нетривиальное информационное пространство — это пространство, образованное двумя состояниями и их элементарным различием. Можно представить эти состояния в качестве двух «битов», 0 и 1. Их природа исчерпывается тем фактом, что эти состояния отличны друг от друга. То есть данное информационное пространство полностью характеризуется структурой его различия.

Другие информационные пространства более сложны. Усложнение может происходить двумя путями: допущением более сложной структуры различия между состояниями или допущением наличия внутренней структуры у самих состояний. Для иллюстрации первой возможности мы могли бы перейти к пространству с четырьмя состояниями — 0,

1, 2 и 3, для иллюстрации второй — к структурированному пространству с такими состояниями, как «110010101». Разумеется, эти способы можно было бы сочетать, что давало бы вдвое усложненные состояния, вроде тех, что имеются в пространстве таких сообщений, как «233102032». Сейчас я более детально рассмотрю две этих разновидности усложнений.

Начнем с первой разновидности усложнения, где очевиднее всего оказывается существование пространств с тремя состояниями, с четырьмя состояниями и т. д., структура различия которых является естественным продолжением структуры различия пространства с двумя состояниями. К примеру, элемент А, В, С или D, выбранный в четырехэлементном пространстве, несет информацию тем же способом, каким несет информацию бит. Природа значков «А», «В» и т. д., конечно, не имеет здесь значения; существенна для данного пространства опять-таки лишь его структура.

Более важным обстоятельством является наличие *континуальных* информационных пространств, состояния которых представляют собой континуум, аналогичный континууму реальных чисел между 0 и 1. В таком пространстве имеется бесконечное множество состояний. Структура различия этого пространства гораздо сложнее, чем в предыдущих случаях: эта структура прямо соответствует топологии континуума, где какие-то состояния находятся между другими состояниями, какие-то состояния — ближе друг другу, чем иные состояния и т. д. Но, как и прежде, мы можем считать информативной любую точку, выбранную из этого континуума.

Структура информационного пространства может быть и двумерным или многомерным континуумом, аналогичным структуре региона n -мерного пространства. Информацию будет нести, к примеру, единичная точка, выбранная из региона трехмерного пространства. В самом общем случае эта структура может быть задана структурой произвольного топологического пространства, по сути, являющегося неким множеством с отношениями «близости» или «соседства». Детали этого, впрочем, не будут иметь такого уж большого значения в дальнейшем, так как я буду говорить об интуитивно привычных структурах, таких как структура континуума.

Второй тип усложнения предполагает состояния с *внутренней структурой*. Они образованы множеством более базовых состояний, которые я буду называть *элементами*. Пример — пространство десятибитных состояний, аналогичных таким «сообщениям», как «100110100». Каждое состояние образовано здесь десятью элементами, и каждый элемент можно рассматривать так, будто он включается в свое собственное *подпространство* двух состояний, соответствующее изначальному пространству, содержащему два состояния. Данное информационное

пространство можно трактовать в качестве некоего продукта десяти подпространств, каждое из которых является самодостаточным информационным пространством.

Могут быть и более интересные внутренние структуры. Так, информационное состояние могло бы иметь континуальную внутреннюю структуру, будучи своего рода континуальным аналогом обсуждавшейся выше десятиэлементной структуры. Подобное состояние обладало бы бесконечным множеством элементов, каждый из которых включался бы в свое собственное подпространство. Соответствующее информационное пространство можно было бы представить как что-то родственное пространству функций в континууме (каждому значению которых соответствовало бы свое подпространство) или в более сложном континуальном пространстве.

Может быть и так, что *подпространства* были бы сложны в первом из упомянутых выше смыслов: элементы каждого подпространства могли бы, к примеру, распределяться в континууме. Так что здесь есть место для двух одновременно существующих уровней сложности. Скажем, каждое состояние могло бы быть образовано континуальной структурой элементов, каждый из которых мог бы принимать те или иные значения в континуальном подпространстве. Информационное состояние этого пространства могло бы рассматриваться в виде волновой формы или какой-то другой функции с континуальной областью определения и амплитудой: это непрерывный аналог описанных выше дискретных «сообщений».

В самом общем случае у информационного пространства будет два вида структуры: каждое комплексное состояние могло бы иметь внутреннюю структуру, и каждый элемент этого состояния принадлежал бы подпространству со своей структурой топологического различия. Первую из них можно было бы назвать *комбинаторной* структурой этого пространства, вторую — *реляционной* структурой подпространств. Каждое подпространство будет, как правило, иметь одну и ту же реляционную структуру, так что мы можем говорить просто о реляционной структуре самого пространства. *Общая* структура пространства задается соединением комбинаторных и реляционных структур. Зачастую я буду ограничиваться рассмотрением информационных пространств с одной лишь реляционной, а не комбинаторной структурой — когда в информационном состоянии есть только один элемент — так как это намного упрощает обсуждение.

Эта модель совершенно обходит стороной понятие семантического содержания, так что обсуждаемая здесь разновидность информации в лучшем случае имеет косвенное отношение к семантической разновидности информации, которая рассматривалась такими философами, как Дретске (Dretske 1981) и Барвайс с Перри (Barwise and Perry 1983).

Не исключено, что данную модель можно было бы расширить таким образом, чтобы она включала семантический элемент — привязывая к каждому информационному состоянию некое семантическое содержание — но в имеющемся виде эта модель независима от семантических соображений.

Данная формализация фиксирует идею Шеннона о том, что информация существенным образом предполагает состояние, выбранное из множества возможностей (в реляционной структуре пространства), а также фиксирует ту идею, что сложная информация может быть составлена из простой (в комбинаторной структуре пространства). Единичный бит, по Шенону, может быть информацией не менее, чем длинное «сообщение» типа «10011010». Шенон также рассматривает случай, когда информация содержится в континуальном пространстве, или в пространстве функций с континуальной областью определения. В каждом случае существен выбор одного элемента в пространстве контрастирующих возможностей.

Сам Шенон много занимается вопросом о *количестве информации* в информационном состоянии, и это количество оказывается мерилом того, насколько *специфичным* является состояние в информационном пространстве. Состояние в информационном пространстве с двумя состояниями несет один бит информации; состояние в пространстве с четырьмя состояниями — два бита; состояние в n -элементном пространстве — $\log_2 n$ битов. Если пространство является собой комбинацию подпространств, информация, которую несет состояние, равна сумме той информации, которую несут его элементы: десятизначное бинарное «сообщение» несет, к примеру, десять бит информации. Эта трактовка применима к дискретным пространствам; в континуальных пространствах количество информации должно определяться более тонким способом. Здесь меня не очень будет интересовать вопрос о количестве информации. Скорее меня будут интересовать сами информационные состояния, отношение которых к количеству информации можно мыслить наподобие отношения материи к массе.

Физически реализованная информация

Согласно данной мной дефиниции информационных пространств, такие пространства *абстрактны*, абстрактными состояниями являются и информационные состояния. Они не составляют какую-то часть конкретного физического или феноменального мира. Но при надлежащем взгляде на вещи, мы сможем отыскать информацию как в физическом, так и в феноменальном мире. Для этого нам надо обсудить различные способы *реализации* в мире информационных пространств и состояний. Сначала я обсажу физическую реализацию, а потом поговорю о феноменальной.

Кажется интуитивно ясным, что информационные пространства и состояния повсеместно реализованы в физическом мире. К примеру, мы можем рассматривать мой переключатель света как то, что реализует информационное пространство с двумя состояниями; его состояния «вверх» и «вниз» как раз и реализуют упомянутые два состояния. Или можно взглянуть на компакт-диск как на то, что реализует комбинаторное информационное состояние, образованное сложной структурой битов. Сходным образом можно усмотреть реализацию информации в термостате, книге или телефонной линии. Как можно было бы осмыслить эти интуиции?

Естественным путем установления связи между физическими системами и информационными состояниями было бы рассмотрение физически реализованной информации в терминах слогана, восходящего к Бейтсону (Bateson 1972): информация есть *различие, производящее различие*. Хотя мой переключатель света может занимать бесконечное множество позиций на непрерывной шкале, большинство этих различий совершенно безразлично для моего света. Находится ли переключатель в самой верхней позиции или на четверть опущен вниз, свет все равно будет гореть. Но если он смещается больше, чем на треть вниз, свет выключается. Пока речь идет о свете, у переключателя есть только два имеющих отношение к делу состояния, которые мы можем назвать «вверх» и «вниз». Различие между этими двумя состояниями — единственное различие, небезразличное для света. Поэтому мы можем рассматривать переключатель как то, что реализует информационное пространство с двумя состояниями, причем одни физические состояния переключателя соответствуют одному информационному состоянию, тогда как другие его физические состояния — другому.

В общем, информационное пространство, связанное с физическим объектом, всегда будет определяться относительно *каузального пути* (в данном случае — каузального пути от переключателя света к самому свету) и пространства возможных *последствий* в конце этого пути (в данном случае речь в этой связи будет идти о включенном или выключенном свете). Физические состояния будут соответствовать информационным состояниям сообразно их воздействиям на каузальный путь. Если два физических состояния оказывают одинаковое воздействие на данный каузальный путь — как это имеет место в случае двух положений электрического переключателя, в равной степени ведущих к включению света — то они будут соответствовать одному и тому же информационному состоянию. Если подобным образом скроить физические состояния, то мы получим базовый набор небезразличных физических различий, составляющих физическую реализацию информационного пространства.

Структура этого информационного пространства будет напрямую соответствовать структуре пространства последствий, которое, в свою очередь, будет либо дискретным пространством, либо континуальным пространством. К примеру, в случае света каузальный путь может вести к двум релевантным последствиям: свет может оказаться включен или выключен. Поэтому переключатель может рассматриваться как то, что реализует информационное пространство, заключающее в себе два состояния.

Сходным образом можно рассматривать и континуальные информационные пространства. Если бы мой переключатель позволял плавно регулировать свет, то поворот его рукоятки в разные позиции давал бы разную интенсивность света на непрерывной шкале. (На практике она могла бы быть дискретной, но я беру идеальную ситуацию.) Последствия, связанные с интенсивностью света, определяют непрерывное информационное пространство, реализованное в моем переключателе света. Физические состояния переключателя, производящие свет одинаковой интенсивности (состояния в тех областях, где рукоятка, возможно, нечувствительна, или состояния, отличающиеся друг от друга по нерелевантным параметрам, таким как цвет переключателя), будут связаны с одним и тем же информационным состоянием. Пространство информационных состояний имеет топологическую структуру континуума, где структура различий состояний соответствует структуре различий воздействий на интенсивность света.

Подобным способом можно проанализировать и информацию, реализованную на компакт-диске. У диска может быть бесконечное множество физических состояний, но если рассматривать его воздействие на проигрыватель компакт-дисков, то оказывается, что он реализует лишь конечное множество возможных информационных состояний. Множество изменений диска — микроскопическое изменение ниже уровня разрешающей способности оптического считающего устройства, небольшая царапина на диске или крупный рисунок на его обратной стороне — не влияют на функционирование этой системы. Для информационного состояния диска значимы лишь те различия, которые отражаются на выходе оптического считающего устройства. Это — различия, связанные с наличием питов и лендов на диске, соответствующих тому, что мы понимаем под «битами». С любым состоянием диска будет связано информационное состояние в большом информационном пространстве. Физические состояния различных оттисков одной и той же записи, если все нормально, будут связаны с одним и тем же информационным состоянием. Оттиски разных записей, да и дефектные оттиски одной записи будут связаны с различными информационными состояниями из-за разности их воздействий.

Это случай, в котором физически реализованное информационное пространство имеет комбинаторную структуру. Каждый «бит» на компакт-диске производит самостоятельное воздействие на проигрыватель компакт-дисков, так что можно считать, что каждый участок диска реализует собственное подпространство с двумя состояниями. Складывая все эти независимые воздействия, мы находим комбинаторную структуру в пространстве совокупных действий компакт-диска и тем самым находим ту же самую комбинаторную структуру в информационном пространстве, реализуемом этим диском. Это информационное пространство может рассматриваться как продукт большого множества подпространств с двумя состояниями, по одному подпространству на каждый пит или ленд диска.

Заметим, что, согласно этой концепции, физически реализованная информация является информацией лишь в той мере, в какой она может быть *обработана*. Как выражается в этой связи Маккей (MacKay 1969), «[И]нформация должна работать». Это согласуется с трактовкой информации и самим Шенноном. Его «количество информации» измеряет то, насколько специфичным является состояние в пространстве состояний, которые могут быть *переданы*, то есть которые могут сыграть особые роли на каком-то другом каузальном пути (именуемом Шенноном коммуникационным каналом). Информация для Шенна — это всегда передаваемое состояние, и объем информационного пространства на деле неявно определяется функцией передающей инстанции. Информация — это различие, которое может произвести различие при его передаче.

Сказанное очевидно из стандартной диаграммы Шенна (рис. 8.1) и соответствующего ее обсуждения. Источником информации является

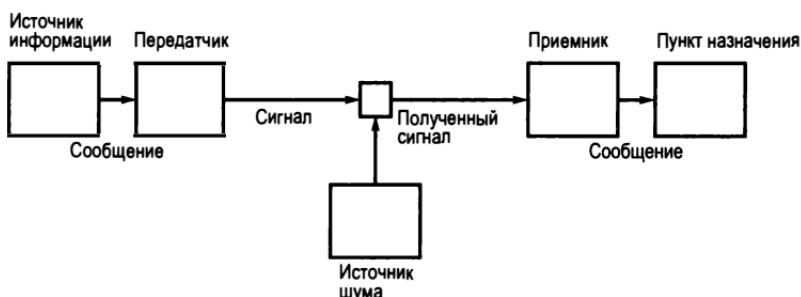


Рис. 8.1. Изображение Шенном информационного канала.
 (Источник: Diagram 1, из Claude E. Shannon and Warren Weaver, *The Mathematical Theory of Communication*, 1963. Copyright © 1949 by the Board of Trustees of the University of Illinois Press.
 Использовано с разрешения University of Illinois Press)

набор «сообщений» (информационных состояний), причем предполагается, что различные сообщения кодируются в отдельные сигналы передающим устройством и что последние соответствуют отдельным сообщениям. На деле это свойство *определяет* условия, при которых состояния могут считаться отдельными сообщениями. Если два разных физических состояния системы конвертируются в один и тот же сигнал, то в них реализуется одно и то же сообщение. Как явствует из правой части диаграммы, на последующих этапах может происходить порча и потеря информации, но для информационного канала существенно то, что различные информационные состояния производят разные воздействия через передатчик. В концепции Шеннона все это скорее подразумевается, чем четко обозначается, так как он напрямую не рассматривает отношение между физическими и информационными состояниями. Но при более внимательном анализе становится ясно, что индивидуализация информационных состояний осуществляется благодаря использованию принципа передаваемости.

Я не буду пытаться выдвинуть строгие критерии реализации информационного пространства в физической системе, но оставлю все на неформальном уровне принципа «различия, производящего различие». Эту неформальную идею можно было бы самыми разными способами формализовать, причем некоторые из них накладывают более сильные ограничения на реализацию, чем другие. На этом этапе было бы преждевременным останавливаться на каком-то из них. Оставляя вещи на неформальном уровне, мы сохраняем некоторые возможности для маневра, когда речь заходит о деталях, которые могут быть прояснены, когда мы составим лучшее представление о подобающем для конкретного применения. Если говорить о теории сознания, то надлежащее уточнение таких деталей будет одним из аспектов процесса уточнения самой этой теории.

Феноменально реализованная информация

Когда мы думаем о внедренности информации в мир, в голову прежде всего приходит ее физическая реализация, но это не единственный способ отыскания в нем информации. Информацию можно найти и в ее *феноменологической* реализации. Опытные состояния напрямую и естественным образом включаются в информационные пространства. Между феноменальными состояниями имеются естественные паттерны сходства и различия, и эти паттерны задают структуру различия информационного пространства. Так что мы можем считать, что феноменальные состояния реализуют информационные состояния тех пространств.

Так, пространство переживаний простого цвета имеет уже обсуждавшуюся нами трехмерную реляционную структуру. Абстрагируя пат-

терны сходства и различия этих переживаний, мы получаем абстрактное информационное пространство с трехмерной реляционной структурой, которая реализуется феноменальным пространством. Простое переживание любого цвета соответствует конкретному расположению в этом пространстве. Конкретное переживание красного — одно феноменально реализованное информационное состояние; конкретное переживание зеленого — другое.

Более сложные переживания, такие как ощущения всего зрительного поля, относятся к информационным пространствам со сложной комбинаторной структурой. К примеру, когда я смотрю на какую-то картину, мое переживание относится к пространству с (по меньшей мере) комбинаторной структурой двумерного континуума, причем каждый элемент этого континуума имеет (по меньшей мере) трехмерную реляционную структуру простого цветового пространства. Структура цветных пятен визуальном поле не *так уж сильно* отличается по своему типу от бинарной структуры десятизначного сообщения, хотя их комбинаторная и реляционная структура значительно превосходят по своей сложности структуру последнего.

При отыскании феноменальных реализаций информационных пространств мы не опираемся на принцип «различия, производящего различие», который мы использовали для отыскания физических реализаций информационных пространств. Скорее мы отсылаем к внутреннему качеству переживаний и упорядочивающей их структуре — отношениям сходства и различия, имеющимся между ними, и к их внутренней комбинаторной структуре. Любое переживание вступает в естественные отношения сходства и различия с другими переживаниями, так что мы всегда сможем найти информационные пространства, в которых они находятся.

Эта трактовка информации выявляет ключевое звено между физическим и феноменальным: везде, где мы находим феноменальную реализацию информационного пространства, мы видим, что то же самое информационное пространство получает и физическую реализацию. И когда опыт реализует какое-то информационное состояние, то же самое информационное состояние реализуется и в физическом субстрате опыта.

Двуаспектный принцип

Возьмем простое переживание цвета, которое реализует информационное состояние того или иного трехмерного информационного пространства. Мы можем обнаружить, что то же самое пространство реализовано мозговыми процессами, фундирующими это переживание, а именно в трехмерном пространстве нейронально закодированных презентаций визуальной коре. Элементы этого трехмерного

пространства прямо соответствуют элементам феноменального информационного пространства.

Мы не знаем в точности, как закодированы эти состояния, и, стало быть, не знаем в точности, как физически реализовано это информационное пространство. Но мы знаем, что оно должно быть реализовано, так как последующие процессы демонстрируют все систематические последствия информационной реализации. Наши отчеты, к примеру, могут систематически меняться в соответствии с локализацией в цветовом пространстве, скажем, когда мы оцениваем цвета как относительно более «темные» и «светлые»; и мы можем соотносить одни объекты с другими в соответствии с их сходствами или различиями по цвету. Так что мы знаем, что в зрительной коре должны иметься релевантные различия, *передаваемые* в другие участки мозга, создающие трехмерное пространство возможных эффектов. Состояния, лежащие в основе двух неотличимых переживаний, будут производить одинаковые релевантные эффекты, даже если с ними связаны несколько иные физические детали — по аналогии с небольшими различиями состояния переключателя света — и состояния, лежащие в основе любых двух *сходных* переживаний, будут производить сходные эффекты. Поэтому мы можем рассматривать зрительную кору как то, что реализует информационные состояния в трехмерном пространстве.

Это же верно и для более сложных переживаний, таких как переживания зрительного поля в целом. Они реализованы в комбинаторном информационном пространстве, и то же самое пространство должно быть физически реализовано в фундирующих их мозговых процессах. Мы знаем, что в каждом участке этого поля различные простые переживания соответствуют различиям в разного рода следующих за ними эффектах, причем эти различия можно распределить сообразно локализации в поле. Так, мы можем давать разные ответы на конкретные вопросы относительно окрашенности того или иного участка; это особое пространство эффектов для каждого участка создает отдельные подпространства для каждого участка. Поэтому где-то в зрительной коре должны быть закодированы информационные состояния — чтобы все релевантные различия могли передаваться для использования в последующих процессах. Пространство релевантных возможных состояний изоморфно здесь пространству возможных переживаний; и поэтому можно видеть, как одно и то же информационное состояние реализуется как физически, так и феноменально.

Информация не обязательно должна кодироваться *локально*, к примеру, в небольшой структуре окружающих нейронов. Она вполне может физически реализовываться в холистской манере, как бывает, в частности, с некоторыми голографическими формами хранения информации. Релевантные различия состояний зрительной коры могли

бы соответствовать различиям, распределенным по всей коре. Но пока они остаются различиями, которые могут быть переданы и имеют соответствующие эффекты, данная информация по-прежнему будет оказываться реализованной.

Естественно предположить, что эта двойная жизнь информационных пространств соответствует какой-то дуальности на глубоком уровне. Можно было бы даже подумать, что эта двойная реализация — ключ к фундаментальному сочленению физических процессов и сознательного опыта. Нам нужен *какой-то* конструкт для установления этой связи, и информация кажется вполне достойным кандидатом. Не исключено, что принципы, связанные с двойной реализацией информации, можно было бы развить в систему базовых законов, связующих физическое и феноменальное.

Это можно было бы сделать, предложив в качестве базового принципа наличие у информации (в актуальном мире) двух аспектов, физического и феноменального. Любое феноменальное состояние реализует какое-то информационное состояние, которое реализуется также в когнитивной системе мозга. И наоборот, по крайней мере некоторые физически реализованные информационные пространства таковы, что при любой физической реализации информационного состояния в этом пространстве данное состояние получает также и феноменальную реализацию¹.

Сам по себе данный принцип далек от конституирования полной психофизической теории. Скорее, он образует что-то вроде трафарета для психофизической теории, предоставляя в наше распоряжение базовую схему, в которую можно вписывать конкретные законы. Для разворачивания этого принципа в теорию надо ответить на множество разных вопросов. К примеру, к *каким* именно физически реализованным информационным пространствам применим этот базовый принцип? Я рассмотрю этот вопрос в четвертом параграфе, но пока оставлю его открытым. Недоопределенность другого рода связана с нечеткостью дефиниции физически реализованной информации: для полностью конкретизированной психофизической теории нам потребовалось бы точно знать, в каком случае можно говорить о физической реализации информационного пространства. Впрочем, процесс создания теории предполагает подобные вещи.

Ряд других важных вопросов касается *онтологии* этого воззрения. Насколько серьезно мы должны относиться к этим рассуждениям о «двуаспектности»? До какой степени эта модель будет *овеществлять* информацию, или трактовать ее как нечто реальное? Утверждается ли в ней, что физическое, феноменальное, или и то, и другое, онтологически зависимы от информационного? Пока я не буду отвечать

на эти вопросы. Позже в этой главе я рассмотрю различные возможные интерпретации этой онтологии. Некоторые из них трактуют информацию в качестве всего лишь полезного конструкта для характеристики психофизических законов; другие отводят ей более фундаментальную онтологическую роль. Соответственно, одни из них относятся к «двуаспектности» серьезней, чем другие.

Пока же я абстрагируюсь от этих метафизических проблем. Данный принцип можно рассматривать просто в качестве закона, связывающего физическое с феноменальным, с онтологическими последствиями, не слишком отличающимися от последствий уже рассмотренных законов. Мы уже знаем, что опыт порождается физическим вследствие определенных законов, приложимых к определенным физическим характеристиками мира. И суть здесь в предположении, что базовым уровнем приложения этих законов к физическому миру является уровень физически реализованной информации.

Конечно, информация может и не быть столь *элементарной* характеристикой физического мира, какими являются масса и заряд, но она и не должна быть элементарной чтобы играть роль в фундаментальных психофизических законах. Базовые физические и феноменальные свойства уже предоставляют нам все необходимые фундаментальные свойства. Нам нужен лишь конструкт, позволяющий соединить эти области. Информация кажется простым и четким конструктом, подходящим для подобного соединения и могущим, как и было обещано, снабдить нас простым и исчерпывающим набором законов. Если подобный набор законов достичим, то тогда мы и впрямь будем располагать фундаментальной теорией сознания.

Может, впрочем, оказаться и так, что саму информацию можно было бы рассматривать как что-то фундаментальное. Кстати, идея о том, что физика в конечном счете работает с информацией, уже обсуждалась некоторыми физиками. При успешности этой идеи могло бы оказаться, что физическое в известном смысле производно от информационного, и можно было бы предельно четко изложить онтологию этой позиции. Позже в этой главе я обсажу некоторые идеи, укладывающиеся в указанное русло.

3. Некоторые подкрепляющие аргументы

У меня нет решающих аргументов, которые позволили бы доказать, что информация — ключ к соединению физических процессов и сознательного опыта, но эту идею можно подкрепить косвенными путями. Я уже обсуждал первый тип подкрепляющих соображений: наблюдение о том, что одни и те же информационные пространства получают

как физическую, так и феноменальную реализацию. Ниже я упомяну другой важный подкрепляющий ресурс и два не столь значимых ресурса подобного рода.

Последние два ресурса связаны с тем фактом, что двухаспектная концепция информации совместима с разработанными ранее психо-физическими принципами, в частности, с принципом структурной когерентности и с принципом организационной инвариантности. Эти принципы являются сильными ограничивающими факторами, и способ сочетания их с фундаментальной теорией неочевиден — так что совместимость информационной концепции с ними является хорошим знаком для данной концепции.

Особенно легко убедиться в совместимости со структурной когерентностью: в известном смысле информационная концепция изготовлена на заказ для соответствия этому ограничению. Структура опыта есть не что иное, как структура феноменально реализованного информационного пространства, а структура осведомленности — не что иное, как структура физически реализованного информационного пространства. Чтобы убедиться в первом, отметим, что то, что я называл имплицитной структурой опыта, соответствует реляционной структуре информационного пространства, а то, что я называл эксплицитной структурой опыта, соответствует комбинаторной структуре этого пространства. Чтобы убедиться во втором, отметим, что различные детали структуры осведомленности по определению являются различиями, небезразличными для последующих процессов, так как они напрямую доступны для глобального контроля и поэтому оказываются физической реализацией информационного пространства. Так как обе они на деле реализуют одно и то же информационное пространство, из этого вытекает принцип структурной когерентности.

Должен заметить, что двухаспектный принцип сам по себе не гарантирует проецирование структуры осведомленности на опыт. Чтобы удостовериться в таком проецировании, мы должны показать, что наличие физическое информационное пространство относится к числу тех пространств, к которым применим двухаспектный принцип. Для этого нам потребовалась бы более детализированная версия этого принципа, надлежащим образом ограничивающая задействуемые информационные пространства, так, чтобы он хотя бы включал информацию, доступную для глобального контроля в привычных нам случаях. В имеющемся виде двухаспектный принцип еще не *предсказывает* полного принципа структурной когерентности, однако по крайней мере *совместим с ним*.

Нетрудно увидеть и то, что двухаспектный принцип совместим с принципом организационной инвариантности. Чтобы убедиться в этом,

заметим, что когда некая система реализует информационное пространство, она делает это вследствие ее функциональной организации. Любая другая система, функционально изоморфная на достаточном уровне детализации, будет иметь тот же паттерн небезразличных различий и поэтому будет реализовывать то же самое информационное пространство. Поэтому если мои переживания возникают вследствие реализации в моем мозге информационных пространств, то те же самые пространства будут реализованы и в функциональном изоморфе, и возникнут те же переживания, что, собственно, и предсказывает принцип инвариантности.

Фундаментальная теория сознания должна будет оперировать такими физическими характеристиками, которые являются одновременно организационно инвариантными и достаточно простыми для того, чтобы играть определенную роль в фундаментальных законах. Организационно инвариантные характеристики по большей части не очень просты, а простые характеристики по большей части не являются организационно инвариантными. Физически реализованная информация может оказаться наиболее естественной характеристикой, удовлетворяющей обоим критериям. Факт соответствия ее обоим упомянутым критериям свидетельствует в пользу информационного подхода к психофизическим законам.

Объясняя феноменальные суждения

Ранее мы видели, что, хотя сознание не может получить редуктивного объяснения, феноменальные суждения — суждения типа «я обладаю сознанием», «не является ли сознание чем-то странным?» и т. п. — могут быть объяснены подобным образом, по крайней мере, в принципе. Это создает некоторую проблему для нередуктивной теории сознания, хотя она и не кажется фатальной. Контринтуитивно то, что эти суждения могли бы быть объяснены без отсылки к самому сознанию, но мы можем научиться жить с этим. Однако мы по-прежнему можем надеяться, что объяснение феноменальных суждений будет каким-то глубоким образом связано с объяснением самого сознания. Полная независимость этих двух объяснений казалась бы чем-то иррациональным и отдаленным во власть случая.

Мы можем сформулировать это в виде требования *объяснительной когерентности* от теории сознания. Исчерпывающая теория ментального должна содержать как (нередуктивное) объяснение сознания, так и (редуктивное) объяснение того, почему мы выносим суждения о наличии у нас сознания, и разумно ожидать, что два этих объяснения будут *когерентны* друг другу. В частности, мы могли бы ожидать, что те процессуальные характеристики, которые главным образом ответственны за продуцирование феноменальных суждений, будут нести

основное бремя ответственности и за само сознание. В таком случае окажется, что даже если само сознание и не привлекается для объяснения феноменальных суждений, в этом объяснении будут задействоваться некие корни сознания.

Конечно, нельзя *доказать*, что теория сознания должна соответствовать этому требованию, но любая теория сознания, удовлетворяющая ему, будет отличаться от других теорий в выигрышную для нее сторону. Если какая-то теория покажет, каким образом объяснение феноменальных суждений существенным образом задействует объяснительный базис сознания, то нам удастся сплести их в более унифицированную картину ментального, отчасти устранив ощущение поразительной случайности их связи.

Мне часто казалось, что это могло бы быть ключом к отысканию теории сознания². Во-первых, мы должны приложить все усилия для понимания того, почему порождаются суждения о сознании. Этот вопрос мог бы оказаться непростым, но здесь не следует ожидать каких-то глубоких метафизических тайн; в принципе этот вопрос относится к сфере когнитивной науки. Затем мы должны выделить ключевые черты такого объяснения и подумать, какую роль они могли бы играть в теории самого сознания. Нет гарантии, что это приведет к удовлетворительной теории сознания, но тем не менее данная стратегия представляется перспективной.

Поиск редуктивного объяснения наших суждений о сознании, скопее всего, будет полезным в любом случае и будет относиться к числу наиболее ценных применений редуктивных методов при разработке теории сознания. Мы могли бы сфокусироваться на том, почему система обработки данных должна продуцировать суждения о наличии сознания и, в особенности, о том, что сознание есть такой странный феномен. Я уже сказал несколько слов на эту тему в главе 5; сейчас я хочу обсудить это более детально. «Объяснение», которое я дам, это не более чем правдоподобно звучащая и ни к чему не обязывающая вариация, но можно надеяться, что с помощью эмпирических исследований ее можно было бы развить до состояния детализированной теории. Обращение когнитивной науки и нейронауки к этим феноменам скорее всего оказалось бы очень результативным.

Итак, оставим на мгновение в стороне само сознание и сконцентрируемся на когнитивной системе обработки данных с позиции третьего лица. Если угодно, представим, что это объяснениедается для зомби. Почему мы могли бы *ожидать*, что данная система должна будет порождать суждения такого рода? Какие процессы, к примеру, должны содействовать суждению о наличии ощущения цвета? Чтобы обдумать это, рассмотрим, что могло бы происходить в ситуации, в которой мы воспринимаем цвета.

Если не вдаваться в микроскопические детали, то происходит примерно следующее. Какой-то спектральный световой пакет попадает в наши глаза и активирует различные типы клеток сетчатки. Три вида колбочек вычисляют информацию в соответствии с количеством света в различных перекрывающихся диапазонах длин волн. Многие различия, присутствующие в изначальной световой волне, сразу же теряются. Эта информация передается по зрительному нерву в зрительную кору, где она претерпевает дальнейшую трансформацию посредством нейронных процессов, превращаясь в информацию, соответствующую значениям на трех осях: к примеру, на красно-зеленой, желто-синей и ахроматитической. То, что происходит в дальнейшем, пока не очень понятно, но дело представляется таким образом, будто информация, соответствующая положению данного цвета в этом трехмерном пространстве, удерживается до ее окончательной категоризации в привычные категории «красного», «зеленого», «коричневого» и т. п. К этим ярлыкам прикрепляются вербальные категории, и в итоге формируется отчет, такой как «сейчас я вижу красное».

Посмотрим теперь с «позиции системы» на то, что происходит. Какого рода суждения она сформирует? Конечно, она сформирует такое суждение, как «здесь имеется красный объект», но, если речь идет о разумной, рефлексивной системе, то мы могли бы также ожидать, что она сможет отрефлексировать сам процесс восприятия. И можно было бы поинтересоваться, какое «воздействие» оказывает восприятие на эту систему?

Ключевой момент состоит здесь в том, что, когда данная система воспринимает красный объект, центральные процессы в ней не имеют прямого доступа к самому объекту и к физическим процессам, фундирующими восприятие. Эти процессы имеют доступ лишь к самой информации о цвете, представляющей собой просто некую локализацию в трехмерном информационном пространстве. Когда надо дать лингвистический отчет о ситуации, система не может сказать: «Это пятно насыщено 500- и 600-нанометровыми отражениями», так как всякий доступ к изначальным длинам волн у нее утрачен. Аналогичным образом она не может отчитаться о нейронной структуре: «В данный момент пикивая частота составляет 50 Гц», так как у нее нет прямого доступа к нейронным структурам. Эта система имеет доступ только к локализации в информационном пространстве.

В самом деле, если говорить о центральных процессах, то она попросту *находится* в каком-то участке этого пространства. Система способна производить различия и она *знает* об этом, но лишена представления о том, каким образом она делает это. Можно было бы ожидать, что через какое-то время она сможет именовать различные

локации, с которыми она работает, — «красное», «зеленое» и т. п. — и что она сможет знать, в каком именно состоянии она находится в данный момент. Но, столкнувшись с вопросом, *как* именно она знает об этом, она сможет сказать лишь: «Я просто знаю, непосредственно». Если спросить ее: «В чем состоит различие между этими состояниями», она сможет ответить лишь «Они просто различны», «Оно одно из *тех*», или «Это — красное, а то — зеленое». Если настаивать и спросить, что это значит, она сможет сказать лишь «Они просто различны, качественно». Что еще она могла бы сказать?

Естественно предположить, что система, которая может непосредственно знать о своей локализации в информационном пространстве, не имея доступа к какому-то дополнительному знанию, будет попросту маркировать свои состояния как изначально и элементарно различные, различающиеся по своему «качеству». И, разумеется, следует ожидать, что эти различия будут «непосредственно» воздействовать на систему: она оказывается в этих состояниях, которые, в свою очередь, оказываются непосредственно доступными для организации последующих процессов; и, к примеру, в знании этой системы о том, в каком состоянии она находится, нет ничего выводного. Нам также следует ожидать, что эти состояния будут совершенно «невыразимы»: система лишена доступа к какой-либо дополнительной релевантной информации, и поэтому она может лишь указывать на сходства состояний и их отличия друг от друга, и на различные связи, которые могли бы у них быть. И уж точно не надо ожидать, что она смогла бы объяснить это «качество» в более фундаментальных терминах.

Можно было бы возразить, сказав, что данная система могла бы быть сконструирована таким образом, что она имела бы доступ к указанной информации на уровне «догадок», примерно так, как у субъектов со слепым зрением. Быть может, она могла бы сказать «Суждение „красное“ просто пришло мне на ум», не высказывая каких-либо утверждений о «качестве». Но это устройство скорее всего было бы неэффективным, так как системе надо было бы ждать догадки. А как быть с ситуациями — когда, к примеру, кто-то играет в теннис — когда на визуальную информацию надо реагировать без формирования суждений? Система, вероятно, сказала бы: «Я просто обнаружила, что знаю, где был мяч и что надо делать, но при этом у меня не было никаких *переживаний* на этот счет». Не исключено, что это непротиворечивый сценарий, но соответствующее устройство не выглядит естественным для когнитивной системы. При проектировании такой системы гораздо более естественным было бы сделать так, чтобы она просто «видела» само различие красного и зеленого, непосредственно основывала свое поведение на этом воспринятом различии и давала уверенные и прямые ответы при постановке соответствующих вопросов. В любом случае, такой

дизайн является по крайней мере *одним* из разумных путей устройства подобной системы, и это все, что требуется нам в данном случае.

Принимая во внимание такую непосредственность доступа к информационным состояниям, естественно будет ожидать, что данная система будет использовать язык «опыта» и «качества» для характеристики ее собственного когнитивного взгляда на восприятие. И неудивительно, что все это будет казаться системе довольно странным: непосредственно известные, невыразимые состояния, которые кажутся столь важными для ее доступа к миру, но которые так трудно пришипить. В самом деле, естественно предположить, что странность этого для рассматриваемой системы будет мало отличаться от странности сознания для нас самих.

Таково начало потенциального редуктивного объяснения наших суждений о сознании: эти суждения продаются потому, что наша система обработки данных расположена в информационном пространстве так, что она имеет прямой доступ к тем местам, в которых она расположена, но ни к чему больше. Это прямое знание будет воздействовать на систему как элементарное «качество»: она знает о различии состояний, но, по сути, не может артикулировать это знание иначе, чем сказав «одно из *тех*». Такой непосредственный доступ к элементарным различиям приводит к суждениям о таинственной элементарной природе этих качеств, о невозможности объяснить их в более фундаментальных терминах и ко многим другим суждениям, нередко высказываемым о сознательном опыте.

И во всем этом ключевую роль играет информация. Разного рода суждения об элементарных «качествах» формируются потому, что данная система имеет доступ только к информационным состояниям. Система попросту оказывается в различных состояниях, и последующие процессы имеют доступ только к структуре различия этих более ранних состояний — и ни к чему более конкретному. Работу здесь производит система небезразличных различий. Именно информация и наш доступ к ней редуктивно объясняют суждения, которые мы выносим о сознании.

Кто-то мог бы поставить здесь точку, заявив об устранении тайны сознания и о наличии объяснения. Разумеется, я не считаю это верным: мы объяснили лишь некоторые суждения, что гораздо более просто. Но теперь мы можем использовать принцип объяснительной когерентности для определенного продвижения по пути создания теории сознания. Если главную ответственность за наши суждения о сознании несут информационные состояния, реализованные в этих процессах, то не исключено, что эти информационные состояния отвечают и за само сознание.

На самом деле именно так я изначально и пришел к информационной концепции сознания. Если объяснительный базис наших феноменальных суждений заложен в структуре небезразличных различий, то естественно предположить, что там же должен быть заложен и объяснительный базис сознания. Это объясняло бы, почему наши суждения так хорошо согласуются с актуально существующими состояниями сознания. Сознательный опыт является собой реализацию какого-то информационного состояния; феноменальное суждение объясняется другой реализацией того же самого информационного состояния. И в известном смысле для того чтобы удостовериться, что эти суждения по-настоящему верны, нам нужно лишь допустить наличие феноменального аспекта информации: качественный аспект этой информации, напрямую проявляющийся в феноменологии, а не только в системе суждений, действительно *существует*. Так что это позволяет сознанию очень хорошо сочетаться с когнитивной структурой и приводит к более интегрированному видению ментального.

Мы можем также отметить хорошее сочетание когнитивной роли информационных состояний и эпистемологии опыта. Непосредственно знакомым нам опытным переживаниям соответствуют физически реализованные информационные состояния, к которым система имеет (прямой) когнитивный доступ. Система формирует свои феноменальные суждения на основе ее прямого доступа к информационным состояниям: эта каузальная связь удачно сочетается с утверждением, что опыт — феноменальная реализация того же информационного состояния — *обосновывает* сформированные ею феноменальные убеждения. Ключевую роль с обеих сторон играет одно и то же информационное состояние; просто в одном случае речь идет о физической реализации, в другом — о феноменальной.

Ничто из сказанного в полной мере не доказывает верности информационного подхода к сознанию. Но оно дополнительно подкрепляет его.

4. Повсеместен ли опыт?

К этому моменту читатели, наверное, уже выстраиваютя в очередь, чтобы сделать возражение в связи с *повсеместностью* информации. Мы находим информацию везде, а не только в системах, обычно признаваемых сознающими. Мой проигрыватель компакт-дисков реализует информацию, двигатель моей машины реализует информацию; даже термостат реализует информацию. В действительности, при данной мной трактовке этого понятия, мы находим информацию везде, где находим причинность. Мы находим причинность повсюду, значит повсюду мы находим и информацию. Но ведь мы уж точно не находим повсюду опыт?

Сторонник информационного подхода мог бы двояко отреагировать на эту ситуацию. Первая и наиболее очевидная реакция состоит в том, чтобы заняться поиском дальнейших ограничений, накладываемых на то, *какого рода* информация оказывается релевантной для опыта. С опытом связано не всякое физически реализованное информационное пространство, а лишь те из них, которые наделены определенными свойствами. Это потребовало бы тщательного рассмотрения того, чем могли бы быть эти дополнительные ограничения и как они могли бы вписываться в фундаментальные законы. Я рассмотрю подобные стратегии позже, но сейчас я хочу поговорить об альтернативном пути. На этом пути мы решаемся признать, что всякая информация связана с опытом. И если это так, то повсеместна не только информация. Опыт тоже оказывается повсеместным.

Если данное предположение верно, то опыт связан даже с очень простыми системами. Эту идею часто называют странной или даже безумной. Я, однако, думаю, что она заслуживает тщательного изучения. Для меня не очевидно, что она ошибочна; в известном смысле она даже привлекательна. Так что здесь я собираюсь рассмотреть причины, по которым можно было бы отвергнуть эту концепцию, чтобы выяснить, насколько они убедительны, — одновременно рассматривая разного рода позитивные соображения, которые могли бы мотивировать нас всерьез относиться к ней.

Каково это — быть термостатом?

Чтобы не размывать картину, возьмем в качестве примера почти что до предела упрощенную систему обработки информации: термостат.

Если рассматривать его в качестве устройства, обрабатывающего информацию, то термостат имеет всего лишь три информационных состояния (одно состояние приводит к охлаждению, другое — к нагреванию, третье — к бездействию). И тезис, таким образом, состоит в том, что каждому из этих информационных состояний соответствует феноменальное состояние. Три этих информационных состояния будут отличаться друг от друга, и изменение информационного состояния будет приводить к изменению феноменального состояния. Мы могли бы задаться вопросом: каков характер этих феноменальных состояний? Иначе говоря, каково это — быть термостатом?

Наверняка быть термостатом не очень интересно. Обработка информации настолько проста, что мы должны ожидать не большей сложности и от соответствующих феноменальных состояний. Здесь будет три элементарно отличающихся друг от друга феноменальных состояния, лишенных какой бы то ни было дополнительной структуры. Не исключено, что эти состояния можно мыслить по аналогии

с нашими ощущениями черного, белого и серого: феноменальное поле термостата может быть полностью черным, полностью белым или полностью серым. И даже это было бы слишком щедрой трактовкой структуры переживаний термостата, так как визуальное поле предполагается имеющим какую-то размерность, а черное, белое и серое имеют сравнительно богатые характеристики. На деле надо было бы ожидать чего-то гораздо более простого, для чего не существует аналога в нашем опыте. Мы едва ли смогли бы лучше симпатически представить эти переживания, чем может представить зрение человек, который лишен его, или чем люди могут представить, каково это — быть летучей мышью; но мы хотя бы можем что-то интеллектуально знать о фундаментальной структуре соответствующих переживаний.

Чтобы эти идеи не казались столь безумными, мы можем поразмыслить о том, что могло бы случиться с опытом при смешении вниз по шкале сложности. Мы начинаем с привычных случаев, с людей, у которых крайне сложные процессы обработки информации порождают хорошо нам известные сложные переживания. Если сместиться к менее сложным системам, то едва ли есть серьезные основания сомневаться в том, что собаки или даже мыши обладают сознанием. Некоторые оспаривали это, но, как мне кажется, зачастую это происходило из-за смешения феноменального сознания и самосознания. Мыши могут быть практически лишены чувства Я, и им может быть недоступна интроспекция, но кажется совершенно правдоподобным, что мыши как-то *переживают* свое существование. Мыши воспринимают свое окружение в таких паттернах информационного потока, которые не так уж сильно отличаются от тех, что имеются в наших собственных мозгах, разве что гораздо менее сложных. И естественно будет предположить, что «перцептивному многообразию» мыши, которое, как мы знаем, у них имеется, соответствует «феноменальное многообразие». Перцептивное многообразие мыши весьма богато — мышь может проводить множество перцептивных различий — и поэтому ее феноменальное многообразие тоже могло бы быть таким. Вероятно, к примеру, что каждому различению, которое может произвести и использовать при восприятии окружения визуальная система мыши, соответствует феноменальное различение. *Доказать*, что так и есть, невозможно, но такое представление о феноменологии мыши кажется наиболее естественным.

При движении вниз по шкале — от ящериц и рыб до слизней — мы видим, что к ним можно применить аналогичные рассуждения. Кажется, нет серьезных оснований предполагать, что феноменология должна исчезнуть при сохранении достаточно сложной перцептивной психологии. Если это происходит, то либо существует радикальный

разрыв между сложными переживаниями и их полным отсутствием, либо где-то на указанном пути феноменология начинает рассогласовываться с восприятием, так что на каком-то этапе относительно богатое перцептивное многообразие сопряжено с гораздо более бедным феноменальным многообразием. Первая гипотеза выглядит малоправдоподобной, а вторая предполагает, что внутренние переживания промежуточных систем будут странным образом оторваны от их когнитивных способностей. Альтернатива уж точно, по крайней мере, не менее правдоподобна. Быть рыбой, вероятно, гораздо менее интересно, чем быть человеком, так как ее более простой психологии соответствует более простая феноменология, но кажется достаточно основательным предположение, что *как-то* быть ею все-таки можно.

Когда же должно исчезнуть сознание при смещении по шкале от рыб и слизней к простым нейронным сетям и, наконец, к термостатам? Феноменология рыб и слизней, скорее всего, будет не примитивной, а относительно сложной, отражая разного рода различия, которые они могут производить. Перед полным исчезновением феноменологии мы, по-видимому, будем иметь дело с какой-то предельно простой феноменологией. Мне кажется, что наиболее естественным местом для такой феноменологии является система с соответствующей ей простой «перцептивной психологией», такая как термостат. Термостат, похоже, реализует предельно выхолощенные формы таких процессов обработки информации, какие имеются у рыб и слизней, так что не исключено, что он мог бы быть наделен аналогичной феноменологией в ее до предела выхолощенной форме. Он производит одно-два релевантных различия, от которых зависит действие; и по крайней мере мне не кажется безосновательным утверждение, что с ними могли бы быть связаны переживаемые различия.

Разумеется, по мере движения вниз по шкале сложности все *могло бы* складываться и как-то иначе, и речь в данном случае не идет о демонстративном доказательстве того, что термостаты *должны* иметь переживания. Но это кажется одной из небезосновательных версий, а если подумать, то и не менее естественной, чем какая-либо из альтернатив. Можно попробовать показать, что рассуждения, используемые нами здесь, являются продолжением рассуждений, опираясь на которые мы приписываем опытные переживания собакам и мышам. По крайней мере, если начать думать о том, что могло бы происходить в опыте мыши, и об основаниях этого в ее перцептивной психологии, то переход к более простым системам покажется гораздо более естественным, чем могло бы казаться вначале.

Те, кто считает «безумным» предположение о том, что термостат мог бы иметь переживания, должны хотя бы объяснить, *по какой*

именно *причине* это предположение является безумным. Дело должно обстоять таким образом, предположительно, потому, что термостат лишен какого-то свойства, с очевидностью требующегося для наличия опытных переживаний; но, на мой взгляд, мы не находим подобных очевидных свойств. Быть может, при обработке информации задействуется какой-то ключевой ингредиент, отсутствующий у термостата и имеющийся у мыши, или отсутствующий у мыши, но имеющийся у людей, но я не вижу такого ингредиента, который с очевидностью требовался бы для опыта, и на самом деле не очевидно, что подобный ингредиент вообще должен существовать.

Конечно, сказать, что термостаты наделены опытом, не означает сказать, что у них есть реальное подобие ментальной жизни. Термостат не будет обладать *самосознанием*; у него не будет даже малейших признаков разума; и я не стал бы утверждать, что термостат может *мыслить*³. Сопротивление идеи сознающего термостата может отчасти объясняться сопряжением опыта и других ментальных способностей; все такие способности почти наверняка требуют гораздо большей сложности. Все эти характеристики включают солидный психологический компонент, и для поддержания релевантных каузальных ролей наверняка потребуется какая-то сложная система. Но когда мы развели феноменальные и психологические свойства, идея сознательного термостата уже не кажется такой страшной. Нам надо вообразить лишь что-то вроде неартикулированной «вспышки» опыта, без понятий, мыслей или сложной обработки данных.

Другой причиной, по которой кто-то может отвергнуть идею сознающего термостата, является то, что он не может найти в этой системе какого-либо *места* для сознания. Она кажется слишком простой, и возникает ощущение, что сознание не может играть в ней никакой роли. Но такая реакция свидетельствует о том, что уроки нередуктивной позиции остались неусвоенными: при тщательном исследовании нам никогда не удастся найти сознание в системе, и мы всегда сможем понять ее процессы, не отсылая к сознанию. Если сознание не является чем-то логически супервентным, нам не следует ожидать того, что мы найдем «место» для сознания в организации той или иной системы; сознание совершенно отличается от свойств системы, связанных с тем, как она обрабатывает свои данные.

Может быть и так, что кто-то не захочет признавать возможность сознающих термостатов просто потому, что мы очень хорошо *понимаем* термостаты. Мы знаем все о процессах обработки их данных, и кажется, что здесь нет оснований задействовать сознание. Но в действительности в этом плане термостаты не отличаются от мозгов. Даже если мы будем в совершенстве понимать процессы обработки данных,

протекающие в мозге, мы, по-видимому, все равно не получим оснований задействовать сознание. Единственное различие состоит в том, что в настоящий момент то, что происходит в мозге, достаточно таинственно для того, чтобы кто-то мог соблазниться предположением, что сознание каким-то образом «локализовано» в тех мозговых процессах, относительно которых у нас еще нет понимания. Но, как я пытался показать, даже понимание этих процессов само по себе не привнесет в имеющуюся у нас картину сознания; так что и в этом случае мозг и термостаты вновь оказываются в одной связке.

Кто-то может быть обеспокоен тем обстоятельством, что он сам мог бы *сделать* термостат, не вкладывая в него сознания. Но это, разумеется, справедливо и по отношению к мозгу, по крайней мере в принципе. Когда мы создаем мозг (скажем, при размножении или развитии), сознание оказывается бесплатным приложением к нему; то же самое будет верным и для термостата. Не следует ожидать, что речь пойдет о размещении сознания как физического компонента системы! Кто-то может беспокоиться из-за того, что термостат не является *живым*; но трудно понять, почему это различие должно иметь такое уж принципиальное значение. Лишенный тела кремниевый мозг, подобный тому, о котором мы говорили в предыдущей главе, судя по всему, не может считаться живым, но мы видели, что он мог бы обладать сознанием. И если аргументы, приведенные в предыдущей главе, верны, то отсутствие в составе термостата биологических компонентов не имеет значения в принципе.

Некое интуитивное сопротивление может быть обусловлено тем фактом, что в термостате, по-видимому, нет места для кого-то или чего-то, могущего *обладать* опытными переживаниями: где в термостате мог бы размещаться *субъект*? Но нам не следует искать на роль субъекта в физических системах какого-то гомункула. Субъект есть система в целом или, точнее, он связан с системой подобно тому, как он связан с мозгом. Найти оптимальную формулировку здесь непросто. Мы не стали бы утверждать, что мой *мозг*, строго говоря, наделен переживаниями, но сказали бы, что эти переживания имеются у *меня*. Но как бы мы ни осмыслили это отношение, можно будет перенести это и на термостаты: строго говоря, лучше, вероятно, говорить не о наличии у термостата переживаний (хотя я и дальше буду выражаться так в неформальных рассуждениях), а о том, что переживания связаны с термостатом. Мы не найдем субъекта «внутри» термостата, но мы не найдем никакого субъекта и внутри мозга.

Вернемся теперь к позитивным моментам, свидетельствующим о том, что простые системы наделены опытными переживаниями: утверждая нечто подобное, мы избегаем необходимости «включения»

сознания на каком-то уровне сложности. В представлении о том, что система, содержащая n элементов, не могла бы обладать сознанием, а система, содержащая $n+1$ элементов, могла бы, есть нечто странное. И мы не можем уходить от решения так, как мы могли бы сделать это, говоря о том, в какой момент кто-то становится «лысым»: в последнем случае, похоже, имеется некая степень семантической неопределенности, но гораздо менее правдоподобным выглядит утверждение, что наличие у системы сознания тоже может быть чем-то неопределенным. (В особенности это справедливо тогда, когда мы занимаем нередуктивную позицию, в соответствии с которой мы не можем эксплицировать факты об опыте в терминах более фундаментальных фактов, подобно тому, как мы эксплицируем неопределенности, связанные с облысением в терминах определенных фактов о количестве волос на голове.) Хотя дело и *могло бы* обстоять таким образом, что сознание включалось бы в какой-то определенный момент, любой такой конкретный момент кажется произвольным, и поэтому теория, избегающая решения этого вопроса, обретает определенного рода простоту.

Последнее соображение в пользу наличия опытных переживаний у простых систем: если опыт действительно является фундаментальным свойством, то его широкое распространение кажется чем-то естественным. Все другие фундаментальные свойства, о которых нам известно, несомненно, присущи даже простым системам и повсюду встречаются во вселенной. Было бы странно, если бы какое-то фундаментальное свойство оказалось впервые реализованным только на относительно позднем этапе истории вселенной и даже тогда — только в отдельных сложных системах. В идее о том, что какое-то фундаментальное свойство должно реализовываться лишь спорадически, нет *противоречия*; но ее альтернатива кажется, при прочих равных условиях, более правдоподобной.

Если с терmostатами связан опыт, то опыт, по-видимому, *повсеместен*: везде, где есть каузальное взаимодействие, есть информация, и везде, где есть информация, имеется опыт. Информацию можно найти в камне — когда он, к примеру, расширяется и сжимается, или даже в различных состояниях электрона. Так что если двухспектрный принцип верен без всяких ограничений, то с камнем или электроном будет сопряжен какой-то опыт.

Панпсихизм до какой степени?

(Я не стал бы утверждать, что у камня *есть* опытные переживания, или что камень *обладает* сознанием, подобно тому, как при неформальных рассуждениях я допускал наличие у терmostата опытных переживаний или обладание им сознанием. Камень, в отличие от терmostата, не рассматривается в качестве системы обработки информации. Он

рассматривается просто как объект, так что связь с опытом оказывается здесь не столь прямой. Быть может, лучше было бы сказать, что в камне *есть* системы, обладающие сознанием: предположительно, в нем имеется много подобных подсистем, опытные переживания которых никак не могут канонически рассматриваться как переживания камня (точно так же, как мои переживания не могут рассматриваться как переживания моего рабочего кабинета). Что же касается термостата, то с ним, напротив, связано каноническое информационное пространство, и поэтому кажется более оправданным говорить о канонических опытных переживаниях термостата. Хотя, конечно, как было отмечено выше, даже эти формулировки несколько неформальны.)

Взгляд, согласно которому опыт есть везде, где имеется каузальное взаимодействие, континтуитивен. Но чем больше мы размышляем об этой позиции, тем более она становится до удивления привлекательной, лучше интегрируя сознание в естественный порядок вещей. Если этот взгляд верен, то сознание не рассыпано осколками в изолированных сложных системах, произвольно продуцирующих богатый сознательный опыт. Оно есть, скорее, какое-то более единообразное свойство универсума, когда очень простые системы имеют очень простую феноменологию, а сложные системы — сложную. Это делает сознание в известном смысле не таким «особенным» и поэтому более рационально фундированым.

Интересный вопрос состоит в том, требуется ли для опыта *действительная* причинность. Может ли термостат иметь опытные переживания, когда он находится в неизменном состоянии (в определенном смысле «каузально обуславливая» какое-то внешнее положение дел, но реально ничего не *делая*)? Или же он имеет переживания только в состоянии изменения? Каузальность, фундирующая опыт в мозге, кажется по большей части деятельной, так как релевантная информация постоянно обрабатывается, нейроны посыпают импульсы и т. д. С другой стороны, не исключено, что различие деятельной и пассивной каузальности не может быть проведено на фундаментальном уровне, и в таком случае их можно было бы рассматривать как равнозначенные. Я не знаю ответа на этот вопрос, но интуитивно кажется, что для опыта требуется какая-то деятельность.

Есть и еще одна возможность, которую я пока не рассматривал, но которую нельзя исключить, а именно возможность того, что простые системы лишены феноменальных свойств, но имеют *протофеноменальные* свойства. В главе 4 я упоминал о возможности существования свойств, более фундаментальных, чем феноменальные свойства, которые могли бы конституировать последние. Если такие свойства и впрямь существуют, то естественной казалась бы их реализация в простых си-

стемах. В таком случае термостаты могли бы и не обладать переживаниями в обычном смысле этого слова, но реализовывать родственный им тип свойства, относительно которого у нас нет полного понимания (некийprotoопыт, быть может). Это сохранило бы упомянутое выше унифицированное представление о мировом порядке и могло бы также содействовать решению проблемы «выключения» (если протофеноменальные свойства фундаментальны, то переживания, конституируемые этими свойствами, все же могли бы постепенно «включаться»). Не утверждая наличия у термостатов развернутых переживаний, эта позиция может также казаться не столь «безумной», как ее альтернатива. Разумеется, ценой этого является допущение класса неизвестных свойств, непонятных нам; но эту возможность нельзя исключать.

Так или иначе, но данная позиция имеет много общего с воззрением, известным как *панпсихизм* — взглядом, согласно которому все наделено ментальными характеристиками. Сам я, как правило, не использую этот термин по ряду причин: (1) поскольку я полагаю, что наличия переживаний может быть недостаточно для наличия того, что мы понимаем под ментальным, хотя их и можно считать ментальным в его простейшей форме; (2) поскольку протофеноменальные свойства могут оказаться еще более далеки от обычного представления о «ментальном»; (3) поскольку я не считаю верным в строгом смысле утверждать, что (к примеру) камни наделены переживаниями — по указанным выше причинам, хотя с камнями и могут быть сопряжены какие-то опытные переживания. Но, возможно, главная причина неадекватности этого термина состоит в том, что он предполагает воззрение, согласно которому переживания, имеющиеся в простых системах, таких как атомы, фундаментальны, а сложные переживания так или иначе оказываются результатом сложения этих более простых переживаний. Хотя так и могло бы быть, нет оснований считать, что дело должно обстоять именно таким образом: сложные переживания могут быть более автономными, чем здесь предполагается. В частности, информационная концепция подразумевает картину, в соответствии с которой сложные переживания детерминируются более холистично, чем тут.

Со всеми этими оговорками, возможно, мы и вправе сказать, что данное воззрение — одна из разновидностей панпсихизма. Должен, однако, отметить, что панпсихизм не является метафизической основой моей концепции; ее основу составляет, скорее, натуралистический дуализм и психофизические законы. Панпсихизм — это всего лишь один из способов возможной реализации естественной супервентности опыта на физическом. В известном смысле, естественная супервентность задает *каркас*, который можно *детализировать* в том числе и с помощью панпсихизма.

Сам я гораздо более уверен в натуралистическом дуализме, чем в панпсихизме. Вопрос о панпсихизме представляется во многом открытым. Надеюсь, однако, что сказанного мной достаточно, чтобы показать, что возможность некоего панпсихизма должна приниматься нами всерьез: против него, как представляется, нельзя выдвинуть решающих аргументов, при том, что существуют различные позитивные доводы в его пользу.

Ограничивающий двуаспектный принцип

Даже если мы готовы признать наличие переживаний у очень простых систем, идея о том, что вся информация связана с опытом по-прежнему могла бы причинять определенное неудобство. Скажем, создается впечатление, что информации, имеющейся в сознательном опыте соответствует лишь небольшая часть информации, которая когнитивно обрабатывается людьми. Не является ли попросту фактом то, что большинство наших процессов обработки информации бессознательно?

Если неограниченный двуаспектный принцип верен, то ответ, предположительно, состоит в том, что вся эта «бессознательная» информация *реализована* в опыте — но только не в *моем* опыте. К примеру, если с одним из моих нейронов связан какой-то опыт — подобно тому, как он связан с термостатом — то нам не следовало бы ожидать, что он будет частью *моего* опыта, так же как мы не ожидали бы, что мой опыт претерпит радикальную трансформацию при замене этого нейрона на маленького сознающего гомункула. Сходным образом опыт может быть связан и с различными «бессознательными» подсистемами в мозге, занимающимися обработкой информации, — но соответствующие опытные переживания принадлежат другому субъекту. В мозге имеется много различных систем обработки информации, и та, что соответствует *мне* — быть может, это система, делающая какую-то информацию доступной для определенного рода глобального контроля и отчета — есть лишь одна из них. Я не мог бы ожидать прямой доступности для меня переживаний других систем, так же как я не жду прямой доступности для меня переживаний других людей.

Кто-то мог бы проявлять беспокойство и по поводу всех относительно сложных систем обработки информации, которые существуют в мире и которые можно найти где угодно, от моего проигрывателя компакт-дисков до моего желудка. Можно ли считать, что все они являются сознательными индивидами, наделенными сложными переживаниями? В ответ стоит заметить, что эти системы лишены даже подобия когерентной когнитивной структуры, присущей нашей системе, так что любые связанные с ними переживания скорее всего совершенно непохожи на наши. Если, к примеру, с проигрывателем компакт-

дисков связаны какие-то переживания, то они, скорее всего, сводятся лишь к «плоской» структуре битов; и если информация, имеющаяся в моем желудке, связана с опытом, то нет оснований считать, что этот опыт соответствовал бы тому, что трактуется нами как ментальное начало. Имеющиеся у нас разновидности опыта возникнут лишь при формировании эволюцией таких систем обработки информации, которые обладают сложными, когерентными когнитивными структурами, отражающими богатое представление о внешнем мире. Вероятно, лишь очень ограниченная группа субъектов опыта обладала бы психологической структурой, требующейся для того, чтобы можно было с полным на то основанием считать их *агентами или личностями*.

Тем не менее столь значительная пролиферация опыта, особенно в пределах одного мозга, могла бы быть причиной дискомфорта. Ситуация усугубляется тем обстоятельством, что при наличии некоего информационного пространства обычно не составляет труда отыскать множество несколько иных информационных пространств — для этого нужно лишь по-другому индивидуализировать релевантный каузальный путь или несколько иначе вычленить релевантные эффекты («различия», производимые информацией). Должны ли мы предполагать, что вокруг всех этих информационных пространств витают различные группы переживаний? Если так, то процессы в моей голове могли бы порождать множество очень близких, но в чем-то и разнящихся феноменальных родственников!

Альтернатива состоит в том, чтобы *ограничить* двухспектрный принцип для сужения класса физически реализованных информационных пространств, имеющих феноменальные корреляты. Наиболее естественной может оказаться стратегия ограничения *способа* обработки информации. В конце концов, я уже говорил, что информация в моей системе, наиболее непосредственным образом соответствующая моему опыту, это информация, напрямую доступная для глобального контроля. В таком виде этот «критерий» наверняка не будет играть никакой роли в фундаментальном законе, так как он выражен слишком размытым высокоДуровневым понятием; в самом деле, мы можем использовать этот принцип лишь в том случае, когда мы уже индивидуализировали высокоДуровневую систему, такую как личность или мозг. Но не исключено, что существует и более точный и простой критерий, который мог бы выполнить подобную работу.

Одна из возможностей состоит в существенности *усиления* информации. Физически реализованная информация реализуется также и в опыте лишь при определенном *усиление* информации, когда ее можно будет использовать для больших изменений в каких-то каузальных путях. Не исключено, что можно было бы даже сказать, что интенсивность опыта соответствует степени усиления, или что-то в этом духе.

Это могло бы хорошо сочетаться с критерием глобальной доступности, хотя здесь могли бы возникнуть другие проблемы: скажем, существует немало примеров усиленной информации, интуитивно не являющейся сознательной; и не очевидно, каким именно образом надо было бы уточнять понятие усиления.

Другая возможность состоит в том, что мы могли бы наложить ограничение на тип *причинности*, реализованный в системе. Мы видели, что везде, где есть причинность, есть и информация; но, быть может, для индивидуализации информационных пространств, фундирующих опыт, подходит лишь определенный тип причинности. Не исключено, к примеру, что для этого годятся лишь какие-то виды «активных» каузальных отношений или какие-то виды «естественных» каузальных отношений. Интуитивно кажется, что многие из информационных пространств, которые можно отыскать в соответствии с выдвигавшимися выше критериями, в известном смысле неестественны; и, быть может, соответствующее ограничение можно было бы как-то прояснить. Это, скорее всего, все равно оставило бы нам очень широкий класс информационных состояний, но могло бы воспрепятствовать их астрономическому умножению.

Я не уверен, каким именно должен быть релевантный ограничительный критерий, но это не означает, что его вообще не могло бы существовать. Могло бы оказаться даже, что ограничительный критерий ограничил бы соответствующие информационные пространства так, что вычеркнул бы информацию в простых системах, таких как терmostаты. Моя интуиция состоит в том, что ограничитель двухспектного принципа может существовать, но он все же мог бы и не вычеркивать информацию в простых системах, таких как терmostаты. Пролиферация множества родственных опытных переживаний в мозге кажется мне более континтуитивной, чем наличие переживаний у простых систем, хотя ни там, ни здесь вопрос нельзя считать предрешенным. В любом случае, при доработке прототеории в теорию были бы возможны самые разные варианты.

5. Метафизика информации

Но остается вопрос: каким образом надо понимать онтологию двухспектной концепции информации? Насколько всерьез мы принимаем эти рассуждения об информационных пространствах и информационных состояниях: есть ли они лишь полезные конструкты, или же они в известном смысле онтологически фундаментальны? Является ли информация первичной, или же первичным в действительности оказывается физическое и феноменальное, а информация — это лишь полезное связующее звено?

Все это можно было бы истолковать множеством разных способов. Наиболее простым и наименее рискованным решением было бы признать физическую реализацию информации и феноменальную реализацию информации совершенно отдельными характеристиками, лишенными онтологической связи помимо законосообразного соединения и некоего структурного изоморфизма. Онтология при этом остается прямолинейной онтологией дуализма свойств, с физическими свойствами, отдельными феноменальными свойствами и законосообразной связью между ними. Рассуждения о «двуаспектности» должны пониматься здесь в дефляционистском ключе: это просто цветастая характеристика двух различных видов коррелятивных свойств со сходной структурой. И информация — всего лишь полезный инструмент, используемый при характеристике этой общей структуры: она не соответствует чему-либо онтологически «глубокому».

Такой взгляд на вещи может быть абсолютно верным, но здесь имеются и более интересные возможности. Большинство из них связано с более серьезным отношением к роли информации. Ниже я рассмотрю один из вариантов реализации этих возможностей. Читатель предупрежден, что обсуждение будет проходить в плоскости спекулятивной метафизики, хотя, если мы хотим разобраться с онтологией сознания, нам едва ли удастся избежать обращения к спекулятивной метафизике.

Физики порой высказывают предположение, что Все из бита информация является чем-то фундаментальным для физики вселенной, и даже что физические свойства и законы могут быть производными от информационных свойств и законов. Эта концепция — «все из бита» — выдвинута Уилером (Wheeler 1989, 1990) и Фридкиным (Fredkin 1990) и рассматривается также в статьях Журека (Zurek 1990) и Метцке (Matzke 1992, 1994). Если это так, мы смогли бы дать информации более серьезную роль в нашей онтологии. Чтобы лучше все это представить, я рассмотрю один из главных путей трактовки информации как чего-то фундаментального в физике. Это не единственный способ преподнесения идей «всего из бита» (в частности, он несколько отличается от трактовки Уилера⁴), но он кажется мне, возможно, самым естественным для обоснования этого понятия. Как мы увидим, эта интерпретация тесно переплетена с «расселиянскими» идеями, обсуждавшимися в четвертой главе (с. 197–200).

Данный подход опирается на то наблюдение, что в физических теориях фундаментальные физические состояния фактически инди-видуализируются как *информационные состояния*. Когда мы смотрим на такие параметры, как масса или заряд, мы видим лишь элементарное пространство небезразличных различий. Физика ничего не говорит

о том, что такое масса, или что такое заряд, она просто указывает ряд различных значений, которые могут иметь эти параметры, и она сообщает нам, какое влияние они могут иметь на другие параметры. Пока речь идет о физических теориях, конкретные состояния массы и заряда с тем же успехом могли бы быть чисто информационными состояниями: значение имеет лишь их локализация в информационном пространстве.

Это отражается в том факте, что физика не накладывает на себя никаких обязательств относительно способа *реализации* этих состояний. Для целей физической теории будет пригодна любая реализация этих информационных состояний, не нарушающая структуру каузальных или динамических отношений между ними. Ведь, в конце концов, пока форма этих отношений остается неизменной, физическое будет *выглядеть* одинаково для наших перцептивных систем: у нас нет доступа к каким-либо дополнительным свойствам, реализованным в физическом мире, кроме как к форме каузальной сети (исключение возможно лишь в том случае, если наши феноменальные свойства напрямую увязаны с реализующими свойствами).

Иногда даже говорили, что вселенная могла бы быть гигантским компьютером. Фридкин (Fredkin 1990) допускал, что вселенная могла бы быть колоссальным клеточным автоматом, реализованным на базовом уровне в громадной структуре битов⁵. Высказывалось и предположение (Leckey 1993), что все пространство-время могло бы основываться на вычислительном процессе, с отдельными регистрами для каждого из фундаментальных параметров, реализованных в мире. Покуда у этих регистров будут сохраняться надлежащие каузальные отношения, существа этого мира не будут замечать изменений. Пример с компьютером иллюстрирует наличие великого множества возможных способов реализации «известных» нам физических сущностей; важно лишь, чтобы какие-то сущности играли надлежащие каузальные роли. Это можно считать частью «метафизики физики»: спекулятивных рассуждений об онтологии, фундирующей каузальную структуру самого пространства-времени.

Сама физика, конечно же, не занимается такого рода метафизикой. Физика может оставаться совершенно нейтральной по отношению к этим вопросам о реализации тех параметров, о которых она говорит, а в действительности и по отношению к вопросам о том, «реализованы» ли подобным образом эти параметры вообще. В той мере, в какой речь идет о физике, состояние мира могло бы и исчерпываться информационным описанием. Если в нем и есть какие-то дополнительные базовые «реализующие» свойства, они не играют прямой роли в физических теориях. Так что могло бы возникнуть искушение вообще избавиться от них.

Это привело бы к картине мира как мира *чистой* информации. Каждому фундаментальному параметру такого мира соответствует информационное пространство, и при допущении физикой реализации этих параметров реализуется и информационное состояние соответствующего пространства. Пока эти информационные состояния соотносятся друг с другом надлежащим образом, все будет, как должно. Согласно этой картине мира, сказать тут больше нечего. Кроме информации, ничего нет.

Именно так я понимаю мир, где «все из бита». Эта концепция не лишена какого-то странного очарования: мир как чистый информационный поток, где нет никакой дополнительной субстанции. (В некоторых вариациях этой концепции может допускаться также *пространство-время* как элементарный каркас с внедренными в него информационными пространствами; в других версиях само пространство-время рассматривается как то, что конституируется отношениями информационных пространств.) Мир есть всего лишь мир элементарных различий, а также каузальных и динамических соотношений между ними. Согласно этой концепции, ошибочным было бы пытаться сказать о мире что-то еще.

С этой картиной мира связаны две главные проблемы. Первая идет от самого сознания. Кажется, что в нем есть что-то выходящее за пределы *чистого* информационного пространства. Феноменальные свойства имеют внутреннюю природу, не исчерпываемую их расположением в информационном пространстве, и возникает ощущение, что чисто информационное видение мира не оставляет места для этих внутренних качеств.

Вторая проблема состоит в неочевидности когерентности понятия чистого информационного потока. Может возникнуть чувство, что при такой картине миру слишком уж недостает субстанциальности, чтобы быть миром. Могут ли существовать *элементарные* различия, не фундированные различиями какого-либо качества, лежащего в их основании? Правдоподобным могло бы казаться утверждение, что каждое конкретное различие в мире должно основываться на чем-то, то есть быть различием *чего-то*.

Эта проблема тесно связана с проблемами концепции «чистого каузального потока», которая обсуждалась в главе 4 (с. 197), вариацией на тему которой является данное возврзение. Та концепция лишала мир всех внутренних свойств, оставляя мир каузальных отношений, в котором, какказалось, ничто не производит каузальную работу. Данная концепция может выглядеть несколько лучше, допуская информационные состояния, соотносимые каузальными отношениями, хотя

**Фонтируя
информацию
в феноменологии**

примечательна несубстанциальность этих состояний, просто *отличных* друг от друга и лишенных собственной природы. Кому-то могло бы показаться, что эта картина мира, лишенного внутренней природы, вообще не является картиной мира.

В самом деле, можно было бы попробовать доказать, что информационные пространства *должны* иметь какую-то дополнительную природу. Два фундаментальных свойства могли бы иметь однотипную информационную структуру, будучи, к примеру, реальными величинами в континууме. Если физика — это чистая информация, то ничто не позволило бы различить существующие образцы двух этих информационных пространств. Но *какое-то* различие между ними должно быть, так как два этих свойства задействованы в разных законах и производят разные воздействия на другие характеристики мира. Поэтому для их различия должны иметься дополнительные параметры; параметры, выходящие за пределы чистой информации. Кажется, что для этого нужно какое-то внутреннее качество.

Эти проблемы можно было бы попробовать разрешить самыми разными способами. Можно было бы сказать, что вторая проблема все же не является фатальной и удовлетвориться физикой чистой информации; а затем можно было бы попытаться инкорпорировать феноменальные свойства, представив их неким образом закономерно связанными с этой информацией. Альтернативный путь предполагает решение второй проблемы допущением внутренних свойств, фундирующих физические информационные пространства, что же касается первой проблемы, то она решалась бы введением феноменальных свойств по независимым соображениям.

Наиболее интригующей стратегией, однако, была бы попытка одновременного решения двух проблем. Первая проблема предполагает, что феноменальные свойства демонстрируют наличие у нас непосредственного знания о некоей внутренней природе мира, не охватываемой чистой информацией; вторая — что мы можем *нуждаться* в какой-то внутренней природе мира для фундирования информационных состояний. И в таком случае могло бы оказаться, что внутренняя природа, требующаяся для фундирования информационных состояний, тесно связана с той внутренней природой, которая открывается нам в феноменологии. Не исключено даже, что одна конституирует другую. Тогда мы получаем на выходе экономную и элегантную онтологию, и одним ударом решаем обе проблемы.

Это, опять же, родственно расселовскому предложению о феноменальности (или протофеноменальности) неизвестных внутренних свойств мира, о котором шла речь в главе 4. Рассел нуждался в этих свойствах для фундирования каузальных отношений, о которых гово-

рит физика, а мы нуждаемся в них здесь для фундирования информационных состояний (небезразличных различий), постулируемых физикой. Это, по сути, одна и та же проблема. В обоих случаях возникает ощущение, что мы получаем два решения по цене одного. Нам нужны какие-то внутренние свойства для осмыслиения физического мира, и нам надо найти какое-то место для внутренних свойств, демонстрируемых феноменологией. Две эти проблемы, как представляется, хорошо стыкуются между собой.

В общем, идея в том, что информационные пространства, требуемые физикой, сами фундированы феноменальными илиprotoфеноменальными свойствами. Каждая реализация подобного информационного пространства на деле является *феноменальной* (или protoфеноменальной) реализацией. Всякий раз, когда реализуется такая характеристика, как масса или заряд, за ними стоит какое-то внутреннее свойство: феноменальное или protoфеноменальное, или, для краткости, *микрофеноменальное* свойство. У нас будет набор базовых микрофеноменальных пространств, по одному на каждое фундаментальное физическое свойство, и именно эти пространства будут фундировать информационные пространства, требуемые физикой. Предельные различия и есть эти микрофеноменальные различия.

Разумеется, эта позиция опять-таки предполагает некую разновидность «возмутительного» панпсихизма, но я уже доказывал, что подобный панпсихизм вовсе не столь безоснователен, как обычно предполагается. Если принять уже сделанное мной допущение о том, что феноменальные свойства могут иметься везде, где есть информация, то мы вполне могли бы предоставить этим свойствам возможность сыграть полезную роль.

Онтология, к которой мы результатом приходим, могла бы быть по праву названа двуаспектной онтологией. Физика нуждается в информационных состояниях, но интересуется только их отношениями, а не их внутренней природой; феноменология нуждается в информационных состояниях, но интересуется только их внутренней природой. Эта концепция допускает наличие одного-единственного базового набора информационных состояний, объединяющих оба этих момента. Можно было бы сказать, что внутренние аспекты этих состояний — феноменальны, внешние — физичны. Или в виде слогана: опыт есть информация изнутри; физика — извне.

Все это замечательно работает в онтологии, хотя, конечно, выглядит и диковато. Но пока мы не зашли слишком далеко, обратим внимание на один колossalный вопрос, остающийся пока без решения: как можно совместить эту онтологию с деталями

А что же макроскопическая феноменология?

психофизической *теории*? В частности, как можно совместить ее с психофизическими регулярностями на макроскопическом уровне? Трудность в том, что двухспектральный принцип применяется здесь на фундаментальном физическом уровне, где *микроскопическая* физически реализованная информация имеет и феноменальную реализацию. Но для теории сознания нам нужно, чтобы феноменальную реализацию имела и *макроскопическая* физически реализованная информация. И во все не очевидно, что подобную «макроскопическую феноменологию» можно извлечь из микроскопической феноменологии.

Наш сознательный опыт вовсе не кажется какой-то суммой микрофеноменальных свойств, соответствующих, к примеру, фундаментальным физическим характеристикам нашего мозга. Наш опыт выглядит гораздо более цельным и гораздо более гомогенным, чем какая-то простая сумма. Это одна из версий «проблемы грануляции», которая была поставлена Селларсом (Sellars 1965) перед материализмом: как опыт мог бы быть тождественным громадному множеству физиологических событий, принимая во внимание гомогенность первого и мелкую гранулированность вторых? Аналогичная проблема особенно актуальна для расселовских воззрений того рода, который я сейчас обсуждаю⁶. Если корни феноменологии исчерпываются микрофеноменологией, то непросто понять, откуда берется гладкая, структурированная макроскопическая феноменология: вместо нее можно было бы ожидать «зазубренной» бесструктурной феноменальной кучи.

Можно по-разному пытаться решить этот вопрос. Во-первых, можно было бы попробовать представить дело так, будто двухспектральная онтология работает на всех уровнях, а не только на микроскопическом уровне. Иначе говоря, даже физические информационные пространства на макроскопическом уровне фундированы какой-то феноменальной реализацией. Можно доказывать, что микроскопический уровеньщен каких-либо привилегий: здесь все проще, но это не значит, что он играет особую онтологическую роль. Аргументы, приведенные нами в поддержку информационного видения физического мира, действенны и на макроскопическом уровне. Можно было бы попытаться показать, что даже на этом уровне есть место для небезразличных макроскопических различий, каждое из которых могло бы быть реализовано в соответствующей ему феноменологии.

Трудность в том, что здесь может не оставаться места для всех этих отдельных феноменальных реализаций. Если фундаментальные физические характеристики реализованы в феноменальных информационных пространствах, то кажется, что этим уже фундируется макроскопическая информация: небезразличные различия здесь и теперь фундированы конфигурациями микроскопических физических черт,

которые, в свою очередь, фундированы микрофеноменологией. И хотя все равно можно было бы попытаться допустить отдельное феноменологическое фундирование, оно было бы избыточным и менее теоретически элегантным, чем аналогичный шаг в микроскопической сфере. Можно было бы попробовать уйти от избыточности, сделав макроскопическое фундирование *изначальным*, но тогда было бы не-просто объяснять изолированные микроскопические системы и т. п. Так что неясно, может ли «фундирующий подход» к двуаспектной онтологии напрямую работать на макроскопическом уровне.

Во-вторых, можно было бы попытаться понять, как макроскопическая феноменология могла бы *конституироваться* этими микрофеноменальными свойствами. На первый взгляд, она не является простой суммой или совокупностью этих свойств; это напрямую вело бы к проблемам с «зазубренностью». Впрочем, как предположил Нагель (Nagel 1986), проблема может быть именно в том, что мы не понимаем отношение части и целого в ментальной сфере. То есть у нас отсутствует точное представление о том, каким образом низкоуровневые феноменальные свойства «складываются» в высокуюровневую феноменологию. Мы склонны мыслить это в терминах физических аналогов, отталкиваясь от того, как микрофизика складывается в макрофизику, но такой способ мышления может быть неверным. Не исключено, что феноменология конституируется совершенно иначе.

Микрофеноменальные свойства могли бы, к примеру, складываться в макрофеноменологию способом, отражающим их общую *информационную*, а не пространственно-временную структуру. Если совокупность этих свойств совместно реализует сложное информационное состояние вследствие имеющихся между ними каузальных отношений, то мы, быть может, могли бы ожидать, что любая производная феноменология будет иметь форму этого информационного состояния. Ведь главной ролью микрофеноменологических свойств является реализация информационных состояний, и поэтому не вызвало бы большого удивления, если бы информационные структуры играли какую-то роль в конститутивных отношениях между данными свойствами. В таком случае любые производные макрофеноменальные состояния имели бы «гладкую» информационную структуру, которая и предсказывается первоначальным двуаспектным принципом. Уяснить это *непросто*, но, в конце концов, мы и не можем ожидать, что наше обыденное представление о физическом будет применимо к области феноменального. Так что более точная картина природы самой феноменологии может быть совместимой с таким конституированием.

Если бы оказалось, что никакое конститутивное отношение не могло бы давать такой результат, можно было бы опробовать третий

вариант, привязывающий макрофеноменологию к микрофеноменологии *законами*⁷. К примеру, это мог бы быть попросту закон, что при реализации микрофеноменальными состояниями определенного информационного состояния вследствие каузальных отношений между ними (принципа «небезразличного различия») будет возникать непосредственная феноменальная реализация того же самого состояния. Это решило бы теоретические проблемы, но ценой усложнения онтологии. У нас больше не было бы простой онтологии, где феноменология оказывается внутренним аспектом физически реализованной информации: какие-то феноменальные вещи «болтались» бы в привязке к этой информации законами, подобно тому, как мы видим это в более привычных версиях дуализма свойств. Привлекательность расселовского видения была бы, таким образом, частично утрачена, хотя концепция по-прежнему была бы вполне когерентной.

В любом случае, я оставил этот вопрос открытым. Несомненно, что эта самая трудная из проблем, стоящих перед любой концепцией расселовского типа; но не очевидно, что она не может быть решена. Если с ней можно справиться, то особенно перспективной кажется вторая стратегия; хотя не исключено, что для решения этой проблемы нужна какая-то совершенно новая идея. Оптимистично глядя на вещи, мы можем рассматривать эту проблему — как сделать психофизическую теорию совместимой как с макроуровневыми фактами о нашей феноменологии и ее физической основе, так и с микроуровневой онтологией расселовской концепции — в качестве одного из главных ограничений, в итоге могущих привести нас к созданию детализированной теории сознания. Одна из трудностей при создании подобной теории состоит в недостатке ограничений. Не исключено, что эта проблема могла бы направить нас в нужное нам русло.

Если бы все рассмотренные стратегии оказались плохими, нам пришлось бы отойти от расселовского видения и обратиться к какой-то другой метафизической концепции. Так, можно было бы попробовать поработать с метафизикой чистой информации как одним из способов понимания физического мира; а затем как-то подключить к ней феноменологию, возможно, путем законосообразной связи с чистой информацией. Или можно было бы просто вернуться к «прирученной» онтологии с отдельными областями физического и феноменального, где каждая из них имеет свою собственную внутреннюю природу, а их связь обеспечивается законосообразными соединениями в духе информационного принципа. Это означало бы, что к рассуждениям о «двуаспектности» надо относиться не столь серьезно, и онтология оказалась бы не такой элегантной, но это по-прежнему могло бы давать нам совершенно удовлетворительную теорию.

6. Открытые вопросы

Представленный мной набросок информационного каркаса психофизических законов оставляет открытым громадное множество вопросов. Эта картина может превратиться в окончательную теорию только при ответе на все эти вопросы. В предыдущем параграфе я упомянул некоторые проблемы, связанные с онтологией этого воззрения. Но можно задать немало вопросов и о форме *законов*, а также о том, как именно должна объясняться наша феноменология. Речь идет в том числе о следующих вопросах:

1. Когда информационное пространство феноменально реализуется, почему оно реализуется так, а не иначе? Скажем, если наше феноменальное цветовое пространство могло бы быть инвертированным, то кажется в чем-то произвольным, что оно именно таково. Нужно ли добавлять законы или допускать контингентные «константы» для решения этой проблемы?
2. Определяется ли характер феноменального информационного пространства структурой этого пространства (пусть и не исключая возможность инверсий)? Могло бы, к примеру, показаться, что как цветовое пространство, так и пространство вкусов являются простыми трех- или четырехмерными пространствами, но, несмотря на сходную форму, их характеры существенно разнятся. Можно попробовать доказать, что сходство их структур — иллюзия, и что если мы встроим их в более крупные структуры — скажем, рассматривая переживания цвета в качестве компонентов цельной, глубокой структуры визуальных переживаний, — сходство исчезнет. Но так и остается вопросом, является ли нечто сходное с нашими переживаниями цвета *единственным* вариантом проецирования визуальной информации о цвете в феноменологию, или тут было возможно что-то совсем другое? Подозреваю, что правильным ответом был бы скорее первый, чем второй, но далеко не ясно, как это можно было бы обосновать.
3. Я использовал этот каркас главным образом для обсуждения простых перцептивных восприятий, таких как восприятие цвета. Не очень понятно, как расширять его для рассмотрения более тонких переживаний, к примеру, сложных эмоций или опыта мыслей о чем-то. Можно ли сделать это?
4. Какого рода формальная структура лучше всего подходит для схватывания структуры феноменальной информации? Какого рода топологические пространства нужны для схватывания реляционной структуры опыта? Нужно ли нам обращаться к более конкретным структурам, таким как метрическое пространство или

дифференцируемое многообразие? Комбинаторная структура опыта еще более интересна: к примеру, простой многомерный континуум был бы, возможно, слишком большим упрощением структуры визуального поля. Как же можно было бы лучше всего ухватить всю его структуру? Следует ли для достижения этой цели модифицировать дефиницию информационного пространства?

5. Как, находясь внутри этого каркаса, объяснить *единство сознания*? Иначе говоря, что делает мои зрительные переживания, слуховые переживания и т. д. переживаниями одного и того же субъекта? Подозреваю, что ответ указывал бы на способ обработки соответствующей информации, так что единство сознания соответствует факту доступности для определенной интеграции релевантной информации. Но детали этого неясны.
6. Каковы в точности критерии того, какая информация в мозге соответствует *моим* переживаниям? Существен ли для этого какой-то особый каузальный путь или тип каузального потока? Ключевую роль здесь, в индивидуализации информации и соответствующих путей, предположительно играет что-то вроде прямой доступности для глобального контроля.

Наличие всех этих вопросов как раз и показывает, как далеки эти схематичные идеи от подлинной теории. В этом можно убедиться, заметив также, как далеки эти идеи от того, чтобы давать нам возможность *предсказывать*, какие именно феноменальные свойства, связанные с физической системой, возникнут из физических свойств этой системы. В ее нынешнем виде данная идея лишена большой объяснительной и предсказательной силы: ее надо хорошо прокачать, чтобы сделать реально полезной.

Для превращения этой идеи в удовлетворительную теорию понадобится немало новых прозрений. Не исключено, что к прорыву могло бы привести рассмотрение проблемы, о которой шла речь в предыдущем параграфе: как совместить феноменальную информацию на макроскопическом уровне со взглядом на информацию как на «внутреннее свойство» на микроскопическом уровне. Другой прорыв мог бы произойти при попытке найти ограничение, выделяющее класс физически реализованных информационных пространств, реализованных и в опыте. К прорывам могли бы привести и соображения, связанные с вещами, которых я вообще не касался.

Данная идея может и вообще оказаться совершенно неверной. Это не удивило бы меня; на самом деле я считаю более вероятным, что ключ к фундаментальной теории лежит где-то в другом месте. Но я выдвинул эти идеи потому, что нам надо начинать размышлять об этих

вещах, и потому, что даже неудачный образец таких размышлений может быть поучительным. Я также надеюсь, что некоторые из попутно рассмотренных идей — о том, как объяснить феноменальные суждения, о повсеместности опыта, а также о связи опыта, информации и внутренних свойств физического — могут оказаться полезными даже при переносе их в другой каркас. Не исключено, что более адекватная теория сознания может сохранить какой-то аромат выдвинутых здесь идей, даже если ее детали будут совсем другими.

Нередко можно услышать, что проблема теорий сознания вроде той, что была здесь рассмотрена, состоит в том, что они излишне спекулятивны и не могут быть проверены. Я, однако, считаю, что настоящая проблема выдвинутой мной «теории» в другом: она слишком неспецифична в своих предсказаниях. *Если бы* у нас была примерно такая же в плане своей простоты теория, которая могла бы *предсказывать* все специфические факты о наших опытных переживаниях, — пусть даже и лишь те факты, которые известны нам с позиции первого лица, — при наличии физических фактов о нашей системе обработки информации, это было бы замечательным достижением и дало бы нам самое серьезное основание считать эту теорию истинной. В настоящий момент у нас нет такой теории, но нет резона полагать, что подобная теория вообще является невозможной.

Часть IV

ПРИКЛАДНЫЕ АСПЕКТЫ

Глава 9

Сильный искусственный интеллект

1. Машинное сознание

Может ли машина быть сознательной? Может ли надлежащим образом запрограммированный компьютер действительно иметь ум? Эти вопросы были предметом громадного множества дискуссий в последние десятилетия. Исследования в области *искусственного интеллекта* (или *ИИ*) в значительной своей части были нацелены на воспроизведение ментального в вычислительных машинах. О серьезном прогрессе говорить пока не приходится, но сторонники этого подхода доказывают, что у нас есть все основания считать, что в итоге компьютеры действительно обретут ум. В то же время, оппоненты доказывают, что у компьютеров есть некие ограничения, которых лишены люди, так что не стоит и вопроса о том, чтобы сознающий ум возник в силу одних лишь вычислительных процессов.

Есть два вида типичных возражений против искусственного интеллекта. Это могут быть *внешние* возражения, когда пытаются показать, что вычислительные системы никогда не смогли бы даже *вести себя* подобно когнитивным системам. Согласно этим возражениям, люди обладают определенными функциональными способностями, которые никогда не могли бы быть у компьютеров. Так, иногда утверждается, что, поскольку эти системы следуют правилам, они никогда не смогут демонстрировать творческое или гибкое поведение, свойственное людям (напр., Dreyfus 1972). Другие доказывали, что компьютеры никогда

не смогут воспроизвести математические интуиции человека, так как вычислительные системы, в отличие от людей, ограничены теоремой Геделя (Lucas 1961; Penrose 1989).

Внешние возражения наталкивались на трудности, связанные с успехами вычислительных симуляций физических процессов в целом. Кажется, в частности, что у нас есть серьезное основание верить в вычислимость законов физики, так что мы по крайней мере должны быть способны осуществить вычислительную *симуляцию* человеческого поведения. Иногда это оспаривается доводами о невычислимом компоненте физических законов (так поступает Пенроуз) или о нефизической причинности (как это делает Лукас), но очевидно, что сторонники подобных доводов ведут неравный бой.

Большее распространение получили возражения, называемые мной *внутренними*. Те, кто высказывает их, хотя бы в целях своей аргументации признают, что компьютеры могли бы симулировать человеческое поведение, но доказывают, что они все равно были бы лишены ментального. В частности, предполагается, что у них не было бы внутренней жизни: ни сознательного опыта, ни подлинного понимания. В лучшем случае компьютер мог бы *симулировать*, но не воспроизводить ментальность. Наиболее известным возражением этого типа является аргумент «Китайской комнаты» Джона Серла (Searle 1980). Согласно этим возражениям, вычислительные системы в лучшем случае были бы наделены пустыми оболочками ментального: они были бы кремниевыми разновидностями зомби.

Сторонники нередуктивного взгляда на сознательный опыт нередко симпатизировали внутренним возражениям против искусственного интеллекта, и многие доказывали, что простой компьютер не может обладать сознанием. Более того, те, на кого произвела впечатление проблема сознания, подчас характеризовали эту проблему, указывая на сознание как на такую черту, которая есть у нас, но отсутствовала бы у любого компьютера! Многим оказалось трудно поверить, что искусственная, небиологическая система могла бы быть тем, что могло бы порождать сознательный опыт.

Однако нередуктивный взгляд на сознание вовсе не автоматически приводит к пессимистическим взглядам на ИИ. Это совершенно разные вопросы. Первый имеет отношение к *прочности* сопряжения физических систем и сознания: конституируется ли сознание физическими процессами, или же оно всего лишь возникает из этих процессов? Второй вопрос касается *формы* этого сопряжения: *какие* именно физические системы порождают сознание? Разумеется, не очевидно, что выполнение надлежащего вычисления должно порождать сознание; но не очевидно и то, что нейронные процессы в мозге должны

порождать его. И, на первый взгляд, не ясно, почему компьютеры должны уступать мозгу в этом отношении. Признав удивительный факт порождения мозгом сознания, мы не испытаем нового удивления, обнаружив, что вычисление могло бы порождать сознание. Так что само по себе принятие нередуктивного взгляда на сознание должно было бы оставлять рассматриваемый нами вопрос открытым.

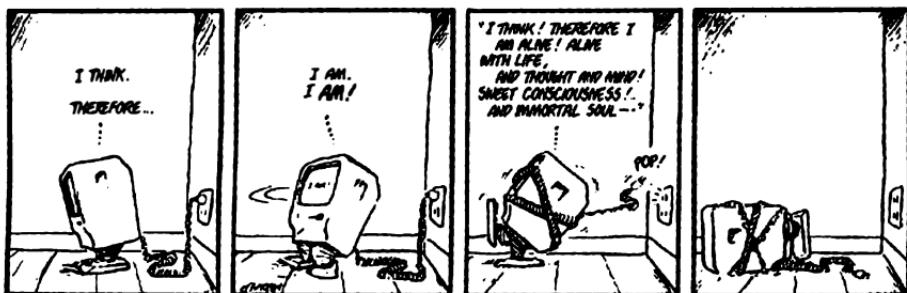


Рис. 9.1. Bloom County о сильном ИИ (© 1985, Washington Post Writers Group. Перепечатано с разрешения).

- (1) — Я мыслю. Следовательно...
- (2) — Я существую. Я существую!
- (3) — Я мыслю! Следовательно, я живой! Я живой, у меня есть жизнь, мысли и психика! Сладкое сознание!.. И бессмертная душа! — Хлоп!

В этой главе я пойду дальше и попытаюсь показать оправданность амбиций сторонников искусственного интеллекта (рис. 9.1). В частности, я буду отстаивать позицию, которую Серл называет *сильным искусственным интеллектом* и которая состоит в том, что существует непустой класс вычислений, такой, что осуществление любого вычисления этого класса достаточно для ментального и в том числе для существования сознательного опыта. Конечно, эта достаточность имеет силу лишь естественной необходимости: осуществление любого из этих вычислений при отсутствии сознания является *логически* возможным. Но мы видели, что это верно и для мозга. При оценке перспектив машинного сознания в актуальном мире нас интересует естественная возможность и необходимость.

(Чтобы этот вывод не казался тривиальностью, учитывая панпсихистские предположения, высказанные в предыдущей главе, замечу, что ничто в этой главе не зависит от тех соображений. В самом деле, я буду доказывать не только то, что осуществление надлежащего вычисления достаточно для сознания, но и то, что осуществление надлежащего вычисления достаточно для богатого сознательного опыта, подобного нашему.)

Я уже выполнил большую часть работы, которая требуется для указанной защиты сильного ИИ, обосновывая в главе 7 принцип организационной инвариантности. Если тот аргумент верен, то он устанавливает, что любая система с надлежащим типом функциональной организации наделена сознанием — неважно, из чего она сделана. Так что мы уже знаем, что, скажем, кремниевый состав не является преградой для обладания сознанием. Остается лишь прояснить связь между вычислением и функциональной организацией, чтобы установить, что осуществление надлежащего вычисления гарантирует наличие релевантной функциональной организации. Если это сделано, то отсюда вытекает сильный ИИ. Я также отвечу на ряд возражений, выдвинутых против проекта сильного ИИ.

2. Об имплементации вычисления

Стандартная теория вычисления имеет дело исключительно с *абстрактными* объектами: машинами Тьюринга, программами на Паскале, конечными автоматами и т. п. Это — математические сущности, населяющие математическое пространство. Когнитивные же системы, существующие в реальном мире, это *конкретные* объекты, имеющие физическое воплощение и каузально взаимодействующие с другими объектами физического мира. Но зачастую мы хотим использовать теорию вычисления для получения выводов о конкретных объектах в реальном мире. Для этого нам нужен мост между областью абстрактного и областью конкретного¹.

Таким мостом является понятие *имплементации* — отношения между абстрактными вычислительными объектами, или, коротко, «вычислениями» и физическими системами, имеющееся тогда, когда физическая система «реализует» вычисление, или когда вычисление «описывает» физическую систему. Вычисления часто имплементируются на синтетических, кремниевых компьютерах, но их можно имплементировать и иначе. К примеру, нередко говорят, что вычисления имплементируют природные системы, такие как человеческий мозг. Вычислительные описания применяются к самым разнообразным физическим системам. И всякий раз в подобных случаях неявно или явно задействуется понятие имплементации.

Понятие имплементации нечасто подвергается детальному анализу; обычно его просто принимают как данность. Но для защиты сильного ИИ мы нуждаемся в его детальном объяснении. Тезис о сильном ИИ выражен в вычислительных терминах и говорит о том, что имплементация надлежащего вычисления достаточна для сознания. Чтобы оценить это утверждение, мы должны знать, в каком именно

случае физическая система имплементирует вычисление. Разобравшись в этом, мы сможем присоединить это знание к проделанному нами ранее анализу психофизических законов, чтобы понять, может ли из этого вытекать интересующий нас вывод.

Некоторые авторы доказывали невозможность дать сколь-либо содержательное объяснение имплементации. В частности, Серл (Searle 1990b) доказывал, что имплементация не является чем-то объективным, а, наоборот, зависит от наблюдателя: любая система может быть истолкована как имплементирующая любое вычисление при надлежащей интерпретации. Серл, к примеру, считает, что о его стене можно рассуждать так, будто она имплементирует его текстовый редактор Wordstar. Если бы это было так, то трудно было бы понять, как вычислительные понятия могут играть какую-то основополагающую роль в теории, в конечном счете имеющей дело с конкретными системами. Что же до сильного ИИ, то он либо оказался бы лишен всякого содержания, либо подразумевал бы признание сильной разновидности панпсихизма. Думаю, однако, что пессимизм такого рода неоснователен: не существует препятствий для объективной характеристики имплементации. В этом параграфе я схематично представлю эту характеристику (она будет несколько технической, но оставшаяся часть главы должна быть понятной, даже если прокопчить эти детали).

Любое объяснение имплементации вычисления будет зависеть от того, о каком классе вычислений идет речь. Существует множество разных вычислительных формализмов, которым соответствуют различные классы вычислений: машины Тьюринга, конечные автоматы, программы на Паскале, коннекционистские сети, клеточные автоматы и т. д. В принципе, нам нужно объяснение имплементации для каждого из этих формализмов. Я объясню имплементацию только для одного формализма, а именно для формализма *комбинаторных автоматов*. Этот класс вычислений достаточно абстрактен для того, чтобы можно было без труда перенести на другие классы связанную с ним концепцию имплементации.

Комбинаторный автомат — более рафинированный родственник конечного автомата. Конечный автомат (КА) специфицируется заданием конечного набора данных на входе, конечного набора внутренних состояний, конечного набора данных на выходе и указанием сопряженного с ними набора отношений *перехода от состоянию к состоянию*. Внутреннее состояние КА — это простой элемент S_1 , лишенный внутренней структуры; это же справедливо для данных на входе и на выходе. Отношения перехода специфицируют для каждой возможной пары данных на входе и внутренних состояний новое внутреннее состояние и имеющийся на выходе результат. Если дано начальное состояние

какого-то КА, то эти отношения перехода от состояния к состоянию специфицируют, каким образом он будет изменяться во времени и что он будет давать на выходе в зависимости от того, что он получает на входе. Вычислительная структура КА представляет собой этот относительно простой набор отношений перехода от состояния к состоянию в наборе неструктурированных состояний.

Конечные автоматы непригодны для репрезентации структуры большинства вычислений, имеющих практическое значение, так как состояния и отношения перехода от состояния к состоянию в них обычно имеют сложную внутреннюю структуру. Никакое описание КА не может, к примеру, передать всю структуру программы на Паскале, машины Тьюринга или клеточного автомата. Полезней поэтому сконцентрироваться на классе автоматов со структуризованными внутренними состояниями.

Комбинаторные автоматы (КоА) ничем не отличаются от КА, за исключением того, что их внутренние состояния структурированы. Состояние комбинаторного автомата представляет собой вектор $[S^1, S^2, \dots, S^n]$. Этот вектор может быть конечным или бесконечным, но я сосредоточусь на первом случае. Элементы этого вектора можно мыслить в качестве *компонентов* внутреннего состояния; они соответствуют клеткам клеточного автомата или ячейкам ленты и состоянию управляющего устройства машины Тьюринга. Каждый элемент S^i может иметь конечное множество значений S_j^i , где S_j^i — j -е возможное значение i -го элемента. Эти значения можно трактовать в качестве «подсостояний» общего состояния. Данные на входе и на выходе имеют аналогичную комплексную структуру: первые являются вектором $[I^1, \dots, I^k]$, вторые — вектором $[O^1, \dots, O^m]$.

КоА детерминируется специфицированием набора векторов внутренних состояний и векторов данных на входе и на выходе и специфицированием набора *правил перехода от состояния к состоянию*, определяющих изменение состояния КоА во времени. Для каждого элемента вектора внутреннего состояния правило перехода от состояния к состоянию определяет то, как его новое значение зависит от прежних значений данных на входе и векторов внутреннего состояния. Для каждого элемента исходящего вектора правило перехода от состояния к состоянию определяет то, как его новое значение зависит от прежних значений вектора внутреннего состояния. Каждый конечный КоА может быть представлен как КА с равными вычислительными возможностями, но КА-описание приносило бы в жертву большинство структурных моментов, существенных для КоА. Эта структура имеет ключевое значение при использовании КоА для уяснения той организации, которая лежит в основе ментального.

Теперь мы уже можем дать объяснение имплементации. Вычисления, такие как КоA — это абстрактные объекты, *формальная структура* которых определяется их состояниями и отношениями перехода от состояния к состоянию. Физические системы — это конкретные объекты, *каузальная структура* которых определяется их внутренними состояниями и каузальными отношениями между ними. В неформальном плане мы говорим, что физическая система *имплементирует* вычисление, если каузальная структура этой системы отражает формальную структуру вычисления. То есть, система имплементирует вычисление, если можно установить такое соответствие между состояниями этой системы и вычислительными состояниями, при котором каузально соотнесенные физические состояния соответствуют формально соотнесенным формальным состояниям.

Эта интуитивная идея может быть напрямую использована для получения представления об имплементации в случае КоA. Физическая система имплементирует КоA, если можно так разложить внутренние состояния системы на подсостояния, и данные системы на входе и на выходе — на подсостояния этих данных, и так соотнести подсостояния системы с подсостояниями КоA, что отношения каузального перехода от состояния к состоянию для физических состояний, данных на входе и на выходе отражают формальные отношения перехода от состояния к состоянию для соответствующих формальных состояний, данных на входе и на выходе.

Формальный критерий имплементации КоA выглядит таким образом:

Физическая система P имплементирует КоA M , если можно разложить внутренние состояния P на компоненты $[s^1, \dots, s^n]$ и установить соответствие f между подсостояниями s^j и соответствующими подсостояниями S^j в M , а также произвести сходную операцию разложения и установления соответствия для данных на входе и на выходе, так, что для каждого правила перехода от состоянию к состоянию

$$([I^1, \dots, I^k], [S^1, \dots, S^n]) \rightarrow ([S'^1, \dots, S'^n], [O^1, \dots, O^l])$$

в M : если P находится во внутреннем состоянии $[s^1, \dots, s^n]$ и получает на входе $[i^1, \dots, i^n]$, соответствующие формальному состоянию и данным на входе $[S^1, \dots, S^n]$ и $[I^1, \dots, I^k]$, то это каузальным образом с неизбежностью приводит к тому, что она переходит к такому внутреннему состоянию и порождает такие данные на выходе, которые надлежащим образом соответствуют $[S'^1, \dots, S'^n]$ и $[O^1, \dots, O^l]$.

Мы можем допустить, что при разложении состояния физической системы на вектор подсостояний значение каждого элемента этого вектора должно быть супервентно на отдельном участке физической системы, чтобы гарантировать, что каузальная организация соотносит

различные компоненты этой системы. В ином случае не очевидно, что детальная каузальная структура действительно присутствует в физической системе. Эти и другие детали приведенной выше дефиниции можно уточнять и дальше. Понятие имплементации не выбито на скрижалях, и оно допускает как расширение, так и сужение — в зависимости от поставленных целей. Но здесь даны главные его контуры, общие для всех концепций имплементации.

Могло бы показаться, что КоA немногим лучше КА. В конце концов, для любого конечного КоA мы можем найти соответствующий ему КА с тем же поведением на входе-выходе. Но между ними все же есть существенные различия. Первое и главное из них состоит в том, что условия *имплементации* для КоA гораздо более ограничены, чем аналогичные условия для КА. Имплементация КоA предполагает сложное каузальное взаимодействие множества отдельных частей; и поэтому КоA-описание может передавать каузальную организацию системы с гораздо большей степенью детализации. Во-вторых, КоA позволяют сформулировать единое объяснение условий имплементации, значимое как для конечных, так и для бесконечных машин. И, в-третьих, КоA может напрямую отражать комплексную формальную организацию вычислительных объектов, таких как машины Тьюринга и клеточные автоматы. В соответствующих КА была бы утрачена большая часть этих структурных моментов.

На деле мы можем использовать эту дефиницию имплементации для прямого получения критериев имплементации для других видов вычислений. К примеру, для конкретизации условий имплементации машины Тьюринга нам нужно просто переописать машину Тьюринга как КоA и применить данную выше дефиницию. Для этого мы описываем состояние машины Тьюринга как громадный вектор. Один из элементов этого вектора репрезентирует состояние ее управляющего устройства, есть элементы и для каждой ячейки ее ленты, репрезентирующие символ в этой ячейке и указывающие, находится ли управляющее устройство на ней. Правила перехода от состояния к состоянию для векторов естественным образом извлекаются из механизмов, специфицирующих поведение управляющего устройства машины и ленты. Разумеется, векторы здесь бесконечны, но условия имплементации для случая с бесконечностью являются непосредственным расширением условий для конечных случаев. При переводе с языка формализма машины Тьюринга на язык формализма КоA мы сможем сказать, что машина Тьюринга имплементируется всякий раз, когда имплементируется соответствующий КоA. Аналогичные переводы можно сделать и для вычислений в других формализмах, таких как клеточные автоматы или программы на Паскале, что позволит сформулировать условия имплементации для вычислений в каждом из этих классов.

Это дает совершенно объективный критерий для имплементации вычисления. Имплементация вычисления не выхолащивается, как полагал Серл. Конечно, некоторые вычисления будут имплементироваться любой системой. К примеру, одноэлементный КоА с одним состоянием будет имплементироваться любой системой, почти так же широко будет имплементироваться и двухэлементный КоА. Верно также, что большинство систем будет имплементировать более чем одно вычисление — в зависимости от того, как мы будем очерчивать состояния этой системы. В этом нет ничего удивительного: вполне ожидаемым представляется то, что моя рабочая станция, как и мой мозг, имплементирует множество вычислений.

Существенно, однако, то, что нет оснований считать, что *любой* КоА будет имплементирован *любой* системой. Если взять любой сложный КоА, то окажется, что существует лишь очень немного физических систем, наделенных каузальной организацией, необходимой для его имплементации. Если мы возьмем КоА, векторы состояний которого имеют тысячу элементов, с десятью опциями для каждого элемента, то аргументы, подобные тем, что выдвигались в главе 7, покажут, что шанс случайного набора физических состояний, имеющих нужные каузальные отношения, чуть меньше 1 из $(10^{1000})^{10^{1000}}$ (на деле гораздо меньше из-за требования прочности отношений перехода от состояния к состоянию²⁾.

Так как же быть с утверждением Серла о том, что вычислительные описания зависимы от наблюдателя? Верно то, что здесь имеется определенная степень такой зависимости: любая физическая система будет имплементировать множество вычислений, и то, на каком из них сосредоточится наблюдатель, зависит от целей этого наблюдателя. Но это не несет в себе угрозу для ИИ или для вычислительной когнитивной науки. По-прежнему верно, что по отношению к любому вычислению можно говорить о фактичности того, что та или иная система имплементирует или не имплементирует его, и в качестве его имплементаций будет выступать лишь ограниченный класс систем. И этого достаточно для того чтобы вычислительные концепции имели метафизическое и объяснительное значение.

Утверждение о том, что физическая система имплементирует комплексное вычисление P , равносильно совершенно нетривиальному утверждению о каузальной структуре этой системы, которое может быть весьма полезным для когнитивных объяснений, а, быть может, и для понимания основы сознания. Лишь системы с очень специфичной разновидностью каузальной организации могут рассчитывать на соответствие сильным ограничительным условиям имплементации. Так что здесь нет опасности выхолащивания, и можно надеяться, что понятие вычисления окажется прочной основой для анализа когнитивных систем.

3. В защиту сильного ИИ

Имплементация КоА поразительно напоминает реализацию функциональной организации. Вспомним, что функциональная организация определяется специфицированием множества абстрактных компонентов, множества состояний каждого компонента и системы отношений зависимости, указующих на то, как состояния каждого компонента зависят от предыдущих состояний и от данных на входе, а также на то, как данные на выходе зависят от предшествующих состояний. Понятие КоА, по сути дела, является непосредственной формализацией указанного понятия.

Действительно, при наличии любой функциональной организации того типа, который был описан в главе 7, из нее можно напрямую извлечь КоА. Нужно лишь допустить, что векторы состояния КоА располагают элементами для каждого компонента этой организации и что формальные переходы состояний КоА соответствуют отношениям каузальной зависимости между компонентами. Реализация функциональной организации практически не отличается от имплементации соответствующего КоА. Небольшие различия, связанные, в частности, с различным обращением с данными на входе и выходе, есть, но они несущественны.

Предложенная мной концепция имплементации, таким образом, проясняет связь между каузальной и вычислительной организацией. Так мы можем увидеть, что при применении вычислительных описаний к физическим системам они, по сути, предоставляют нам формальное описание каузальной организации системы. Вычислительный язык оптимально подходит для спецификации подобной абстрактной каузальной организации. Можно даже попробовать показать, что именно поэтому вычислительные понятия получили столь широкое распространение в когнитивной науке. В объяснении поведения сложной когнитивной системы наибольшее значение имеет абстрактная каузальная организация системы, и вычислительные формализмы оказываются идеальным каркасом, с помощью которого может быть описана и проанализирована подобная организация³.

Указанная связь упрощает защиту сильного искусственного интеллекта. Я уже приводил доводы в пользу принципа организационной инвариантности, согласно которому относительно любой системы, наделенной сознательными переживаниями, верно, что система, обладающая такой же высокодетализированной функциональной организацией, будет обладать качественно идентичными сознательными переживаниями. Мы, однако, знаем, что любая функциональная организация может быть абстрагирована в КоА, всегда имплементируемый

при реализации этой организации. Из этого следует, что для данной сознательной системы M ее высокодетализированная функциональная организация может быть абстрагирована в КоA M , такой, что любая система, имплементирующая M , будет реализовывать такую же функциональную организацию и поэтому будет обладать сознательными переживаниями, качественно неотличимыми от переживаний исходной системы. А это и есть сильный искусственный интеллект.

К примеру, на основе нейронного описания мозга мы могли бы получить абстрактный КоA, располагающий элементами вектора состояния для каждого нейрона и подсостояниями для каждого элемента, отражающими соответствующий диапазон состояний для каждого из нейронов. Правила перехода от состояния к состоянию в КоA отражают тип зависимости состояния каждого нейрона от состояния других нейронов и соотнесенность нейронных состояний и данных на входе и на выходе. Если значимы не только нейронные компоненты мозга, то можно включить и их. Любая физическая система, имплементирующая этот КоA, будет иметь такую высокодетализированную функциональную организацию, которая будет дублировать функциональную организацию мозга на нейронном уровне. Согласно принципу инвариантности, эта система будет обладать переживаниями, неотличимыми от тех, которые связаны с мозгом.

Заманчиво представлять компьютер просто как устройство с входом и выходом, между которыми есть место лишь для формальных математических манипуляций. Подобный взгляд на вещи, однако, игнорирует тот ключевой факт, что внутри компьютера — так же как и внутри мозга — имеется богатая каузальная динамика. В самом деле, в обычном компьютере, понейронно имплементирующем симуляцию моего мозга, есть место для реальной каузальности напряжений различных цепей, в точности отражающей паттерны каузальности между нейронами. Каждому нейрону будет соответствовать участок памяти, репрезентирующий этот нейрон, и каждый из этих участков будет физически реализован напряжением в каком-то физическом месте. За возникающий сознательный опыт отвечают именно эти каузальные паттерны электрических цепей, так же как за его возникновение отвечают каузальные паттерны нейронов мозга.

Тезис сильного искусственного интеллекта можно защищать и напрямую, используя аргументы от блекнувших и от скачущих квалиа. При наличии любых двух имплементаций КоA между ними будет спектр случаев, в которых физические компоненты имплементации заменяются по одному при сохранении паттерна их каузального взаимодействия с остальной частью системы. Если одна из систем обладает сознанием, и если КоA является абстрактным выражением ее высокодетализированной функциональной организации, то из упомянутых

аргументов следует, что вторая система должна обладать сознанием и что она должна быть наделена неотличимыми сознательными переживаниями. Если бы вторая система не обладала сознанием, то существовала бы промежуточная система с блекнувшими квалиа. Если бы вторая система не обладала сознанием или имела другие осознанные переживания, то мы смогли бы сконструировать промежуточную систему со скачущими квалиа. Такие последствия неправдоподобны по причинам, указанным в главе 7. Но если принять, что квалиа не могут блекнуть или скакать подобным образом, то из этого следует, что вторая из изначально имевшихся у нас систем обладает переживаниями, неотличимыми от переживаний первой системы, что и означает признание истинности тезиса о сильном искусственном интеллекте.

Здесь есть небольшая тонкость. Аргумент предполагает, что организация мозга может быть абстрагирована в КоA-описание. Для этого нужно лишь, чтобы релевантная организация могла быть описана в терминах конечного множества частей, каждая из которых имеет конечное множество релевантных состояний. Однако это можно было бы оспорить. Не исключено, к примеру, что для каждого нейрона требуется бесконечное множество состояний, чтобы передать непрерывность соответствующих процессов, имеющую важное значение. И можно было бы утверждать, что переходы между этими бесконечными состояниями могли бы оказаться невычислимыми. Позже я еще рассмотрю подобное возражение; пока же я готов принять вывод, что если когнитивная динамика вычислима, то надлежащая вычислительная организация будет порождать сознание. Иначе говоря, тут меня больше интересуют внутренние, а не внешние возражения. Тем не менее позже в этой главе я отвечу на ряд внешних возражений.

4. Китайская комната и другие возражения

Противники сильного ИИ, конечно же, подчас выдвигали конкретные доводы против этой позиции. Наиболее известным из них мы обязаны Джону Серлу, излагающему их, в частности, в статье 1980 г. «Сознания, мозги и программы». Я отвечу на них в рамках очерченной модели.

Китайская комната

В ходе знаменитого аргумента против сильного ИИ, Серл (Searle 1980) доказывает, что любая программа может быть имплементирована без порождения ментального. Доказательство построено на демонстрации того, что он считает *контрпримером* для тезиса о сильном ИИ: Китайской комнаты, внутри которой субъект, манипулирующий символами, симулирует человека, понимающего китайский язык. Китайская комната задумана как

иллюстрация системы, имплементирующей программу — какой бы они ни была — при отсутствии релевантного сознательного опыта.

В изначальной версии Серл направляет свой аргумент скорее против машинной *интенциональности*, чем против машинного сознания, доказывая, что у Китайской комнаты отсутствует «понимание». Ясно тем не менее, что, по сути, речь идет о сознании. Ядро этого аргумента, если он верен, напрямую устанавливает, что система Китайской комнаты лишена сознательных состояний, таких как сознательное переживание понимания китайского языка. Серл считает, что интенциональность предполагает сознание, так что этого достаточно для того, чтобы убедиться, что комната лишена и интенциональности. Другие, впрочем, отрицают это. В любом случае, мы можем отвлечься от вопроса о связи между сознанием и интенциональностью и сформулировать проблему исключительно в терминах сознания. При этом можно чуть лучше высветить спорные моменты.

(Иначе говоря, мы можем разделить выводы Серла на две части: (1) программы не достаточно для сознания; и (2) программы не достаточно для интенциональности. Серл полагает, что из (1) следует (2), но другие отрицают это. Яснее всего будет, если сфокусировать аргумент о сильном ИИ на (1): все стороны соглашаются, что, если (1) верен, то наиболее интересная разновидность сильного ИИ обречена, и даже Серл согласился бы, что опровержение (1) показало бы, что аргумент Китайской комнаты не проходит. Связь между сознанием и интенциональностью можно выделить тогда в отдельный вопрос, не критичный для аргумента против ИИ.)

Действуя подобным образом, мы избегаем ситуации, при которой оппоненты выступают против (2), не беспокоясь об опровержении (1). К примеру, возражения, сфокусированные на связи между Китайской комнатой и ее окружением (Fodor 1980; Rey 1986), и возражения, процедурно или функционально объясняющие интенциональность (Boden 1988; Thagard 1986), могут в лучшем случае прояснить вопрос об интенциональности, но они никак не повышают правдоподобность вывода о сознательности Китайской комнаты. В результате они оставляют такое чувство, будто проблема, которую ставит перед ИИ этот сценарий, так и не была рассмотрена. Максимум, что они делают, так это оспаривают вспомогательную посылку о том, что интенциональность предполагает сознание⁴.)

Аргумент Китайской комнаты разворачивается следующим образом. Возьмем любую программу, которая, как предполагается, передает какой-то из аспектов сознания, к примеру, понимание китайского языка или наличие ощущения красного. Далее эта программа может быть имплементирована англоязычным субъектом, не знающим других языков, — будем называть его *демоном* — в черно-белой комнате. Демон

вручную следует всем правилам, специфицированным программой, ведя записи о всех релевантных внутренних состояниях и переменных на клочках бумаги, при необходимости стирая и обновляя их. Мы можем также вообразить, что демон соединен с роботизированным телом, получая цифровые данные от перцептивных датчиков, манипулируя ими согласно программным спецификациям и посылая результирующие данные на моторные эффекторы. Здесь можно говорить о совершенной имплементации программы. Тем не менее кажется очевидным, что у демона нет сознательного понимания китайского и что он лишен ощущения красного. Поэтому имплементация программы не достаточна для этих сознательных переживаний. Сознание предполагает не только имплементацию релевантной программы.

Зашитники сильного ИИ обычно возражали так, что они соглашались, что демон не понимает по-китайски, но доказывали, что понимание и сознание должны быть атрибутированы *системе*, включающей в свой состав демона и клочки бумаги. Серл заявлял об очевиднейшей неубедительности такого возражения. Действительно, в утверждении о том, что система, состоящая из субъекта и клочков бумаги, обладает коллективным сознанием, есть нечто континтуитивное. На этой стадии аргумент заходит в тупик. Защитники ИИ доказывают, что система обладает сознанием, оппоненты находят этот вывод нелепым, и дальнейшее продвижение кажется затруднительным. Думаю, однако, что приведенные ранее доводы дают возможность выйти из тупика к сильному ИИ.

Допустим, что программа, о которой мы говорим, в действительности является комбинаторным автоматом, отражающим на нейронном уровне организацию носителя китайского языка, смотрящего на сочное красное яблоко. Демон в комнате имплементирует этот КоА, располагая клочками бумаги для каждого элемента вектора состояния и обновляя их на каждом временном шагу в соответствии с правилами перехода от состояния к состоянию. Мы можем развернуть аргументы блекнувших и скачущих квалиа, сконструировав спектр случаев, занимающих промежуточное положение между исходным носителем китайского языка и Китайской комнатой⁵. Сделать это нетрудно. Вначале мы можем вообразить, что нейроны в голове носителя китайского языка последовательно заменяются крошечными демонами, каждый из которых дуплицирует входящую-исходящую функцию нейрона⁶. При получении стимулов со стороны соседних нейронов демон производит соответствующие вычисления и, в свою очередь, стимулирует соседние нейроны. По мере замены все большего количества нейронов демоны берут верх, пока весь череп не оказывается заполнен миллиардами демонов, реагирующих на сигналы друг друга и на входящие чувственные

данные, производящих вычисления, сигнализирующих другим демонам и, в свою очередь, стимулирующих моторные окончания. (Если кто-то будет возражать, сказав, что все эти демоны никогда не смогут поместиться в череп, то взамен можно вообразить сценарий с устройствами радиопередачи, находящимися вне черепа.)

Далее, мы постепенно уменьшаем число демонов, удваивая их работу. Сперва мы заменяем два соседних демона одним демоном, делающим их работу. Новый демон будет фиксировать внутреннее состояние обоих симулируемых им нейронов — можно представить, что эта запись будет храниться на кусочке бумаги в каждой из локаций. Каждый из этих кусочков бумаги будет обновляться в зависимости от сигналов, поступающих от соседних нейронов и от состояния другого кусочка бумаги. Консолидация демонов будет продолжена, пока в конце концов на их месте не окажется один-единственный демон и миллиарды крошечных кусочков бумаги. Можно представить, будто каждый из этих кусочков находится в месте изначального расположения соответствующего ему нейрона, а демон мечется по мозгу, обновляя каждый из этих кусочков в качестве функции состояний соседних кусочков и, при необходимости, чувственных данных.

Несмотря на все эти изменения, итоговая система имеет ту же функциональную организацию, что и исходный мозг. Каузальные отношения между нейронами в исходном случае отражаются в каузальных отношениях между демонами в промежуточном случае и между кусочками бумаги — в конечном случае. В этом конечном случае каузальные отношения опосредуются действиями демона — кусочек бумаги воздействует на состояние демона, который воздействует на соседний кусочек бумаги — но они тем не менее являются каузальными отношениями. Если мы посмотрим на функционирование этой системы в ускоренном режиме, то мы увидим вихрь каузальных взаимодействий, точно соответствующий нейронному вихрю.

Мы, таким образом, сможем применить аргументы от блекнувших и скачущих квалиа. Если итоговая система лишена сознательного опыта, то должна существовать промежуточная система с блеклыми сознательными переживаниями. А это неправдоподобно по уже известным нам причинам. Мы можем также вообразить, что мы поочередно подключаем то нейронную цепь, то соответствующую ей резервную цепь, имплементированную демонами или одним демоном и кусочками бумаги. Как и прежде, это привело бы к скачущим квалиа при неизменной функциональной организации, так что система не смогла бы заметить разницу. И опять-таки гораздо более правдоподобным выглядит допущение, что квалиа не будут меняться.

Оправданным, стало быть, кажется заключение, что итоговая система имеет точно такие же сознательные переживания, что и исходная

система. Если нейронная система порождала переживания чего-то ярко-красного, то их будет порождать и система демонов, равно как и сеть клочков бумаги, опосредованных демоном. Но этот итоговый случай, разумеется, всего лишь копия системы, находящейся в Китайской комнате. Мы тем самым нашли позитивное основание считать, что та система действительно обладает сознательными переживаниями, в частности, теми, которые связаны с пониманием китайского языка и переживанием красного.

Подобный взгляд на вещи проясняет два момента, которые могли бы оставаться затемненными серловским описанием Китайской комнаты. Во-первых, «клочки бумаги» в комнате — это не просто кучи формальных символов. Они образуют конкретную динамическую систему, каузальная организация которой прямо соответствует организации исходного мозга. Это обстоятельство затемняется медлительностью связываемой с ними манипуляции символами, а также присутствием демона, осуществляющего эту манипуляцию; и тем не менее сознательный опыт порождает именно конкретная динамика этих клочков бумаги. Во-вторых, демон играет сугубо второстепенную роль. Ключевая каузальная динамика связана с клочками бумаги, соответствующими нейронам в исходном случае. Демон всего лишь оказывает вспоможение этим каузальным отношениям. Образ демона, мечущегося внутри черепной коробки, делает очевидным то обстоятельство, что атрибутирование переживаний системы *демону* было бы серьезным смешением уровней. Факт сознательности демона может искушать на предположение, что если переживания системы вообще где-то существуют, то они существуют именно в демоне; но в действительности сознательность демона не имеет никакого отношения к функционированию системы. Функции демона могла бы выполнить простая таблица соответствий. Ключевым аспектом системы является динамика символов.

Аргумент Серла интуитивно впечатляет нас из-за причудливости программной имплементации, затемняющей реализацию релевантной каузальной динамики. Но при избавлении от образов, навеваемых присутствием нерелевантного демона и медлительностью перетасовки символов, мы увидим, что каузальная динамика в комнате в точности отражает каузальную динамику в черепной коробке. И тогда предположение о том, что эта система порождает опыт, уже не будет казаться столь неправдоподобным.

В одной из версий серловского аргумента демон запоминает правила вычисления и внутренним образом имплементирует программу. Конечно, на практике люди не могут запомнить даже сотню правил и символов, не говоря уже о многих миллиардах, но можно вообразить, что демон с модулем суперпамятии был бы в состоянии запомнить все пра-

вила и состояния всех символов. В этом случае опять-таки можно ожидать, что система породит сознательные переживания, которые не будут переживаниями демона. Серл доказывает, что если кто-то и должен иметь переживания, то им может быть только демон, так как все процессы обработки информации протекают внутри него, однако эта ситуация должна рассматриваться как пример реализации двух ментальных систем в одном и том же физическом пространстве. Устройство, порождающее китайские переживания, никак не совпадает с организацией, порождающей переживания демона. Понимающее китайский устройство состоит в каузальных отношениях миллиардов локаций модуля суперпамяти; демон же, повторю, действует лишь в качестве каузального вспоможения. Это становится очевидным при рассмотрении спектра случаев, когда демон, мечущийся в черепной коробке, постепенно запоминает правила и символы — до тех пор, пока не интернализирует все их. Соответствующая структура постепенно перемещается из черепной коробки в суперпамять демона, но опыт остается неизменным и совершенно обособленным от переживаний демона.

Можно было бы предположить, что поскольку мой аргумент предполагает дуплицирование организации мозга, он устанавливает лишь слабую разновидность сильного ИИ, тесно увязанную с биологией. (При обсуждении возражения, называемого им «Симулятором мозга», Серл выражает удивление тем обстоятельством, что сторонник ИИ выдвигает возражение, подразумевающее детальную симуляцию человеческой биологии.) Однако в таком случае мы недооценили бы силу этого аргумента. Программа симуляции мозга — это просто тонкий край нашего клина. Узнав, что *одна* программа может порождать ментальное даже при ее имплементации в духе Китайской комнаты, мы полностью нейтрализуем аргумент Серла, претендующий на универсальный характер: мы знаем, что демон и бумага в Китайской комнате действительно могут фундировать независимое сознание. И тогда мы открываем шлюзы для целого спектра программ, которые могли бы оказаться кандидатами на роль программ, порождающих сознательный опыт. Границы этого спектра — это открытый вопрос, но Китайская комната не является препятствием на этом пути.

Второй аргумент, выдвинутый Серлом (Searle 1984), звучит так:

Синтаксис и семантика

1. Компьютерная программа имеет синтаксический характер.
2. Синтаксис недостаточен для семантики.
3. Сознания семантичны.
4. Поэтому имплементации программы недостаточно для сознания.

Опять-таки, это выдвигается как аргумент об интенциональности, но он может рассматриваться и как аргумент о сознании. Серл считает, что главной разновидностью интенциональности является феноменологическая интенциональность, внутренне присущая сознанию.

Этот аргумент можно интерпретировать и критиковать разными способами, но главная его проблема состоит в том, что он не придает должного значения ключевой роли имплементации. *Программы* — это абстрактные вычислительные объекты, и они сугубо синтаксичны. Можно быть уверенным, что ни одна программа как таковая не может претендовать на наличие в ней ментального. Но *имплементации программ* — это конкретные системы с каузальной динамикой, и они не есть нечто сугубо синтаксичное. Имплементация имеет каузальный вес в реальном мире, и благодаря этому весу возникает сознание и интенциональность. Синтаксична программа; имплементация же имеет семантическое содержание.

Серл мог бы попробовать доказать, что и имплементации в определенном смысле синтаксичны — возможно, потому, что динамика имплементаций детерминируется формальными принципами. Однако «синтаксичность» имплементаций имеет смысл, никак не связанный с тем, в котором можно правдоподобно утверждать, что синтаксис недостаточен для семантики. Хотя тезис о том, что статичный набор абстрактных символов лишен внутренних семантических свойств, и может быть правдоподобным, гораздо менее очевидным кажется утверждение, что формально специфицированные каузальные процессы не могут фундировать ментальное.

Мы можем следующим образом спародировать этот аргумент:

1. Рецепты синтаксичны.
2. Синтаксис недостаточен для рассыпчатости.
3. Пирожные рассыпчаты.
4. Значит, имплементации рецепта недостаточно для пирожного.

В таком виде указанный дефект бросается в глаза. Аргумент не проводит различия между рецептами, являющимися синтаксическими объектами, и имплементациями рецептов, полноценными физическими системами в реальном мире. Вся работа, опять-таки, проделывается отношением имплементации, соотносящим области абстрактного и конкретного. Рецепт имплицитно специфицирует класс физических систем, которые могут рассматриваться как *имплементации* этого рецепта, и именно эти системы обладают такими характеристиками, как рассыпчатость. Аналогичным образом программа имплицитно специфицирует класс физических систем, которые могут рассматриваться в качестве ее имплементаций, и именно они порождают такие характеристики, как ментальное.

Одно из популярных возражений против искусственного интеллекта (напр., Searle 1980; Harnad 1989) состоит в том, что симуляция феномена не тождественна его воспроизведению. К примеру, при компьютерной симуляции пищеварения никакая пища в действительности не переваривается. Симуляция урагана — это не настоящий ураган; при его симуляции на компьютере никто не промокает. При симуляции тепла реальное тепло не выделяется. Так почему же мы должны ожидать появления настоящего сознания при его симуляции? Почему мы должны ожидать, что именно в этом случае вычислительный процесс не просто симуляция, но и нечто реальное?

Симуляция — это всего лишь симуляция

Нет сомнений, что для многих свойств симуляция не является воспроизведением. Симулированное тепло — это не реальное тепло. Имеются, однако, и такие свойства, для которых симуляция *есть* воспроизведение. К примеру, симуляция системы с каузальной петлей *есть* система с каузальной петлей. Так что реальный вопрос состоит здесь в том, как отличить типы *X*, при которых симуляция *X* действительно есть *X*, от иных?

Ответ, как я полагаю, может быть таким: симуляция *X* оказывается *X* именно в тех случаях, когда свойство быть *X* является *организационным инвариантом*. Дефиниция организационного инварианта — такая же, как и раньше: свойство есть организационный инвариант, когда оно зависит только от функциональной организации фундирующей системы, а не от каких бы то ни было иных деталей. Вычислительная симуляция физической системы может ухватывать ее абстрактную каузальную организацию и гарантировать ее воспроизводимость при любой имплементации, вне зависимости от состава имплементирующей инстанции. Подобная имплементация будет в таком случае *воспроизводить* любые организационные инварианты изначальной системы, однако другие свойства будут утрачены.

Свойство быть ураганом не является организационным инвариантом, так как оно отчасти зависит от таких неорганизационных свойств, как скорость, форма и физических состав фундирующей системы (система с теми же каузальными взаимодействиями, очень медленно имплементированная на большом количестве бильярдных шаров, не была бы ураганом). Пищеварение и тепло сходным образом зависят от различных аспектов фундирующих физических субстратов, не являющихся всецело организационными. Мы могли бы постепенно заменять биологические компоненты пищеварительной системы — так, чтобы реакции, основанные на действии кислоты, заменялись каузально изоморфными взаимодействиями кусочков металла, и это уже не было бы пищеварением. Так что нам не стоит ожидать, что симуляция систем с этими свойствами сама будет обладать ими.

Но феноменальные свойства не таковы. Как я доказывал в главе 7, эти свойства представляют собой организационные инварианты. И если это так, то отсюда следует, что надлежащая симуляция системы с феноменальными свойствами сама будет обладать ими — благодаря воспроизведению высокодетализированной функциональной организации изначальной системы. Организационная инвариантность принципиально отличает сознание от других упомянутых свойств и открывает путь сильному ИИ.

5. Внешние возражения

По большей части меня интересовали внутренние возражения против сильного искусственного интеллекта, так как они наиболее значимы в контексте этой книги, однако я хотя бы упомяну некоторые внешние возражения. Я уже отмечал, что позиции противников внешних возражений против искусственного интеллекта изначально кажутся сильными: есть все основания считать, что законы физики, по крайней мере в их нынешнем понимании, вычислимы, и что человеческое поведение определяется физическими законами. Если так, то из этого следует, что вычислительная система может симулировать человеческое поведение. Тем не менее возражения время от времени выдвигаются, так что я кратко обсуджу их.

Возражения от следования правилам

Возможно, самое давнее внешнее возражение против ИИ состоит в том, что вычислительные системы всегда следуют правилам и поэтому неизбежно будут лишены ряда человеческих способностей, вроде креативности или гибкости. Во многих отношениях это самое слабое из внешних возражений, в частности из-за его явной нечеткости и неконкретности. В самом деле, на него можно легко ответить, сказав, в свою очередь, что на нейронном уровне человеческий мозг может быть вполне механичным и рефлекторным, но это никоим образом не препятствует креативности и гибкости на макроскопическом уровне. Конечно, оппонент опять-таки всегда может не согласиться с утверждением о механичности нейронного уровня, но в любом случае не видно хорошего аргумента в пользу тезиса о том, что вычислительная динамика на базовом каузальном уровне несовместима с креативностью и гибкостью на макроскопическом уровне.

Подобное возражение может подкрепляться неявным отождествлением вычислительных систем с *символьными* вычислительными системами: системами, производящими символные манипуляции с высокоуровневыми концептуальными репрезентациями — в предельном случае, с системами, жестко выводящими заключения из посылок

логики предикатов. Не исключено, что в этой области указанное возражение не лишено оснований, хотя даже это не очевидно. Но в любом случае класс вычислительных систем гораздо шире. К примеру, низкоуровневая симуляция мозга представляет собой некое вычисление, но не символьное вычисление того рода. Если говорить о промежуточном уровне, то к несимвольным вычислениям обращались коннекционистские модели в когнитивной науке. В этих случаях на каком-то уровне система может следовать правилам, но это напрямую не отражается на поведенческом уровне; и в самом деле, коннекционисты часто говорят, что их метод позволяет получить гибкость на высоком уровне из низкоуровневой механистичности. Как выразился Хоффштадтер (Hofstadter 1979), уровень, на котором я мыслю, не обязательно совпадает с уровнем, на котором я существую⁷.

Иногда утверждается, что теорема Геделя показывает, что вычислительным системам свойственны ограничения, которых нет у людей. Теорема Геделя говорит нам, что в любой непротиворечивой формальной системе, достаточно богатой для произведения арифметических операций, будет существовать некое истинное предложение — *Геделевское предложение* системы — которое эта система не сможет доказать. И аргумент состоит в том, что поскольку мы, однако же, можем понять, что оно истинно, мы обладаем некоей способностью, отсутствующей у этой формальной системы. Из этого следует, что никакая формальная система не может в точности передавать человеческие способности. (Подобные аргументы выдвигали среди прочих Лукас (Lucas 1961) и Пенроуз (Penrose 1989, 1994).)

Возражения от теоремы Геделя

Краткий ответ на эти аргументы состоит в том, что нет оснований считать, что и люди могут знать об истинности соответствующих Геделевских предложений. В лучшем случае мы можем знать, что если система непротиворечива, то ее Геделевское предложение истинно, но нет оснований полагать, что мы можем установить непротиворечивость произвольных формальных систем⁸. В особенности это справедливо в случае сложных формальных систем, таких как система, симулирующая реакции человеческого мозга: задача определения непротиворечивости подобной системы вполне может выходить за пределы наших возможностей. Так что вполне может оказаться так, что каждый из нас может симулироваться сложной формальной системой F , такой, что мы не в состоянии установить, является ли она непротиворечивой. И если это так, то мы не сможем узнать, будут ли истинными наши собственные Геделевские предложения.

Существует множество вариаций этого геделевского аргумента, с реакциями оппонентов на это предположение и ответными репликами,

нацеленными на то, чтобы обойти эти возражения. Здесь я не буду обсуждать их (хотя я подробно обсуждаю их в Chalmers 1995c). Эти вопросы связаны со множеством интересных и стимулятивных моментов, но, думаю, мы вправе сказать, что тезис о том, что геделевские ограничения не применимы к людям, никогда не был убедительным образом обоснован.

Возражения от невычислимости и континуальности

Приведенные выше возражения являются собой «высокоуровневые» аргументы о невычислимости когнитивных процессов. Но можно было бы попробовать атаковать позиции ИИ

и на низком уровне, доказывая невычислимость физических процессов. К примеру, Пенроуз (Penrose 1994) доказывает, что в адекватной теории квантовой гравитации мог бы быть невычислимый элемент. Единственным основанием для такого вывода, однако, оказывается у него вышеупомянутый геделевский аргумент. В самой физической теории нет ничего, что фундировало бы этот вывод; так что если отбросить геделевский аргумент, то исчезает основание верить в невычислимые физические законы. В самом деле, можно было бы попробовать показать, что если каждый элемент мозга, такой как нейрон, имеет лишь конечное множество релевантных состояний, и если существует лишь конечное множество релевантных элементов, то релевантная каузальная структура мозга должна выражаться вычислительным описанием.

Это ведет нас к последнему возражению, которое состоит в том, что процессы в мозге могут быть существенным образом *непрерывными*, тогда как вычислительные процессы дискретны, и что эта континуальность может быть существенной чертой нашей когнитивной компетентности, так что никакая дискретная симуляция не смогла бы воспроизвести эту компетентность. Быть может, создавая приблизительную копию нейрона с помощью элемента, имеющего лишь конечное множество состояний, мы утрачиваем нечто жизненно важное в плане реализации его функций. Оппонент может ссылаться, к примеру, на «чувствительную зависимость от изначальных условий» в определенных нелинейных системах, означающую, что даже небольшая округляющая ошибка на одной из стадий процесса может вести к масштабным макроскопическим различиям на более поздней стадии. Если процессы в мозге именно таковы, то любая дискретная симуляция мозга будет приводить к результатам, отличающимся от тех, которые получаются в континуальной реальности.

Имеется, однако, серьезное основание полагать, что абсолютная континуальность не может быть существенной характеристикой нашей когнитивной компетентности. Наличие фонового шума в биологических системах означает, что никакой процесс не может зависеть

от требования такого уровня точности, который выходит за определенные пределы. За пределами, скажем, 10^{-10} по соответствующей шкале неконтролируемые флуктуации фонового шума будут препятствовать дальнейшему уточнению. Иначе говоря, если мы создаем приблизительную копию состояния системы с таким уровнем точности (быть может, для надежности, еще более продвинутую — к примеру, на уровне 10^{-20}), то ее работа будет приносить те же результаты, которые реально могли бы быть и у той системы. Конечно, вследствие нелинейных эффектов эта приблизительная копия может продуцировать поведение, отличное от поведения, продуцируемого той системой по данному поводу, — но она продуцировала бы поведение, которое *могла бы* продуцировать и та система при несколько ином биологическом шуме. При желании мы можем даже приблизительно смоделировать сам процесс шума⁹. В результате симулирующая система будет обладать такими же поведенческими *способностями*, что и изначальная система, даже если она и продуцирует иное конкретное поведение в конкретных случаях. Мораль такова, что, когда речь идет о дуплировании наших когнитивных способностей, близкое сходство не хуже тождества.

Верно то, что система с безграничной степенью точности могла бы обладать когнитивными способностями, которые никогда не смогла бы получить в свое распоряжение какая-либо дискретная система. Можно было бы, к примеру, закодировать аналоговую величину, соответствующую реальному числу, n -ое бинарное значение которого равно 1, если и только если n -ая машина Тьюринга останавливается при любых данных на входе. Используя эту величину, совершенная континуальная система могла бы решить проблему остановки, с которой не может справиться никакая дискретная система. Наличие шума, однако, означает, что никакие биологические процессы не смогли бы надежно имплементировать эту систему. Биологические системы предполагают лишь лимитированную точность, и поэтому человеческие и животные мозги должны ограничиваться такими способностями, которые могут быть у дискретных систем.

6. Заключение

Вывод таков, что, похоже, не существует принципиальных преград, которые могли бы сдержать амбиции искусственного интеллекта. Внешние возражения не выглядят очень уж сильными. Внутренние возражения могли бы доставлять большее беспокойство, но анализ аргументов, подкрепляющих эти возражения, показывает, что они не являются убедительными; более того, если аргументы, которые я приводил в предыдущих главах, верны, то у нас имеется серьезное позитивное основание

считать, что имплементация надлежащего вычисления повлечет за собой появление сознательного опыта. Так что перспективы машинного сознания можно признать хорошими — пусть и не на практике, но хотя бы в принципе.

Я мало говорил о том, какого *рода* вычисления, скорее всего, достаточны для сознательного опыта. В большинстве аргументов я использовал для иллюстрации понейронную симуляцию мозга; но вероятно, что для этого могло бы быть достаточным и множество других видов вычислений. Могло бы, к примеру, быть так, что вычисление, отражающее каузальную организацию мозга на гораздо более грубом уровне, передавало бы тем не менее те моменты, которые релевантны для возникновения сознательного опыта. Вероятно и то, что вычисления совершенного иного вида, соответствующие совершенно другим типам каузальной организации, при их имплементации тоже могли бы порождать богатые сознательные переживания.

Эта картина в равной мере совместима как с символным, так и с коннекционистским подходом к познанию, а также и с другими вычислительными подходами. Действительно, можно было бы попробовать доказать, что центральная роль вычисления в исследовании познания связана с тем, что вычислительные конструкции могут передавать практически *любую* разновидность каузальной организации. Мы можем рассматривать вычислительные формализмы в качестве источника идеального формализма для выражения паттернов каузальной организации и, более того (в сочетании с методами имплементации), в качестве идеального инструмента для их воспроизведения. Какая бы каузальная организация ни оказалась ключевой для познания и сознания, мы можем ожидать, что какая-то вычислительная конструкция сможет точно передать ее. Можно было бы даже попытаться показать, что именно эта гибкость скрывается за часто упоминаемой *универсальностью* вычислительных систем. Сторонники искусственного интеллекта не обязаны подписываться под каким-то одним видом вычисления, только и достаточным для ментальности; тезис ИИ столь правдоподобен именно из-за широты класса вычислительных систем¹⁰.

Так что вопрос о том, какой именно класс вычислений достаточен для воспроизведения человеческой ментальности, остается открытым; но у нас есть серьезное основание верить, что этот класс не является пустым.

Глава 10

Интерпретация квантовой механики

1. Две тайны

Проблема квантовой механики почти столь же трудна, как проблема сознания. Квантовая механика дает нам удивительно точные формулы для предсказания результатов эмпирических наблюдений, но картина мира, которую она при этом рисует, лишь с очень большим трудом поддается осмыслению. Как наш мир может быть таким, каким он должен быть, чтобы предсказания квантовой механики оказывались успешными? В ответе на этот вопрос нет ничего даже отдаленного напоминающего консенсус. Как и в случае с сознанием, нередко кажется, что ни одно из решений проблемы квантовой механики не может быть удовлетворительным.

Многие полагали, что между двумя этими наиболее загадочными проблемами могла бы существовать какая-то глубокая связь (напр., Bohm 1980; Hodgson 1988; Lockwood 1989; Penrose 1989; Squires 1990; Stapp 1993; Wigner 1961). Если есть две тайны, то соблазнительно предположить, что у них имеется общий исток. Это искушение усиливается тем обстоятельством, что проблемы квантовой механики кажутся тесно сопряженными с понятием наблюдения, сущностным образом предполагающим отношение между опытом какого-то субъекта и всем остальным миром.

Чаще всего высказывалось предположение, что квантовая механика могла бы оказаться ключом к физическому объяснению сознания. Но, как мы видели, этот проект обречен на неудачу. Квантовые «теории» сознания в итоге страдают от такого же провала в объяснении, что и классические теории. В любом случае опыт должен рассматриваться как нечто выходящее за пределы физических свойств мира. Квантовая механика, возможно, могла бы содействовать характеристике психо-физической связи, но одна лишь квантовая теория не может сказать нам, почему существует сознание.

Впрочем, две эти проблемы могут быть увязаны и более тонким образом. Даже если квантовая механика и не объясняет сознание,

не исключено, что теория сознания могла бы пролить свет на проблемы квантовой механики. В конце концов, многие признают, что эти проблемы каким-то образом связаны с наблюдением и опытом. Естественно предположить, что теория опыта могла бы помочь нам разобраться в этом. Некоторые говорили об активной роли сознания в квантовой теории — допуская, к примеру, что сознание осуществляет «коллапс волновой функции» — но я буду отстаивать позицию, согласно которой роль сознания во всем этом не столь непосредственна. В частности, я попытаюсь показать, что мы можем переосмыслить проблемы квантовой теории, представив их в качестве проблем отношения между физической структурой мира и нашим опытом мира, и что, следовательно, надлежащая теория сознания может подкрепить неортодоксальные интерпретации квантовой механики.

2. Каркас квантовой механики

Основным каркасом квантовой механики является исчисление для предсказания результатов экспериментальных измерений. Здесь я охарактеризую одну из версий этого исчисления, оставляя в стороне множество технических деталей, чтобы дать простое описание, покрывающее большинство существенных черт. В этом параграфе я трактую этот каркас только как исчисление для эмпирических предсказаний, оставляя открытым вопрос о том, дает ли он прямое описание физической реальности. Глубокие интерпретационные проблемы обсуждаются в следующем параграфе.

В классической модели состояние физической системы может быть выражено в очень простых терминах. Состояние частицы, к примеру, выражается путем указания определенных значений каждого из множества свойств, таких как положение и импульс. Мы можем назвать простое значение такого рода *базовым значением*. В квантовой модели все не так просто. В общем, состояние системы должно быть выражено *волновой функцией* или *вектором состояния*. Релевантные свойства не могут выражаться здесь простыми значениями, а должны выражаться некоей комбинацией базовых значений. Квантовое состояние может рассматриваться как *суперпозиция* более простых состояний.

Простейшим примером является такое свойство, как спин, имеющий только два базовых значения¹. Эти базовые значения могут быть поименованы как «вверх» и «вниз». Впрочем, в квантовой механике спин частицы не всегда оказывается «вверх» или «вниз». В общем, спин частицы должен выражаться *комбинацией* «вверх» и «вниз», в каждом случае со своей комплексной магнитудой. Спин частицы поэтому лучше всего рассматривать как вектор в двумерном векторном простран-

стве. И его наиболее естественной визуализацией будет суперпозиция состояния спина-«вверх» и состояния спина-«вниз», с различными соответствующими им магнитудами.

Это же верно для положения и импульса, разве что они имеют бесконечное множество базовых значений. Положение и импульс классической частицы могут принимать любое из бесконечного множества значений в континууме. Соответственно, положение квантовой частицы должно выражаться бесконечномерным вектором с различной магнитудой для каждого из этих местоположений. Этот вектор лучше всего трактовать как *волну* с различными амплитудами в различных местах пространства; функция, соотносящая место с соответствующей амплитудой, есть волновая функция. Импульс квантовой частицы тоже может рассматриваться в качестве волны с различными амплитудами при разных базовых значениях импульса. Положение или импульс такой частицы можно опять-таки мыслить в качестве суперпозиции базовых значений положения или импульса.

Поскольку эти состояния — всего лишь векторы, они могут быть разложены на свои компоненты разными способами. Хотя зачастую полезно рассматривать вектор двумерного спина в качестве суммы компонента «вверх» и компонента «вниз», его можно разложить и множеством других способов, в зависимости от того, что принимается за базис векторного пространства. Все эти базисы в равной степени «естественны»; ни один из них не является предпочтительным для природы. На деле оказывается, что один и тот же вектор репрезентирует как положение, так и импульс частицы. При разложении по одному базису мы получаем «позиционные» амплитуды; при разложении по другому — «импульсные» амплитуды. В общем, релевантное в том или ином случае разложение зависит от того, к какой из величин мы проявляем интерес, и, в частности, от того, какую из величин мы решаем *измерить*, о чем я вскоре еще поговорю.

Состояния систем из более чем одной частицы несколько более сложны, но главная идея остается неизменной. Возьмем систему, состоящую из двух частиц, *A* и *B*. Состояние этой системы, как правило, нельзя выразить простой комбинацией волновой функции *A* и волновой функции *B*; состояния двух этих частиц часто будут *нераздельными*. Состояние данной системы должно быть выражено, скорее, в виде волновой функции в более сложном пространстве. Однако эта волновая функция может рассматриваться как некая суперпозиция более простых состояний системы с двумя частицами, так что общая картина все равно будет применима к данному случаю. Это же верно и для более сложных систем, состояние которых по-прежнему лучше всего будет репрезентировать волновой функцией, соответствующей суперпозиции состояний.

Все это континтуитивно, но еще не парадоксально. Если мы буквально истолкуем этот формализм как описание реальности, то осмыслить его будет не *слишком* трудно. Некоторые полагали, что он несовместим с «объективным» представлением о мире, так как из него вытекает, что вещам в мире не может быть присуще объективного, определенного состояния. Однако это вовсе не следует из сказанного. Согласно данной картине, состояние некоей сущности лучше всего выражается волновой функцией, а не дискретными величинами, но оно тем не менее является совершенно определенным состоянием. Эта картина просто говорит нам, что на базовом уровне реальность подобна волне. Это требует нового образа мысли, но мы можем привыкнуть к нему. В конце концов, базовый уровень микроскопической реальности очень далек от макроскопического уровня, с которым мы обычно имеем дело, и не кажется чем-то совершенно удивительным то, что здесь могут встречаться какие-то необычные свойства. Проблемы возникают из-за *других* характеристик квантовой механики.

Ядро квантовой механики составляют два принципа, определяющие динамику волновой функции: *уравнение Шредингера* и *постулат изменения*. Два этих совершенно разных принципа совместно определяют изменение волновой функции той или иной системы с течением времени.

Основное содержание квантовой механики заключено в уравнении Шредингера. Это дифференциальное уравнение, определяющее изменение волновой функции системы *почти* при любых обстоятельствах. Детальная структура этого уравнения не важна для наших целей. Наиболее важной характеристикой является в данном случае то, что оно представляет собой линейное дифференциальное уравнение: если имеются два состояния, *A* и *B*, такие, что *A* изменяется в *A'*, а *B* — в *B'*, то состояние, оказывающееся суперпозицией *A* и *B*, будет трансформироваться в суперпозицию *A'* и *B'*. Стоит также отметить, что динамика уравнения Шредингера такова, что относительно дискретные состояния имеют тенденцию к размытию с течением времени. Состояние, изначально являющееся суперпозицией значений в каком-то ограниченном диапазоне, как правило, трансформируются в суперпозицию значений в гораздо большем диапазоне. Наконец, уравнение Шредингера является абсолютно детерминистическим.

Уравнение Шредингера весьма прозрачно и не вызывает больших вопросов. Именно в нем заключена содержательная часть квантовой теории. При применении квантовой теории к практическим или экспериментальным проблемам дело по большей части сводится к вычислению эволюции различных состояний согласно шредингеровской динамике.

Уравнение Шредингера, однако, *не исчерпывает* сути дела. Согласно этому уравнению, подавляющее большинство физических состояний

вскоре трансформируется в суперпозицию широкого диапазона состояний. Но это не сочетается с наблюдаемым нами миром. Когда мы измеряем положение частицы, мы получаем какое-то определенное значение, а не суперпозицию значений, которую предсказывало бы уравнение Шредингера. Если бы квантовая динамика сводилась к уравнению Шредингера, то даже на макроскопическом уровне мир оказался бы в состоянии дикой суперпозиции. Но наш опыт говорит, что этого не происходит. Стрелки расположены определенным образом, движущиеся объекты наделены определенным и поддающимся измерению импульсом и т. п. Так что здесь должно быть еще что-то, позволяющее нам переходить от этого уравнения к таким дискретным событиям, которые характеризуют наш опыт.

Второй частью картины в стандартном формализме является *постулат измерения* (известный также как постулат коллапса или проекции). В нем утверждается, что при определенных обстоятельствах шредингеровская динамика оказывается неприменимой. А именно, утверждается, что при проведении *измерения* волновая функция *коллапсирует* в нечто более определенное. Тип ее коллапсирования зависит от того, какое свойство подвергается измерению. К примеру, если мы измеряем спин частицы, то, даже если до этого она находилась в состоянии суперпозиции, она коллапсирует в состояние, в котором спин будет либо вверх, либо вниз. Если мы измеряем положение частицы, то ее волновая функция коллапсирует в состояние с вполне определенным положением². Итоговое состояние по-прежнему соответствует волновой функции, но эта волновая функция такова, что вся ее амплитуда сконцентрирована в определенном положении; во всех остальных местах она оказывается равной нулю. Любой величине, которую мы могли бы измерить, соответствует оператор, и при измерении данное состояние коллапсирует в состояние, являющееся *собственным состоянием* этого оператора. Собственное состояние оператора — это всегда состояние, в котором соответствующая ему измеряемая величина имеет какое-то определенное значение. Из этого следует, что когда мы измеряем величину, результатом всегда будет определенное значение этой величины, что в точности согласуется с имеющимся у нас опытом.

Динамика коллапса имеет скорее вероятностный, чем детерминистический характер. Если частица находится в состоянии, являющемся суперпозицией каких-то положений, то при измерении ее положения мы знаем, что она коллапсирует в состояние с определенным положением, но не знаем, каким оно будет. Скорее для каждого потенциально коллапсированного состояния постулат измерения конкретизирует *вероятность* того, что данная система коллапсирует в это состояние. Эта вероятность³ задается квадратом амплитуды волновой функции в по-

ложении, соответствующем определенному значению, о котором идет речь. К примеру, если спин частицы — суперпозиция спина-«вверх» (с амплитудой $1/2$) и спина-«вниз» (с амплитудой $\sqrt{3}/2$), то при измерении спина он коллапсирует в состояние спина-«вверх» с вероятностью $1/4$ и в состояние спина-«вниз» с вероятностью $3/4$. Амплитуды волновой функции всегда таковы, что сумма соответствующих вероятностей равна 1.

3. Интерпретируя квантовую механику

Два этих принципа вместе составляют исключительно мощное исчисление для предсказания результатов экспериментальных измерений. Для предсказания результатов эксперимента мы выражаем состояние системы в виде волновой функции и вычисляем, как эта волновая функция изменяется во времени в соответствии с уравнением Шредингера до момента измерения. При измерении мы используем амплитуды вычисленной волновой функции для определения вероятности появления различных коллапсированных состояний и того, что измерение даст нам ту или иную величину. Экспериментальные результаты однозначно подкрепляли предсказания данной теории; мало какие научные теории были столь же успешны в плане предсказаний. В качестве исчисления эта теория очень прочна.

Проблемы возникают, когда мы задаемся вопросом о том, как могло бы работать это исчисление. Что могло бы случиться в реальном мире для обеспечения столь высокой точности предсказаний данного исчисления? Это проблема *интерпретации* квантовой механики. Существует множество вариантов решения этой проблемы, но все они не лишены недостатков.

Вариант 1:
Понимать
исчисление
буквально

Первой естественной реакцией было бы истолковать формализм квантовой механики в буквальном смысле, как мы это делаем с большинством научных теорий. Данное исчисление содержит волновую функцию, задаваемую динамикой уравнения Шредингера и постулатом измерения, и оно работает, так что мы должны предполагать, что оно рисует нам непосредственную картину происходящего в мире. Иначе говоря, состояние системы на самом деле является именно этим волновым состоянием, выражаемым волновой функцией и изменяющимся сообразно динамике, выражаемой двумя упомянутыми базовыми принципами. В основном это состояние изменяется в соответствии с уравнением Шредингера, но при измерении оно изменяется в соответствии с постулатом измерения. Согласно этому воззрению,

мир состоит из волн, обычно линейно изменяющихся в суперпозиции и иногда коллапсирующих в более определенное состояние при осуществлении измерения.

Осмыслить эту картину, однако, не так просто. Все проблемы идут от постулата измерения. В соответствии с этим постулатом, коллапс происходит при выполнении измерения, но что считать измерением? Как *природа* узнает о том, в какой момент производится измерение? «Измерение» уж точно не является каким-то базовым термином законов природы; если постулат измерения должен хотя бы отдаленно напоминать фундаментальный закон, понятие измерения должно быть заменено чем-то более ясным и основополагающим. Если коллапс волновой функции — объективно существующий процесс в мире, то нам нужны ясные, объективные критерии, позволяющие установить, когда этот процесс имеет место.

Очевидно ошибочным было бы сказать, что коллапс происходит в тех случаях, когда квантовая система взаимодействует с *измерительным прибором*. Проблема здесь в том, что наличие понятия «измерительного прибора» в базовых законах столь же неправдоподобно, как и наличие в них понятия «измерения». Раньше нам нужны были критерии для измерения; теперь нам требуются критерии для измерительного прибора.

На заре квантовой механики часто говорили, что измерительный прибор — это *классическая* система, и что измерение происходит при любом взаимодействии квантовой системы с классической системой. Ясно, однако, что это неудовлетворительное решение. Квантовая теория имеет, как предполагается, универсальный характер, и она должна быть применима к процессам внутри измерительного прибора ничуть не меньше, чем ко всем остальным. Если мы не готовы допустить, что в мире существует два фундаментально различных вида физических объектов — а это допущение потребовало бы создания совершенно новой теории — то выражение «*классическая система*» не может быть термином, используемым в фундаментальном законе природы — ничуть не отличаясь этим от «измерения».

Указанной идеи аналогично и предположение, что измерение происходит всякий раз, когда квантовая система взаимодействует с *макроскопической* системой. Ведь столь же очевидно, что «макроскопическое» не является понятием, которое может фигурировать в базовом законе. Оно должно быть заменено чем-то более точным: чем-то вроде «система, имеющая массу в один грамм или больше». Однако крайним произволом было бы, если бы в базовом законе фигурировало нечто подобное.

Нет такого физического критерия, который казался хотя бы отдаленно пригодным для определения коллапса. Критерии, выдвинутые

на микроскопическом уровне — допускающие, к примеру, что коллапс происходит, когда система взаимодействует с протоном — исключаются экспериментальными данными. Альтернатива состоит в том, что критерий должен содержать отсылку к высокоуровневому физическому свойству, так что коллапс происходит, когда система имеет определенную высокоуровневую конфигурацию. Однако любое подобное высокоуровневое свойство казалось бы произвольным, и хороших кандидатов пока не предлагалось. В предположении, что шредингеровская динамика микроскопических систем неожиданно отменялась бы при попадании этих систем в контекст каких-то конкретных конфигураций, есть что-то очень странное.

Из всех выдвигавшихся критериев какие-то шансы есть лишь у того, согласно которому измерение происходит, когда квантовая система затрагивает *сознание* какого-то существа. В отличие от предыдущих критериев, он хотя бы определен и непроизведен⁴. Соответствующая интерпретация исчисления весьма элегантна и проста, и это единственная буквальная интерпретация данного исчисления, получившая достаточно широкое распространение. Она была впервые предложена Лондоном и Баузером (London and Bauer 1939), но чаще всего ассоциируется с Вигнером (Wigner 1961).

Заметим, что эта интерпретация *предполагает* дуализм сознания и тела. Если бы сознание было всего лишь еще одним физическим свойством, то возникли бы все упомянутые выше проблемы. Эта позиция оказалась бы одной из разновидностей концепции «высокоуровневого свойства», согласно которой попросту случается так, что волновые функции физических систем коллапсируют в контексте каких-то сложных физических конфигураций. Но если дуализм верен, то критерий коллапсирования может быть подлинно фундаментальным. Кроме того, тот факт, что причина коллапса является чем-то внешним для физических процессов, позволяет создать гораздо более простую теорию. Все чисто физические системы управляются теперь одной лишь шредингеровской динамикой, а совершенно иная динамика измерения имеет независимый источник.

У этой интерпретации, однако, имеется ряд контринтуитивных следствий. Возьмем измерительное устройство, к примеру стрелку, позволяющую измерить состояние электрона, и предположим, что это состояние изначально характеризуется суперпозицией. Когда поблизости нет сознания, вся система будет подчиняться линейной шредингеровской динамике: если различные дискретные состояния электрона дают различные дискретные состояния стрелки, то из этого следует, что состояние электрона, характеризующееся суперпозицией, будет порождать такое состояние стрелки, которое тоже будет *характеризоваться суперпозицией*. Иными словами, эта теория предсказывает, что

стрелка одновременно указывает на множество разных мест! Она указывает на какое-то определенное место лишь тогда, когда я *смотрю* на эту стрелку.

Сценарий с кошкой Шредингера ведет к еще более странным последствиям. В этом сценарии некая кошка закрыта в камере, какое-то устройство измеряет спин электрона, и система настроена так, что кошка гибнет лишь тогда, когда спин электрона оказывается в положении «вверх». (Допустим, что кошка находится под анестезией, так что сознание не принимает участие во всех этих процессах.) Если электрон изначально находится в состоянии суперпозиции, то кошка попадет в состояние, являющееся суперпозицией жизни и смерти! Только когда сознающее существо заглянет в камеру, определится, является ли кошка мертвой или живой.

Согласно этой картине, *любой* макроскопической системе, как правило, будет свойственна масштабная суперпозиция при отсутствии proximity сознания. До возникновения сознания в процессе эволюции весь мир находился в состоянии гигантской суперпозиции, пока первая крупинка сознания, судя по всему, не привела его к внезапному коллапсу. Это может звучать безумно, но таково прямое следствие единственной внятной буквалистской интерпретации принципов квантовой механики. Надеюсь, что это поможет уяснить, насколько странной является квантовая механика и насколько серьезны проблемы, связанные с ее интерпретацией.

Не исключено, что указанные контриintuitивные следствия можно было бы и принять, но я тем не менее не являюсь сторонником этой интерпретации. Начать с того, что она не сочетается с той позицией, которую я отстаивал, а именно с идеей повсеместности сознания. Если сознание связано даже с очень простыми системами, то, согласно этой интерпретации, коллапс будет происходить на самом базовом уровне и очень часто. Это не стыкуется с физическими данными, предполагающими, что низкоуровневые суперпозиции зачастую продолжительное время существуют в неколлапсированном состоянии. Вторая проблема состоит в том, что не существует даже намека на основательную теорию о том, *какого рода* воздействие на сознание порождает коллапс, равно как и о том, какой вид принимает итоговый коллапс. Можно по-разному отвечать на эти вопросы, но ни одна из конкретизаций не кажется особо убедительной.

Другие проблемы связаны с самим понятием коллапса. Начать с того, что коллапс должен быть *нелокальным*: если две частицы имеют сцепленные состояния, измерение первой частицы вызовет одновременный коллапс состояния второй частицы. Это приводит к определенной нестыковке с теорией относительности. К примеру, кажется,

что нелокальный коллапс подразумевает отсылку к привилегированной системе отсчета. Без нее время коллапса второй частицы будет недостаточно определенным, так как у нас нет четкой дефиниции одновременности в разных местах.

Более общая трудность состоит в том, что процесс коллапса в целом не очень хорошо сочетается со всей остальной физикой. Если брать его в буквальном смысле, то это мгновенный, дискретный, темпорально асимметричный нелокальный процесс, радикально отличный от всех других процессов, в существовании которых нас убеждает физическая теория. Кажется странным, что такой необычный процесс должен сосуществовать с прямолинейным, континуальным, темпорально симметричным и локальным уравнением Шредингера. В самом деле, сравнивая коллапс с элегантностью и мощью уравнения Шредингера, составляющего ядро квантовой теории, мы можем счесть его каким-то случайным привнесением. И что-то несообразное есть в представлении о том, что мир на своем базовом уровне содержит два столь совершенно разных вида динамики.

Эти аргументы, конечно, далеко не решают дело, и интерпретация, согласно которой сознание коллапсирует волновую функцию, заслуживает очень серьезного отношения. Тем не менее, на мой взгляд, имеется серьезное основание поискать другую интерпретацию, рисующую более простую и непосредственную картину фундаментальных природных процессов.

Вариант 2:

**Попытка получить
постулат измерения
задаром**

Все проблемы буквалистской интерпретации связаны с признанием постулата измерения фундаментальным законом. Заманчиво поэтому предположить, что этот постулат мог бы *не иметь базового характера*, а оказываться

следствием более фундаментальных принципов. Тут есть два возможных пути. Мы могли бы попытаться ввести *дополнительные* базовые принципы, не столь проблематичные, как постулат измерения, но приводящие к тем же результатам. Это стратегия четвертого варианта. Или же мы могли бы попробовать получить те же результаты в качестве следствия уже известных базовых принципов, таких как уравнение Шредингера. То есть мы могли бы попытаться получить постулат измерения даром.

Интуитивную мотивацию в пользу этой стратегии нетрудно понять. Интуиция говорит нам, что эффекты суперпозиции проявляются по большей части на микроскопическом уровне и что они могли бы каким-то образом «отменяться» на макроскопическом уровне. Быть может, при наличии множества микроскопических суперпозиций они взаимодействуют так, что порождают относительно определенное

макроскопическое состояние. Благодаря ряду математических свойств сложных конфигураций мы могли бы понять, каким образом эффективный коллапс мог бы оказываться следствием микроскопической неопределенности. Базовый пробабилистический коллапс был бы в таком случае заменен эмерджентным статистическим процессом, протекающим в той или иной сложной системе.

Предпринималось множество попыток проработать математическую сторону этого предположения, с целью чего нередко обращались к статистическим принципам термодинамики (напр., Daneri, Loinger, and Prosperi 1962). К сожалению, все эти попытки провалились, и сегодня многие считают, что они и должны проваливаться. Поскольку шредингеровская динамика линейна, всегда можно сконструировать ситуации, в которых микроскопические суперпозиции ведут к макроскопическим суперпозициям. Если электрон в положении «вверх» ведет к одному макроскопическому состоянию, а электрон в положении «вниз» — к другому, то электрон в состоянии суперпозиции должен вести к макроскопической суперпозиции (Альберт (Albert 1992, с. 75) очень четко обосновывает это заключение). Таким образом, без введения дополнительных базовых принципов приходится ожидать, что мы обнаружим суперпозицию на макроскопическом уровне.

Эти стратегии небесполезны. Подобное использование статистики, а также более поздние работы о «декогеренции» (Gell-Mann and Hartle 1990 и др.) показывают, что волновая функция с суперпозицией будет не какой-то хаотичной мешаниной, а зачастую будет разрешаться в сравнительно четко очерченную суперпозицию раздельных макроскопических состояний. Эти макроскопические состояния «декогерируют» друг от друга лишь с минимальными интерференционными взаимовлияниями. Это по крайней мере помогает нам отыскать какой-то элемент привычного классического мира в волновой функции с характерной для нее суперпозицией. Но суперпозиция волновой функции никуда не исчезает, и подобные работы ничего не говорят нам о том, почему лишь один из элементов макроскопической суперпозиции должен быть актуальным. Так что для решения этой базовой проблемы требуются дополнительные усилия. Они могут оказаться наиболее продуктивными в сочетании с другими вариантами решения, в частности с тем, который рассмотрен тут под номером 5.

Возможно, самой распространенной позицией среди практикующих физиков является та, согласно которой мы попросту не должны задаваться вопросом о том, что происходит в реальном мире за пределами исчисления квантовой механики. Исчисление работает, и этого достаточно. Есть две разновидности этой позиции. В соответствии с первой

Вариант 3:
О чем нельзя
говорить...

из них, в мире, возможно, *что-то* и происходит, но мы никогда не сможем узнать, что именно. Исчисление дает нам всю в принципе доступную эмпирическую информацию, так что все остальное есть чистая спекуляция. Мы вполне могли бы оставить свое беспокойство и продолжать вычисления. Эта позиция оправдана в практических целях, но она не может удовлетворить тех, кто хочет, чтобы физика рассказывала нам о базовом уровне реальности. Если исчисление работает, то мы хотели бы иметь хоть какое-то представление о том, *как* оно могло бы работать. Не исключено, что мы никогда не сможем узнать об этом наверняка, но спрашивать об этом все же имеет смысл.

Вторая разновидность указанной позиции имеет более жесткий характер. Здесь утверждается, что вопросу о том, что происходит в мире, не соответствует никакого фактического содержания. Согласно этому вззрению, факты исчерпываются фактом работоспособности исчисления. Оно не часто формулируется в столь эксплицитном виде, так как при подобной формулировке в эти утверждения почти невозможно поверить. Оно предлагает нам такую картину реальности, в которой нет места миру! Эта позиция ведет к одной из версий идеализма, в соответствии с которым все, что существует, — это наши перцепции или что-то вроде того. Перед тем, как мы открыли контейнер с кошкой Шредингера, она не мертва, она не жива, и она не находится в состоянии суперпозиции; она вообще не находится ни в каком состоянии. Отрицая факты за пределами наших измерений, эта концепция отрицает независимо существующую реальность.

«Копенгагенскую интерпретацию», выдвинутую Бором и его коллегами, нередко считают одной из вариаций этого вззрения, хотя сочинения Бора несколько неоднозначны и сложны для толкования. В них можно найти намеки и на какие-то моменты, характерные для первого варианта решения, равно как и на эпистемологическую версию данного варианта. Бор всячески подчеркивал «классическую» природу измерительных приборов, и его позицию можно понять так, будто он считал, что только классические (или макроскопические) объекты имеют объективные состояния. На вопросы же о реальном состоянии объекта, описываемого с помощью суперпозиции, попросту накладывается запрет. Это, однако, предполагает разведение классических и квантовых систем, которое трудно осуществить по объективным основаниям; трудно также представить, что реальность просто «расстворяется» по мере того, как мы переходим от макроскопического к микроскопическому уровню. Многим казалось, что если принимать идеи Бора всерьез, то они приведут к жесткому операционализму, обсуждавшемуся в предыдущем абзаце. Они тоже рисуют такую картину базового уровня реальности, которая ничего не изображает.

Если буквалистская интерпретация постулата измерения неприемлема и если он не может быть выведен из имеющихся физических принципов, то естественно предположить, что эти принципы не исчерпывают того, что здесь происходит. Быть может, если мы постулируем *дополнительные базовые физические принципы*, то мы сможем менее проблематично объяснить эффективность квантово-механического исчисления.

Первый способ сделать это — удержать идею коллапса, но дать ей иное объяснение. Подобная стратегия сохраняет допущение, что базовые физические состояния — это волновые функции, задаваемые уравнением Шредингера, но вводит новые принципы для объяснения того, как микроскопические суперпозиции превращаются в макроскопическую дискретность.

Самым известным примером данной стратегии является «ЖРВ»-интерпретация, которой мы обязаны Жирарди, Римини и Веберу (Ghirardi, Rimini, and Weber 1986; см. также Bell 1987a)⁵. В этой интерпретации постулируется фундаментальный закон, в соответствии с которым вектор состояния любой элементарной частицы может претерпевать микроскопический «коллапс» в любой момент с очень небольшой вероятностью (шанс того, что частица коллапсирует в данную секунду, примерно равен 1 из 10^{15}). Когда подобный коллапс происходит, это, как правило, приводит к коллапсу состояния той макроскопической системы, в которую она включена, вследствие наличия там неразрывных связей. В любой макроскопической системе имеется множество таких частиц, так что из сказанного следует, что любая макроскопическая система в любое время, как правило, будет находиться в относительно дискретном состоянии. И можно показать, что итоговый результат будет очень близок к тому, чтобы воспроизвести предсказания постулата измерения.

Альтернатива состоит в том, чтобы исключить необходимость в коллапсе путем отрицания того, что базовый уровень реальности репрезентируется волновой функцией с присущей ей суперпозицией. Если такие свойства, как положение, имеют определенные значения даже на базовом уровне, то в коллапсе не будет никакой необходимости. Подобная теория постулирует «скрытые параметры» на базовом уровне, напрямую объясняющие дискретность реальности на макроскопическом уровне. Ценой этого предположения оказывается то, что теперь мы нуждаемся в новых принципах, которые позволили бы объяснить, почему принципы изменения волновой функции и коллапса *кажутся* настолько эффективными.

Вариант 4:

Вводить дополнительные базовые физические принципы

Наиболее известной в этой связи является теория, разработанная Бомом (Bohm 1952). Согласно этой теории, положение базовых частиц всегда является чем-то определенным. Волновая функция сохраняет роль некоей «волны-пилота», направляющей изменение положения частицы, и сама волновая функция подчиняется уравнению Шредингера. Вероятностные предсказания постулата измерения получают новую интерпретацию и трактуются в качестве *статистических законов*. В соответствии с этой теорией оказывается, что мы никогда не можем знать точного положения частицы до соответствующего измерения, а можем знать только ее волновую функцию. Постулат измерения говорит нам о том, какая *пропорция* частиц при данной волновой функции будет иметь то или иное положение. Поэтому он дает нам наилучшие статистические предсказания, которые мы только можем ожидать, учитывая отсутствие у нас знания об этом.

Все предложения указанного типа сталкиваются с проблемами. Как ЖРВ-интерпретация, так и интерпретация Бома подчеркивают определенность *положения*, нарушая тем самым симметрию между положением и импульсом в квантово-механическом исчислении. Это оправдано в целях предсказания, так как можно попробовать показать, что наши суждения о макроскопической определенности всегда предполагают отсылку к каким-то определенным положениям (примером может быть положение стрелки), но делает теорию менее элегантной. По сходным причинам очень непросто согласовать эти подходы и с теорией относительности.

ЖРВ-теория сталкивается и с дополнительными трудностями, наиболее серьезной из которых, возможно, является то, что из нее, строго говоря, не следует, что макроскопический мир вообще дискретен. Макроскопическое состояние по-прежнему репрезентируется волновой функцией с суперпозицией: хотя ее амплитуда по большей части концентрируется в одном месте, она не равна нулю везде, где не равна нулю амплитуда неколлапсированной волновой функции. И это возобновляет проблемы, связанные с суперпозицией. Стрелка по-прежнему показывает на множество мест, даже после измерения. Конечно, амплитуда, связанная с большинством этих мест, очень невелика, но трудно понять, почему низкоамплитудная суперпозиция хоть в чем-то более приемлема, чем высокоамплитудная.

Теория Бома имеет меньше технических проблем в сравнении с ЖРВ-интерпретацией, но из нее вытекает ряд странных следствий. Наиболее поразительным из них является ее крайняя *нелокальность*. (Любая теория, говорящая о скрытых параметрах и сообразная с предсказаниями данного исчисления, должна быть нелокальной по причинам, указанным Беллом, — см. Bell 1964⁶.) Дело не сводится к тому,

что свойства одной частицы могут мгновенно влиять на свойства другой частицы, находящейся на расстоянии от нее. Оказывается, что при определении траектории частицы мы, возможно, должны принимать во внимание волновые функции частиц, существующих в других галактиках! Все они содействуют формированию глобальной волновой функции, и эта волновая функция одновременно управляет траекториями частиц во всей вселенной.

Но, возможно, наиболее фундаментальным основанием с подозрением относиться к этим интерпретациям является то, что они постулируют *наличие чего-то сложного за чем-то простым*. Невзирая на проблемы квантово-механического исчисления, оно исключительно просто и элегантно. Эти же интерпретации вводят сложные дополнительные принципы, имеющие к тому же характер допущений, сделанных сравнительно *ad hoc*, для замены и объяснения этого простого каркаса. Сказанное несколько в меньшей степени применимо к ЖРВ-интерпретации, дополнительное усложнение которой состоит только во введении двух новых фундаментальных констант и нарушении симметрии между положением и импульсом; но остается удивительной «удачей» то, что значения констант оказываются именно такими, что они могут почти воспроизводить предсказания стандартной модели. Дополнительные усложнения, связанные с интерпретацией Бома, хуже тех; она постулирует определенность положений и волновую функцию, сложный новый фундаментальный принцип, посредством которого волновая функция определяет положение частиц, а также нарушает симметрию изначальной модели.

Мы могли бы сказать, что эти интерпретации создают такое впечатление, будто мир был создан Декартовым злым демоном, так как они заставляют нас верить, что мир устроен не так, как в действительности. Как выражаются Алберт и Левер (Albert and Loewer 1989), Бог у Бома не играет в кости, но наделен извращенным чувством юмора. Сценарий, в котором сложная бомовская интерпретация воспроизводит предсказания простой модели лишь по своей степени отличается от того случая, когда данные, поступающие в мозг в бочке, выстраиваются таким образом, чтобы создавать видимость непосредственно данного внешнего мира. Она напоминает «интерпретацию» эволюционной теории, согласно которой Бог в готовом виде создал окаменелости несколько тысяч лет назад, обеспечив тем самым воспроизведение предсказаний эволюционной теории. Простота объяснительной модели была принесена в жертву ради сложной гипотезы, которой удается воспроизвести результаты изначальной теории.

Квантово-механическая модель настолько проста и элегантна, что базовая теория, не воспроизводящая эту простоту и элегантность, никогда не будет удовлетворять нас или не будет абсолютно убедительной.

Если бы в квантовой теории были какие-то аномалии, какие-то экспериментальные результаты, в полной мере не предсказуемые в этой модели, то более правдоподобным мог бы оказаться вывод, что эта простота является всего лишь верхушкой сложного айсберга. В настоящий момент, однако, эта модель кажется настолько надежной, что представление о необходимости постулировать сложный механизм для объяснения простых предсказаний в рамках этой модели кажется чем-то экстраординарным.

Учитывая, что *ни одна* из интерпретаций квантовой механики не лишена проблем, эти интерпретации заслуживают серьезного отношения. Но естественным будет поискать какую-то более простую картину мира.

Вариант 5:

Достаточно одного лишь уравнения Шредингера

Основой квантовой механики является уравнение Шредингера, и оно так или иначе присутствует в любой интерпретации квантовой механики. Все рассмотренные нами интерпретации добавляли что-то к уравнению Шредингера для

объяснения макроскопической дискретности мира. Но самой простой из имеющихся интерпретаций является та, которая говорит о верности уравнения Шредингера и заявляет, что больше ничего и не требуется. Иными словами, физическое состояние мира полностью характеризуется волновой функцией, и ее изменение целиком описывается уравнением Шредингера. Это интерпретация Эверетта (Everett 1957, 1973).

Одна из очерченных выше стратегий (вариант 2) также предполагает возможность ограничиться уравнением Шредингера, но при этом утверждает, что это можно совместить с дискретностью на макроскопическом уровне. Мы видели, что это не должно проходить по ясным математическим соображениям. Интерпретация Эверетта гораздо более радикальна. Уравнение Шредингера не подвергается здесь никаким сомнениям, и состояние мира описывается волновой функцией на каждом из его уровней. Из этого следует, что в противоположность имеющейся видимости, мир находится в состоянии суперпозиции даже на макроскопическом уровне.

4. Интерпретация Эверетта

Мотивы, стоящие за этой интерпретацией, очевидны. Ядром квантовой механики является уравнение Шредингера. Постулат измерения и все другие предлагавшиеся принципы кажутся какими-то добавками. Так почему бы не избавиться от них? В равной степени очевидна и проблема, связанная с этой интерпретацией. Если все исчерпывается уравнением Шредингера, то на каждом из своих уровней мир оказывается в состоянии суперпозиции. Но он не *выглядит* таковым: мы

никогда не наблюдаем стрелки, оказывающиеся в суперпозиции двух состояний. Почему же?

Эта интерпретация в самом лучшем случае крайне континтутивна. Согласно этому воззрению, суперпозицией лучше всего описывать не только состояние электрона, но и состояние измеряющей его стрелки! Если рассуждать объективно, то, строго говоря, неверно утверждать, что стрелка показывает вверх или вниз. Скорее, она находится в суперпозиции состояний указывания вверх и указывания вниз. Это же верно для макроскопических состояний почти всего, что находится в состоянии, описываемом волновой функцией, которая почти никогда не соответствует единичному «дискретному» состоянию. Согласно этому воззрению, суперпозиция существует повсюду. Но почему же тогда мир кажется дискретным?

Ответ Эверетта состоит в том, чтобы *расширить суперпозицию до такой степени, чтобы она затрагивала и сознание*. Если мы принимаем уравнение Шредингера всерьез, то если стрелка, измеряющая электрон, находится в суперпозиции состояний, то суперпозиция будет свойственна и мозгу субъекта, наблюдающего эту стрелку. Состояние мозга будет описываться как суперпозиция состояния, в котором он воспринимает стрелку, направленную вверх, и состояния, в котором он воспринимает стрелку, направленную вниз. Ключевой шаг Эверетта состоит в предположении, что каждое из этих двух состояний должно связываться с отдельным наблюдателем. В результате такого измерения порождаются два наблюдателя. Один из них ощущает стрелку-«вверх», другой — стрелку-«вниз». Это значит, что *каждый из них будет ощущать дискретное состояние мира*.

Далее Эверетт показывает, что при такой схеме эти наблюдатели будут наделены большинством из тех свойств, которые мы ожидаем у них найти, и что отсюда можно извлечь большинство предсказаний квантово-механического исчисления. Нетрудно, к примеру, увидеть, что каждое из двух состояний, накладывающихся друг на друга, не будет иметь здесь доступа к другому состоянию такого рода, так что суперпозиция сознания не будет чем-либо выдавать себя ни в одном из состояний. Можно даже показать, что когда наблюдатель, производящий измерение, воспринимает другого наблюдателя, измеряющего ту же самую величину, воспринятые результаты измерения будут согласовываться друг с другом, так что мир будет казаться чем-то вполне когерентным. Одним словом, любой наблюдатель будет воспринимать мир во многом именно так, как мы ожидаем, хотя сам мир и находится в состоянии суперпозиции.

Эту интерпретацию не надо смешивать с интерпретацией *расщепляющихся миров*, согласно которой мир в буквальном смысле расщепляется на множество отдельных миров всякий раз, когда производится

измерение. Существует один мир, в котором стрелка направлена вверх, и совершенно обособленный мир, в котором стрелка направлена вниз. При таком рассмотрении эта позиция менее всего напоминает что-то простое. Начать с того, что она требует наличия нового и предельно фундаментального принципа для описания процесса «расщепления». Далеко не очевидно, когда именно должно осуществляться «расщепление» (проблема «измерения» возобновляется в новом виде), и крайне неясно, что должны представлять собой миры, возникающие в результате расщепления. Для осуществления буквального расщепления волновая функция должна «разделиться» на множество компонентов, но имеется множество вариантов декомпозиции волновой функции, и квантовая механика ничего не говорит о предпочтительности какого-либо базиса для подобной декомпозиции. Эта интерпретация кажется даже более сложной — и в большей степени имеющей *ad hoc* характер — чем разного рода «коллапсионные» интерпретации, и не существует серьезных оснований принимать ее.

Концепция расщепления часто приписывается Эверетту (во многом благодаря изложению работ Эверетта в трудах Девитта (DeWitt 1970, 1971)), но ее невозможно найти в его сочинениях. Собственная позиция Эверетта не вполне ясна, но гораздо более естественно толковать ее в моем ключе; и именно такую интерпретацию рекомендуют Альберт и Левер (Albert and Loewer 1988), а также Локвуд (Lockwood 1989). Согласно этой трактовке, никакого объективного «расщепления» не происходит. Скорее, волновая функция трансформируется в суперпозицию состояний, при которой состояния, находящиеся в отношении суперпозиции, лучше всего рассматривать в качестве компонентов одного-единственного мира. Концепция Эверетта иногда называется интерпретацией *множественных миров* (наводя тем самым на мысль о расщепляющихся мирах), но обсуждаемую мной позицию точнее назвать интерпретацией *одного большого мира*. Есть только один мир, но он больше, чем мы могли бы подумать⁷.

В соответствии с этой концепцией, если расщепление и происходит, то оно происходит только в умах наблюдателей. Поскольку суперпозиции влияют на состояние, в котором находится мозг субъекта, возникает множество обособленных сознаний, соответствующих компонентам суперпозиции. Каждое из них воспринимает отдельный дискретный мир, соответствующий тому типу мира, который воспринимаем мы — назовем его *мини-миром*, в противоположность *максимиру* суперпозиции. Реальный мир — это макси-мир, а мини-миры существуют лишь в сознаниях субъектов. Эверетт называет эту позицию интерпретацией *относительности состояния*: состояние мини-мира, в котором стрелки указывают на дискретные точки, считается

его состоянием лишь *относительно* спецификации наблюдателя. Объективное состояние мира — это суперпозиция.

Ключевой элемент этой концепции, однако, остается без анализа. Почему мы вправе отождествлять каждый компонент такого состояния мозга с отдельным наблюдателем? Почему вместо этого не существует один-единственный наблюдатель со спутанным суперпозиционным ментальным состоянием? И вообще, почему подобное некогерентное состояние мозга порождает сознания? Эверетт обходит эти ключевые вопросы. Более того, может показаться, что увязывая волновую функцию состояния мозга с множеством сознаний, каждое из которых воспринимает какое-то дискретное состояние, Эверетт незаконным образом отсылает к *предпочтительному базису*, так же как и интерпретация расщепляющихся миров. Волновая функция не является объективно делимой на какие-то компоненты, а может быть разложена самыми разными способами, в зависимости от того, какой базис для соответствующего векторного пространства мы выбираем. Зачастую в наших целях естественным может быть разложение волновой функции каким-то одним способом, опираясь на конкретный базис, но подобное разложение не отражает какого-либо объективного свойства волновой функции. Там, где состояние мозга может быть разложено на состояния «восприятия-верхнего» и «восприятия-нижнего», оно в равной степени может быть разложено на два таких состояния, с каждым из которых будет связаны спутанные восприятия. Создается впечатление, что, постулируя объективную декомпозицию, Эверетт выходит за те пределы, которые определяются уравнением Шредингера.

Ключевым элементом, упущенном Эвереттом при рассмотрении этих проблем, является анализ отношения между сознанием и телом. Эверетт допускает, что суперпозиционное состояние мозга будет связано со множеством отдельных субъектов опыта, но он даже не пытается обосновать это утверждение. Ясно, что этот вопрос существенным образом зависит от теории сознания. Сходное предположение высказывает Пенроуз:

В частности, я не понимаю, почему сознающее существо должно осознавать лишь «одну» из альтернатив линейной суперпозиции? Что есть такого в сознании, что требует признания невозможности «осознания» той заманчивой линейной комбинации мертвой и живой кошки? Думаю, что для согласования концепции множественных миров с тем, что мы действительно наблюдаем, нам потребовалась бы некая теория сознания (Penrose 1989, с. 296).

В самом деле, главный вопрос квантовой механики можно понять как вопрос об отношении между физическими процессами и опытом. Центральное место в квантовой механике занимает картина, на которой микроскопическая реальность описывается суперпозиционной

волновой функцией, изменяющейся в соответствии с уравнением Шредингера. Но в *опыте* мир дискретен. И главный вопрос: как это получается? Различные интерпретации дают разные ответы на него. Некоторые (к примеру, бомовская) отрицают первую посылку, полагая, что реальность дискретна даже на базовом уровне. Другие выдвигают базовые принципы (постулат измерения или закон коллапса в интерпретации ЖРВ) для опосредования перехода от того, чему свойственна суперпозиция, к дискретному. Ряд теорий (см. второй вариант) пытается объяснить, как суперпозиционные микроскопические состояния могут статистически порождать дискретную макроскопическую реальность. Все три последние стратегии — *непрямые*, старающиеся объяснить дискретность опыта объяснением дискретности макроскопической реальности, которая лежит в основании этого опыта.

Альтернативная стратегия предполагает *прямой* ответ на вопрос об опыте. Если мы всерьез принимаем тезис о примате уравнения Шредингера, то главный вопрос состоит в том, почему при *такой* физической структуре мира мы воспринимаем его именно *так*? Но это не что иное, как вопрос о том, как определенные физические структуры порождают опыт. Иначе говоря, это вопрос того типа, который я обсуждал на протяжении всей этой книги и на который должна отвечать теория сознания.

Если мы должны постулировать *ad hoc* теорию сознания для ответа на этот вопрос, то это существенно уменьшает привлекательность интерпретации Эверетта. Лучшей ее чертой всегда была простота, но новые и произвольные психофизические законы придали бы ей такой же *ad hoc* характер, который свойствен и бомовской интерпретации. Но если ответ на этот вопрос может дать *независимо мотивированная* теория сознания, то интерпретация Эверетта действительно покажется привлекательной.

Теория сознания, которую я отстаивал, может дать ответ на этот вопрос, причем надлежащего вида. Дело обстоит так, что эта теория *предсказывает*, что суперпозиционное состояние мозга должно быть связано с множеством раздельных субъектов дискретного опыта. Чтобы убедиться в этом, примем, что *максимальное феноменальное состояние* — это феноменальное состояние, характеризующее совокупный опыт субъекта в определенное время. *Максимальным же физическим состоянием* мы будем считать физическое состояние, в полной мере характеризующее внутреннее физическое состояние системы в определенное время. Для получения вывода достаточно признать следующий *принцип суперпозиции*:

Если теория предсказывает, что система в максимальном физическом состоянии P порождает связанное с ним максимальное феноменальное состояние E , то эта теория предсказывает, что система суперпозиции P с ортогональными физическими состояниями также будет порождать E .

Если этот принцип верен, то суперпозиция ортогональных физических состояний будет по меньшей мере порождать максимальные феноменальные состояния, которые отдельно порождались бы этими физическими состояниями. Именно этого и требует интерпретация Эверетта. Если мозг находится в суперпозиции состояний «восприятия верхнего» и «восприятия нижнего», то он породит хотя бы двух субъектов опыта, один из которых будет воспринимать стрелку, направленную вверх, а другой будет воспринимать стрелку, направленную вниз. (Разумеется, они будут двумя *раздельными* субъектами опыта, так как каждое из феноменальных состояний является максимальным феноменальным состоянием субъекта.) Это верно и в общем случае. Суперпозиция всегда будет порождать группу субъектов, что и требуется интерпретацией Эверетта.

Итак, нам нужно установить, что из очерченной мной теории вытекает принцип суперпозиции. Самый простой способ сделать это — обратиться к модели из главы 9 и, в частности, к тезису о том, что сознание возникает из имплементации надлежащего вычисления. Чтобы использовать его для установления данного принципа, нам надо установить, что, если вычисление имплементируется системой в максимальном физическом состоянии P , оно также имплементируется системой, представляющей собой суперпозицию P с ортогональными физическими состояниями.

Допустим, таким образом, что изначальная система (в максимальном физическом состоянии P) имплементирует вычисление C . Иначе говоря, между физическими подсостояниями системы и формальными подсостояниями C имеется такое соответствие, что каузальные отношения между физическими подсостояниями соответствуют формальным отношениям между формальными подсостояниями. Тогда одна из версий подобного соответствия будет также фундировать имплементацию C в системе с суперпозицией. Для некоего подсостояния S изначальной системы мы можем отыскать соответствующее подсостояние S' системы с суперпозицией, использовав очевидное отношение проекции: система с суперпозицией находится в S' , если система, получающаяся в результате ее проекции на гиперплоскость P , находится в S . Поскольку система с суперпозицией представляет собой суперпозицию P с ортогональными состояниями, из этого следует, что если изначальная система находится в S , система с суперпозицией находится в S' . А поскольку уравнение Шредингера линейно, то из этого также вытекает, что отношения перехода между подсостояниями S' в точности отражают отношения между изначальными подсостояниями S . Мы знаем, что эти отношения, в свою очередь, в точности отражают формальные отношения между подсостояниями C . Из этого следует, что система с суперпозицией также имплементирует C , что нам

и нужно было установить. В соответствии с принципом организационной инвариантности, если изначальная система порождает субъект опыта, система с суперпозицией тоже будет порождать качественно неотличимый от него субъект опыта.

Принцип суперпозиции можно доказывать и на основе использования двуспектрной теории информации и того довода, что релевантная информация, воплощенная в изначальном физическом состоянии, имеется также и в суперпозиции. Однако поскольку эта теория лишена однозначности, этот аргумент не столь ясен, как предыдущий, так что я не буду здесь разворачивать его. Важно то, что теория сознания, которую я частично разработал, так или иначе *предсказывает* результаты, требующиеся интерпретацией Эверетта. То есть она предсказывает, что, даже если мир находится в громадной суперпозиции, все равно будут существовать субъекты, воспринимающие его в опыте в качестве дискретного мира.

При отсутствии других проблем, из этого следует, что сочетание уравнения Шредингера с независимо мотивированной теорией сознания может предсказывать тот образ мира, который является нам в опыте. Иначе говоря, единственным физическим принципом, нужным для квантовой механики, оказывается уравнение Шредингера, а постулат измерения и другие базовые принципы — это всего лишь необязательный багаж.

Конечно, нам нужны и психофизические принципы, но они нужны в любом случае, и выясняется, что эти принципы, кажущиеся правдоподобными по независимым причинам, могут выполнять здесь необходимую работу. Это усиливает идею всерьез отнестися к интерпретации Эверетта.

5. Возражения против интерпретации Эверетта

Интерпретацию Эверетта часто критикуют, хотя вес критических замечаний неодинаков. Я распределию эти возражения на несколько групп.

Возражения, основанные на «расщеплении»

Многие возражения идут от интерпретации — верной или неверной — позиции Эверетта как концепции «расщепляющихся миров». Это понятно, учитывая, что ее нередко называют интерпретацией «множественных миров». К примеру, Белл (Bell 1976) говорит о неясности того, когда должно происходить «ветвящее» событие, неясности, связанной с неясностью понятия измерения, а также об от-

существии предпочтительного базиса для разделения на миры. Понятно, однако, что эти возражения неприменимы к данной интерпретации, не требующей ни объективного «ветвления», ни предпочтительного базиса. Хьюз (Hughes 1989) в сходном ключе возражает против «онтологического ливня», проливающегося в процессе расщепления, а Хили (Healy 1984) отмечает, что создание новых миров нарушает сохранение массы-энергии! Жаль, что «расщепляющая» интерпретация концепции Эверетта обрела такую большую популярность, так как ее очевидные трудности привели к тому, что более интересной интерпретации не уделялось того внимания, которого она заслуживала.

Некоторые из возражений против интерпретации расщепляющихся миров опираются на то, что она нуждается в предпочтительном базисе, но так обстоит дело и с возражениями против одномирной версии. В частности, возникает вопрос: почему сознания, связанные с суперпозиционным состоянием мозга, соответствуют только его декомпозиции сообразно некоему предпочтительному базису? Почему нет сознаний, порождаемых другими декомпозициями или вообще суперпозиционным состоянием в целом? Это правомочное возражение против версии самого Эверетта, которая, как кажется, требует такой канонической декомпозиции. Но оно не может быть выдвинуто против той версии, о которой я говорил, из которой следует, что суперпозиция порождает соответствующих субъектов дискретного опыта без необходимости постулирования предпочтительного базиса. И я не нуждался в допущении, что это *единственные* сознания, порождаемые подобной суперпозиционной системой.

Возникает вопрос: «А связаны ли с суперпозицией другие сознания?». Ответ: «возможно». Если принимается двуспектная теория информации, то мы уже знаем, что с низкоуровневыми процессами в такой системе могут быть сопряжены переживания. Субъекты опыта могут быть сопряжены с процессуальной структурой и при суперпозиции. Не исключено, что сознания сопряжены с другими декомпозициями системы или что большое суперпозиционное сознание сопряжено со всей суперпозиционной системой. Допущение таких сознаний зависит от деталей теории сознания, но трудно понять, как их существование может быть проблемой.

Можно было бы попробовать построить возражение против данной теории на тезисе о возможности суперпозиционных сознаний.

**Возражения,
связанные
с предпочтительным
базисом**

**А как насчет
суперпозиционных
сознаний?**

Возражение: почему *мое* сознание не находится в состоянии суперпозиции? Ответ: потому что я есть тот, кто я есть. Теория предсказывает существование несуперпозиционных сознаний, и мое сознание оказывается одним из них. Спрашивать, почему мое сознание не является одним из суперпозиционных сознаний, это все равно что спрашивать, почему я не мышь. Это просто часть нередуцируемой индексикальности моего существования. Мышиные сознания существуют, суперпозиционные сознания могут существовать, но мое сознание не является одним из них. Возражение: почему у меня нет никакого доступа к суперпозиционным сознаниям, к примеру, к воспоминаниям о суперпозиционных переживаниях? Ответ: данная теория предсказывает, что дискретные сознания, о которых идет речь, будут воспринимать мир всецело дискретным, и что у них не будет прямого доступа к другим частям суперпозиции. Все их воспоминания, к примеру, будут воспоминаниями о дискретных наблюдениях.

В любом случае, *интересными* сознаниями, связанными с суперпозиционной системой, будут, похоже, только привычные дискретные сознания. Эти сознания сложны и когерентны, и их опыт отражает структуру рациональных процессов. Любые другие сопряженные с ней сознания будут относительно некогерентны и по большей части лишены интересной структуры. Этот вывод подкрепляется «декогерационной» моделью Гелл-Манна и Хартла (Gell-Mann and Hartle 1990) и др. Согласно этой модели, интересная структура в волноподобной, сложной адаптивной системе обычно обнаруживается среди компонентов «естественной» декомпозиции; система естественным образом «декогерирует» в определенном направлении. В рациональных системах, таким образом, когерентная когнитивная структура может быть обнаружена только в компонентах этой естественной декомпозиции, и только они будут порождать сложные, когерентные сознания. Любые другие субъекты опыта в такой системе не будут причисляться к субъектам, которых можно называть личностями.

Возражения, основанные на личном тождестве

Целый кластер интуитивных беспокойств связан с тождеством наблюдателя. Возьмем сознание M_1 , которое, как я помню, существовало примерно в это время вчера. Сегодня в различных «ветвях» суперпозиции имеется множество сознаний, восходящих к тому сознанию. Мое сознание M_2 есть лишь одно из них. И я вполне мог бы спросить: почему я оказался здесь, а не в какой-то другой ветви? Как выражается Хофштадтер (Hofstadter 1985b):

Почему мое единое чувство самого себя распространяется по *этой* случайной ветви, а не по какой-то другой? Какой закон определяет случайный выбор, выделяющий ту ветвь, за которой, как я чувствую, я следую? Почему мое

чувство себя не сопровождает другого меня после расщепления и не идет другими путями? Что привязывает *яйность* к точке зрения этого тела, изменяющегося в этой ветви универсума в данный момент времени?

В ответ на это мы опять-таки должны сослаться на нередуцируемую индексикальность: мое сознание — *вот это*, и это вся история. Есть чувство, что за этим должно стоять что-то более глубокое, и что факт того, что вчерашнее сознание M_1 превратилось в сегодняшнее сознание M_2 , а не в какое-то другое, немаловажен для мира. Но если смотреть с объективной позиции, в этой ветви нет ничего особо привилегированного. Даже с позиции M_1 все сегодняшние сознания равнозначны. Ни одно из них не является единственным законным наследником M_1 : все они в равной степени несут «яйность» M_1 . M_2 кажется привилегированным только с *этой* позиции, с позиции M_2 (разумеется, другие мои двойники в суперпозиции ощущают то же самое по отношению к самим себе). Привилегированная роль M_2 есть лишь еще один индексикальный феномен, подобный тому факту, что я — Дэвид Чалмерс, а не Рольф Харрис. *Это* сознание — здесь, а не там. Индексикальные факты вообще загадочны, но этот факт не выбывает-ся из их ряда и не привносит в мир дополнительной асимметрии.

У нас есть сильная интуиция того, что личное тождество всегда должно быть связано с каким-то фактом: если из моего нынешнего состояния возникает множество сознаний, должен существовать некий факт, говорящий о том, какое из них будет *мной*. Но эта идея была подвергнута мощной критике Парфитом (Parfit 1984), убедительно доказывающим, что фактическая сторона личного тождества исчерпывается фактами о психологической континуальности, памяти и т. п. Если мы соглашаемся с этим анализом, то все завтрашние сознания в одинаковой степени могут претендовать на то, чтобы называться *мной*, и нет такого факта, который позволил бы различить их. В этом выводе есть что-то неприятное, превращающее определенный «поток» личного тождества в иллюзию, но анализ Парфита дает основание полагать, что такой поток всегда и был ею.

Наиболее существенным возражением против интерпретации Эверетта является то, что с ее помощью нельзя осмыслить *вероятности*, поставляемые постулатом измерения⁸. В каком-то случае постулат измерения может говорить нам, что при проведении определенного измерения шанс обнаружения стрелки, указющей вверх, равен 0,9, а шанс обнаружения стрелки, указющей вниз, равен 0,1. По интерпретации Эверетта, на деле происходит то, что как стрелка, так и состояние мозга наблюдателя оказываются в суперпозиции, из чего получается (по меньшей

Интерпретация вероятностей

мере) два субъекта опыта. Один из них ощущает стрелку, указывающую вверх, другой — указывающую вниз. Точно такую же картину мы наблюдали бы, если бы вероятности были 50:50. Конечно, в случае 90:10 амплитуда суперпозиционной волновой функции по большей части сконцентрирована в области «верхнего» состояния мозга, но как это связано с вероятностями?

Эверетт решает этот вопрос, вводя меру на пространстве наблюдателей, соответствующую вероятностям, заданным постулатом измерения (то есть соответствующую квадрату амплитуды соответствующей части волновой функции). Используя эту меру, он доказывает, что в пределе большинство наблюдателей (то есть подмножество наблюдателей с мерой 1) будет обладать воспоминаниями о наблюдениях, согласующимися с частотами, предсказанными вероятностями в постулате измерения. К примеру, среди наблюдателей, проводящих измерение, подобное тому, которое было описано выше множество раз, большинство будет вспоминать, что они обнаружили стрелку, указывающую вверх, 90 процентов времени, вниз — 10 процентов времени. Так мы и находим роль для вероятностей. Возникает, однако, вопрос: как обосновать эту меру на пространстве наблюдателей? Если бы мы измеряли это пространство иначе, могли бы возникнуть весьма отличные от тех вероятности. К примеру, если бы мы назначали равные меры всякий раз, когда два субъекта возникают из суперпозиции (независимо от амплитуды), большинство наблюдателей вспоминали бы пропорцию «верхнего» и «нижнего» равной 50:50. Ни уравнение Шредингера, ни психофизические законы не гарантируют «верность» той или иной меры.

Альберт и Левер (Albert and Loewer 1988) решают эту проблему, отказываясь от мер. Взамен они постулируют более радикальные психофизические законы, по которым с каждым состоянием мозга сопряжено бесконечное множество сознаний. С каждым сознанием, постулированным предыдущей концепцией, эта теория соотносит бесконечное множество качественно неотличимых сознаний. Далее, там, где теория Эверетта предсказывает, что сознание разойдется на два сознания, эта теория говорит, что любое сознание пойдет одним или другим путем, с вероятностью, заданной постулатом измерения. Поэтому, если мы берем любое сознание, связанное с состоянием мозга перед вышеупомянутым измерением, то можно будет говорить о 90-процентной вероятности, что оно трансформируется в состояние «восприятия верхнего», и 10-процентной вероятности, что оно трансформируется в состояние «восприятия нижнего». Так сохраняются вероятностные предсказания квантово-механического исчисления.

Ясно, что здесь уже нет той простоты. Новые психофизические законы лишены какой-либо независимой мотивации, и данная теория нуждается также в дополнительных «интрапсихических» законах, управ-

ляющих *трансформацией* сознаний. Вводя эти *ad hoc* постулаты, теория приносит в жертву ряд ключевых достоинств интерпретации Эверетта. Можно также попробовать показать проблематичность этих интра-психических законов, поскольку они постулируют фундаментальные нередуцируемые факты относительно тождества личности во времени. Трудно понять, что делать с этими фактами. Так, их признание потребовало бы отказа от парфитовского анализа личного тождества. У них нет даже естественной супервентности на физических фактах, и это усложняет метафизическую картину. Этую интерпретацию нельзя сбрасывать со счетов, но ее цена велика.

Альтернатива состоит в том, чтобы обойтись без дополнительных ухищрений и посмотреть, нельзя ли сохранить вероятности как-то иначе. Заманчиво истолковать это как проблему индексикальности. Почему из всех мест волновой функции, в которых я мог бы оказаться, я оказался там, где мои воспоминания согласуются с предсказаниями исчисления? Одна из возможностей состоит попросту в том, чтобы признать это элементарным индексикальным фактом: в этой области есть *какие-то* сознания и мне случается быть одним из них. Это, однако, кажется неудовлетворительным, так как примечательная регулярность вычисления оказывается удивительным совпадением. Нам нужно показать, что это не так.

Даже если мы признаем, что мое попадание *сюда* — счастливое совпадение, тут неявно присутствует идея некоей меры на пространстве сознаний. Суть в том, что антецедентно более вероятным является то, что я должен оказаться сознанием одного, а не другого типа, возможно, из-за относительной избыточности тех классов. Такие неявные меры присутствуют во многих наших рассуждениях о мире. Когда я индуктивно рассуждаю, отталкиваясь от каких-то данных и делая тот или иной вывод, я знаю, что для *некоторых* наблюдателей в сходном эпистемическом положении вывод не будет верным, но я допускаю, что для *большинства* таких наблюдателей он будет верным, даже если в каждом из классов имеется бесконечное множество членов. То есть я допускаю, что антецедентно более вероятно, что я окажусь скорее в одном, чем в другом классе. Подобное рассуждение неявно предполагает некую меру на пространстве сознаний.

Так что не исключено, что мы можем обосновать вероятности, эксплицитно вводя подобную меру. Наверняка большая часть амплитуды волновой функции сосредоточена там, где воспоминания наблюдателей соответствуют предсказания исчисления. Возможно, то, что мое сознание должно оказаться в высокоамплитудной, а не в низкоамплитудной области, является более вероятным. В частности, если мы допустим, что антецедентная вероятность того, что я окажусь одним, а не другим сознанием, пропорциональна квадрату амплитуды

соответствующей части волновой функции, то из этого следует, что частотность моих воспоминаний почти наверняка будет предсказана квантово-механическим исчислением.

Но почему объективно соответствует эта мера? Надо ли признавать ее в качестве базового факта о распределении самостей? Можно ли каким-то образом обосновать ее в качестве канонической меры на этом пространстве? Это трудные вопросы, тесно связанные с тайной самой индексикальности — почему я оказываюсь *этой*, а не какой-то другой личностью? Это одна из главных тайн, и совершенно неясно, как можно разрешить ее. Тем не менее идея меры на пространстве сознаний не лишена перспектив и может даже понадобиться для других целей, таких как обоснование индукции. Пока же интерпретация вероятностей остается самой серьезной трудностью, с которой сталкивается интерпретация Эверетта.

6. Заключение

Надо признать, что в интерпретацию Эверетта практически невозможно поверить. Она постулирует, что мир намного больше, чем мы когда-либо думали. Согласно этой интерпретации, на деле мир представляет собой громадную суперпозицию состояний, изменяющихся различными путями с начала времен, и в опыте нам дана лишь крохотная его часть. Она также постулирует неопределенность моего будущего: через минуту появится множество сознаний, с равным правом могущих претендовать на то, чтобы считаться *мной*. С тех пор, как я написал предыдущее предложение, прошла минута; и кто знает, чем заняты сейчас эти другие сознания?

Ясно, с другой стороны, что *все* ныне существующие интерпретации квантовой механики в известном отношении безумны. В этом фундаментальный парадокс квантовой механики. Три главных кандидата на роль интерпретации — это, возможно, интерпретация Вигнера, согласно которой сознание производит коллапс, бомовская интерпретация, допускающая нелокальные скрытые параметры, и интерпретация Эверетта. Если брать эти интерпретации, то из вигнеровской вытекает, что макроскопические объекты часто находятся в суперпозиции — до тех пор, пока случайный взгляд наблюдателя не приводит к их коллапсу. Интерпретация Бома ведет к представлению о том, что траектория каждой частицы в универсуме зависит от состояния каждой из имеющихся в нем частиц. А из концепции Эверетта вытекает, что мир гораздо больше, чем мы могли бы подумать.

Из них концепция Бома, возможно, наименее безумна, Эверетта — наиболее, а вигнеровская в этом плане находится где-то между ними.

Но если выстраивать их по теоретической ценности, то все будет наоборот. Концепция Бома неудовлетворительна из-за ее сложности и искусственности. Концепция Вигнера вполне элегантна, и два ее базовых динамических закона отражают квантово-механическое исчисление, при условии проработки всех ее деталей. Но самой простой оказывается концепция Эверетта. Она постулирует лишь уравнение Шредингера, принцип, признающийся всеми интерпретациями квантовой механики. К ее достоинствам, отсутствующим у других, относится то, что это полностью локальная теория, а также то, что она напрямую совместима с теорией относительности.

Стоит также отметить, что континтуитивные моменты интерпретации Эверетта содержатся и в двух других интерпретациях. Согласно концепции Вигнера, мы должны признавать, что мир менялся по образцу гигантской суперпозиции Эверетта — и что в нем могли быть суперпозиционные звезды и камни, если и не суперпозиционные кошки — по крайней мере, до тех пор, пока в нем не эволюционировало первое сознательное существо, произведшее коллапс волновой функции. Согласно концепции Бома, неколлапсированная волновая функция Эверетта присутствует в качестве «волны-пилота», определяющей положение различных частиц. Вся структура, имеющаяся в других компонентах, таким образом, наличествует в состоянии мира, хотя ее большая часть нерелевантна для эволюции частиц. Поскольку эти концепции также отводят центральные роли неколлапсированной волновой функции, можно было бы попробовать доказать, что это снижает относительную неправдоподобность концепции Эверетта.

Разумеется, никогда нельзя исключать возможности создания новой теории, которая превзойдет перечисленные концепции по своей правдоподобности и теоретической ценности. Но это не кажется столь уж вероятным. Полное отсутствие экспериментальных аномалий находит на мысль, что квантово-механическое исчисление сохранится как предсказательная теория. Если это так, то мы не можем ожидать решения этой проблемы под влиянием эмпирических факторов. Не исключено, что новая и усовершенствованная интерпретация могла бы возникнуть вследствие концептуальных разработок, но может оказаться так, что самые перспективные ниши в концептуальном пространстве к настоящему времени уже освоены. В таком случае мы можем застрять в спектре возможностей, похожем на тот, что уже имеется — с существенными уточнениями, быть может, но с плюсами и минусами, сходными с нынешними по своим качественным параметрам. И если говорить об этих возможностях, то интерпретация Эверетта во многих отношениях кажется наиболее привлекательной, хотя в то же время верно, что труднее всего согласиться именно с ней.

В этой работе я отстаивал ряд контринтуитивных концепций. Долгое время я сопротивлялся дуализму сознания и тела, но теперь я одобряю его, и не только в качестве единственной позиции, которую можно защитить, но и удовлетворительной самой по себе. Нельзя исключать, что я где-то запутался или проглядел новую, радикальную возможность; но я не чувствую дискомфорта, когда говорю, что дуализм скорее всего верен. Я также рассуждал о возможности панпсихизма. Подобно дуализму сознания и тела, он поначалу контринтуитивен, но эта контринтуитивность исчезает со временем. Я не уверен, истинна или нет эта концепция, но она по крайней мере интеллектуально привлекательна, а если поразмыслить над ней, то и не столь безумна, чтобы с ней нельзя было согласиться.

Безумность интерпретации Эверетта — величина иного порядка. Я легко допускаю, что это наиболее интеллектуально привлекательная интерпретация квантовой механики, но признаюсь, что не могу искренне поверить в нее. Если бы Бог вынудил меня поставить на кон мою жизнь в зависимости от истинности или ложности концепций, которые я отстаивал, я достаточно уверенно сказал бы о фундаментальности опыта и с гораздо меньшей уверенностью — о повсеместности опыта. Что же касается интерпретации Эверетта, я разрывался бы на части и, возможно, в конце концов мне не хватило бы смелости сделать ставку на нее⁹. Быть может, она просто слишком уж необычна, чтобы в нее можно было поверить. Неясно тем не менее, можно ли в конечном счете придавать большое значение этим интуитивным сомнениям. Эта концепция проста и элегантна, и она предсказывает существование наблюдателей, которые видят мир именно так, как его вижу я. Разве этого недостаточно? Не исключено, что мы никогда не сможем эмоционально одобрить ее, но мы, по крайней мере, должны всерьез относиться к тому, что она может оказаться истинной.

Примечания

Глава 1

1. См. (Nagel 1974). Обычно считается, что в философском контексте это выражение впервые использовал Фаррел (Farrell 1950). См. также (Sprigge 1971).
2. Разные авторы по-разному используют термин «квалиа». Я использую его, как мне кажется, в стандартном смысле для указания на те свойства ментальных состояний, которые типологизируют эти состояния через то, каково это — иметь их. Употребляя этот термин, я не собираюсь предрешать ход дальнейшего обсуждения таких, к примеру, вопросов, как вопрос о том, является ли знание о квалиа безошибочным, являются ли они интенциональными свойствами и т. д. Квалиа могут быть свойствами как «внутренних» ментальных состояний, так и ощущений. Зачастую удобно выражаться так, как если бы квалиа были свойствами, напрямую реализуемыми субъектом, а не ментальными состояниями этого субъекта; эта практика не приносит вреда, и ее оправдывает тот факт, что квалиа по самой сути соответствуют типам ментальных состояний.
3. Такие выражения, как «красное ощущение», «зеленый опыт» и т. п., будут использоваться мной на протяжении всей книги. Разумеется, этим я не хочу сказать, что опытные переживания реализуют тот же тип цветовых свойств, который реализуется объектами (яблоками, деревьями) во внешнем мире. Подобные высказывания всегда можно переформулировать: «опыт того типа, который обычно имеется у меня (в актуальном мире) при взгляде на красные объекты» и т. п., но краткая формулировка выглядит более естественной.
4. Положите 12 стаканов сухой спаржевой фасоли в кипящую соленую (4 столовых ложки соли) воду. Когда фасоль размягчится, поместите ее в холодную воду. Смешайте 2 нарезанных кубиками красных перца, 5 таких же зеленых перцев, 2 нарезанных кубиками большие луковицы, 3 стакана изюма и пучок нарезанной кинзы, добавьте в заправку для салата 1,5 стакана кукурузного масла, 0,75 стакана винного уксуса, 4 столовых ложки сахара, 1 столовую ложку соли, 4 столовых ложки черного перца, 5 столовых ложек порошка карри и 0,5 столовой ложки черешков гвоздики. Подавайте охлажденным. Спасибо Лизе Томас и Encore Café.
5. Богатый материал для размышлений о видах конкретных переживаний см. в *A Natural History of the Senses* Экермана (Ackerman 1990). Эта книга может надолго затянуть тех, кто увлечен своим сознательным опытом.

6. Интересно, что Декарт нередко исключал ощущения из категории ментального, связывая их с телом, так что, с другой стороны, не всякое феноменальное состояние (по крайней мере в моей трактовке этого понятия) можно было бы счесть ментальным.
7. Эта обычная интерпретация Райла не учитывает тонкостей его теории, но это хотя бы полезная фикция.
8. Существуют и другие виды функционализма, к примеру патнэмовский (Putnam 1960). Я не рассматриваю их здесь, так как они предлагались в качестве эмпирических гипотез, а не анализов ментальных понятий.
9. Нагель (Nagel 1970) выдвигает сходное возражение против Армстронга, отсылая к проблеме других сознаний.
10. Аргумент Серла основан на тезисе о том, что при отсутствии сознания невозможно было бы объяснить «аспектуальную форму», демонстрируемую интенциональностью, как в случае, когда кто-нибудь убежден в чем-то относительно Венеры в ее «утренне-звездном» аспекте, но не в ее «вечерне-звездном» аспекте. Для меня не очевидно, что аспектуальную форму нельзя было бы объяснить иными способами; можно было бы даже попробовать доказать, что современные компьютеры демонстрируют что-то вроде нее, сохраняя информацию обо мне в одном «аспекте», но не в другом (к примеру, в аспекте моего имени, но не моего номера социального обеспечения). На это можно было бы возразить, что в данном случае идет речь только о «как бы» аспектуальной форме, поскольку подлинной аспектуальной формой является только феноменальная аспектуальная форма; но такой шаг, похоже, привел бы к тривиализации данного аргумента.
11. Указания на эту позицию можно вычитать из некоторых замечаний Локвуда (Lockwood 1989) и Нагеля (Nagel 1986), хотя я не уверен, что кто-то из них в действительности одобряет именно такое решение.
12. Можно было бы сослаться на аргументы, вроде тех, что выдвигает Крипке (Kripke 1982), призванные показать, что содержание убеждения не определяется психологическими и феноменальными свойствами. Эти аргументы спорны, но в любом случае стоит отметить, что вывод, к которому они подводят, состоит не в том, что такое содержание есть далее не редуцируемый элемент ментального, а скорее в том, что само содержание является чем-то неопределенным. На деле суть происходящего в том, что соображения, вроде тех, что высказаны в тексте, дают нам серьезное основание считать, что не существует никакого третьего, допускающего независимые изменения, аспекта ментального; поэтому любые вопросы, которые не могут быть уложены с привлечением первых двух элементов, вообще не могут быть уложены.
13. Это «тематически нейтральный» анализ конкретных ментальных понятий, в чем-то сходный с анализами Плейса (Place 1956) и Смарта (Smart 1959). Ощущение апельсина как опыт — это, грубо говоря, ощущение такого рода, которое обычно порождается апельсинами. Плейс: «Характеризуя остаточный образ в качестве зеленого... мы говорим, что мы имеем опыт такого рода, который, как правило, имеется у нас тогда, когда мы смотрим

на зеленое пятно, и который мы научились описывать подобным образом» (с. 49); Смарт: «Когда кто-то говорит: „Я вижу желтовато-оранжевый остаточный образ“, он говорит примерно следующее: „Происходит нечто подобное тому, что происходит, когда у меня открыты глаза, я не сплю и передо мной лежит хорошо освещенный апельсин“» (с. 150). Однако из-за присутствия непроанализированного понятия «опыта» этот анализ недостаточен для прямого отождествления феноменальных и физических состояний, как это предлагали сделать Плейс и Смарт. Смарт обходит эту проблему, выводя «опыт» за пределы своего анализа, вставляя взамен двусмысленную фразу «происходит нечто». Если эта фраза трактуется достаточно широко, чтобы покрыть состояния любого типа, то анализ оказывается неадекватным, а если узко — как некий опыт, то анализ будет ближе к делу, но недостаточным для нужного вывода.

14. Джекендоф проводит различие между «феноменальным ментальным» и «вычислительным ментальным». Это различие очень сходно с моим различием феноменального и психологического, хотя я не хотел бы предрешать вопрос о том, являются ли психологические процессы вычислительными по своему характеру.
15. Нилкин (Nelkin 1989) проводит различие между CN (сознанием в «нагелевском» смысле), C1 (первопорядковым состоянием обработки информации) и C2 (второпорядковым прямым невыводным доступом к другим сознательным состояниям). В другой статье (Nelkin 1993) он проводит сходное различие между феноменальностью, интенциональностью и интроспективностью. Байсия (Bisiach 1988) различает C1 (феноменальный опыт) и C2 (доступ частей или процессов системы к другим ее частям или процессам). Нетсоулас (Natsoulas 1978) выделяет множество смыслов термина «сознание». Деннет (Dennett 1969) проводит различие между двумя видами «осведомленности»; первый связан с вербальными отчетами, второй — с контролем за поведением вообще, хотя ни одно из этих понятий не может быть названо с очевидностью феноменальным.
16. Розенталь четко отделяет сознание от «сенсорного качества» и говорит, что его теория касается только первого из них, что могло бы означать, что он не обсуждает феноменальные аспекты. Однако он также утверждает, что состояние сознательно, когда можно говорить о том, каково это — находится в этом состоянии, и это значит, что предметом его рассмотрения является все же феноменальное сознание. Однако в концепции Розенталя мы почти не найдем указаний на то, как можно было бы объяснить феноменальное сознание. Почему существование высокоуровневой мысли о каком-то состоянии ведет к наличию чего-то, что выражает, каково это — находится в данном состоянии? Помимо рассуждений о том, что два этих феномена, похоже, существуют на практике, Розенталь не предлагает никакого ответа на этот вопрос.
17. Исключение составляет психофизика, которая, похоже, может прояснить различные черты сознательного опыта, пусть ей и не удается полностью объяснить его. В главе 6 я еще вернусь к этому вопросу.

Глава 2

1. Идея супервентности была введена Муром (Moore 1922). В печатном виде этот термин был впервые упомянут Хейром (Hare 1952). Дэвидсон (Davidson 1970) был первым, кто применил это понятие к проблеме соотношения ментального и физического. Позже Ким (Kim 1978, 1984, 1993), Хорган (Horgan 1982, 1984c, 1993), Хелман и Томпсон (Hellman and Thompson 1975) и другие разработали детальную теорию супервентности.
2. Я использую выражение «A-факт» для краткой записи «реализации A-свойства». Обращение к фактам делает обсуждение менее тяжеловесным, хотя все рассуждения о фактах и их отношениях в конечном счете могут быть выражены в терминах схематики совместной реализации свойств: при необходимости я детализирую это в примечаниях. В частности, следует отметить, что тождество индивида, реализующего A-свойство, не имеет значения для A-факта в моем понимании: важна лишь реализация свойства. Если бы тождество индивида частично конституировало A-факт, то любой A-факт влек бы за собой факты о сущностных свойствах этого индивида, и тогда дефиниция супервентности вела бы к континтуитивным выводам.
3. Я допускаю, возможно, несколько искусственно, что индивиды имеют четкие пространственно-временные границы, так что их физические свойства — это свойства, реализованные в таком регионе пространства-времени. Если в контексте рассуждений о локальной супервентности мы должны были бы счесть физически тождественными пространственно-раздельные объекты, то нам надо было бы исключить из супервентностной базы любые свойства, касающиеся абсолютного пространственно-временного положения (хотя можно было бы избежать необходимости обращения к пространственно раздельным объектам, рассматривая всего лишь возможные объекты, находящиеся в том же самом положении). К тому же я всегда говорю так, будто низкоуровневые и высокоуровневые свойства реализуются индивидами одного типа, так что, к примеру, стол реализует микрофизические свойства — будучи охарактеризован через распределение этих свойств. Строго говоря, более точно, возможно, было бы сказать, что микрофизические свойства реализуются только микрофизическими сущностями, но мои формулировки упрощают картину. В любом случае самые важные проблемы, с которыми мы будем иметь дело, связаны не с локальной, а с глобальной супервентностью.
4. Можно по-разному уточнять, что означает тождество миров относительно некоего множества свойств; это не имеет большого значения для нашего обсуждения. Быть может, лучше всего было бы сказать, что два мира тождественны относительно их A-свойств, если существует взаимно-однозначное соответствие между классами индивидов, реализующих A-свойства в обоих мирах, при котором каждый из двух соответствующих индивидов реализует те же A-свойства. Говоря о глобальной супервентности, мы в таком случае должны допустить, что соответствия, вследствие которых два мира рассматриваются как A-неразличимые, так и B-неразличимые, совместимы друг с другом; то есть ни один индивид не соотнесен

с одним коррелятом при А-соответствии, с другим — при В-соответствии. Дефиниция глобальной супервентности обретает следующий вид: любые два А-тождественных мира (при некоем соотнесении) В-тождественны (в экстенсионале этого соотнесения).

Более привычный способ сделать это состоит в допущении, что А-тождественные миры должны включать в себя совершенно одинаковых индивидов, реализующих те же самые свойства, однако, как указывает Маклафлин (McLaughlin 1995), это неоправданно сильное допущение: оно приводит к тому, что кардинальное число мира и сущностные свойства индивидов с неизбежностью оказываются супервентными вообще на любых свойствах. Предлагаемая мной дефиниция обходит эту проблему, обеспечивая то, чтобы отношение детерминации заключало в себе исключительную схематику А-свойств и ничего более.

5. С одним исключением: Бог не мог бы создать мир, который не был создан Богом, хотя мир, не созданный Богом, судя по всему, является чем-то логически возможным! Я проигнорирую такого рода усложнения.
6. Отношение такого рода возможности к выводимости в формальных системах весьма неочевидно. Можно попробовать доказать, что аксиомы и правила вывода конкретных формальных систем обосновываются именно в терминах уже имеющихся у нас понятий логической возможности и необходимости.
7. Интуитивное понятие естественной возможности концептуально предшествует дефиниции в терминах законов природы: регулярность может считаться законом именно тогда, когда она имеется во всех ситуациях, которые могли бы сложиться в природе, то есть во всех естественно возможных в интуитивном смысле ситуациях. Иногда это формулируется следующим образом: чтобы нечто могло считаться законом, это нечто должно иметься не только в актуальных, но и в контрфактических ситуациях, и для определения релевантных контрфактических ситуаций требуется более фундаментальное понятие естественной возможности.
8. Для отсылки к необходимости такого рода нередко используются также термины «физическая необходимость» и «каузальная необходимость», но я не хочу предрешать вопрос о том, являются ли все законы природы физическими или каузальными.
9. Важное различие логической и естественной супервентности нередко замалчивается или игнорируется в литературе, зачастую не конкретизирующей модальность отношений супервентности. Естественная (или нормологическая) супервентность обсуждается ван Клевом (van Cleve 1990), использующим ее для объяснения одной из разновидностей эмерджентности. Сигер (Seager) проводит сходное различие между тем, что он называет конститтивной и коррелятивной супервентностью. Они напрямую соответствуют логической и естественной супервентности, хотя Сигер анализирует данные понятия несколько иначе.
10. Слабая супервентность требует лишь, чтобы в мире — но не в мирах — не было «В-различий без А-различий» (детали см. в (Kim 1984)). Модальная слабость этого отношения делает его слишком слабым для боль-

шинства целей. В лучшем случае оно может пригодиться для выражения концептуальных ограничений нефактуальных рассуждений (как в (Hague 1984)), хотя, как указывает Хорган (Horgan 1993), даже эти ограничения, похоже, предполагают кроссмировую зависимость. Сигер (Seager 1988) обращается к слабой супервентности для выражения некоей систематической внутримировой корреляции, не являющейся строго необходимой, но естественная супервентность является гораздо лучшим средством для решения этой задачи.

11. Глобальная естественная супервентность без локализованной регулярности — когерентное понятие, если исходить из неюмистской концепции законов, хотя, быть может, и нет, если опираться на юмистскую концепцию (основанную на понятии регулярности). Однако, даже если исходить из неюмистской концепции, трудно понять, какого рода данные могли бы свидетельствовать о существовании подобного отношения.
12. Хорган (Horgan 1982), Джексон (Jackson 1994) и Льюис (Lewis 1983b) обсуждают сходную проблему в контексте определения материализма.
13. Измененная дефиниция может быть более точно сформулирована в духе прим. 4. Пусть $B(W)$ — класс индивидов с В-свойствами в мире W . Можно сказать, что W' В-превосходит W , если имеется взаимно-однозначное соответствие $f: B(W) \rightarrow B(W')$ (то есть между $B(W)$ и подмножеством $B(W')$), такое что для любого $a \in B(W)$, $f(a)$ реализует каждое В-свойство, которое реализует a . В таком случае В-свойства логически супервентны на А-свойствах в W , если любой мир, А-неотличимый от W , оказывается В-превосходящим W , где релевантные В-соответствия опять-таки ограничиваются экстенсионалами А-соответствий.

Чтобы убедиться в необходимости этого ограничения, вообразим, что в нашем мире имеется исчислимое бесконечное множество психологически тождественных умов, один из которых реализован в эктоплазме, тогда как другие реализованы физически. Интуитивно ясно, что в этом мире психологическое не является супервентным на физическом, но всякий физически неотличимый мир будет психологически превосходящим. Хотя мы ожидаем, что любой такой мир без эктоплазмы будет свидетельствовать против супервентности, между психологическими аспектами того мира и нашего мира существует взаимно-однозначное соответствие. Проблема в том, что это соответствие не связано с соблюдением физического соответствия, соотнося физическую сущность с эктоплазматической сущностью; поэтому нам нужно дополнительное ограничение.

14. Для целей данной дефиниции отношение содержания миров друг в друге может рассматриваться как что-то изначальное. Льюис (Lewis 1983a) и Джексон (Jackson 1993) отмечали бесплодность продолжения анализа такого рода понятия до бесконечности. Что-то надо принимать в качестве исходного, и отношение содержания представляется не менее ясным, чем какие бы то ни было другие отношения. Кто-то мог бы предпочесть рассуждения не о мирах, содержащих W в качестве собственной части, а о мирах, содержащих качественный дубликат W в качестве своей собственной части; это работало бы не менее эффективно.

15. Обратим внимание, что, согласно данной дефиниции, существуют позитивные факты, не являющиеся реализациями позитивных свойств. Возьмем, к примеру, реализацию свойства быть бездетным-или-кенгуру. Быть может, надо более строго определить позитивный факт как реализацию позитивного свойства, но, насколько я могу судить, более слабая дефиниция не дает каких-либо нежелательных эффектов.
16. Тезис о логической супервентности свойств в нашем мире, по-видимому, должен отсылать к закономерности. Если бы дело обстояло так, что при несколько ином стечении обстоятельств в нашем мире (быть может, при несколько иных случайных флуктуациях) в нем жили бы нефизические ангелы, несмотря на соблюдение законов природы, то логическая супервентность биологических свойств на физических свойствах была бы просто исторической случайностью. Более сильный и интересный метафизический тезис получается при замене в дефинициях логической супервентности отсылок к нашему миру и актуальным индивидам на отсылку к естественно возможным мирам и индивидам. Это исключает сценарии, подобные упомянутому выше. В качестве бонуса это позволяет нам определять, являются ли логически супервентными такие нереализованные свойства, как быть небоскребом высотой в милю. Согласно предыдущей дефиниции, все подобные свойства бессодержательно супервентны.

Это дает следующую дефиницию: В-свойства логически супервентны на А-свойствах, если и только если для любой естественно возможной ситуации X и любой логически возможной ситуации Y , если X и Y А-неотличимы, то Y В-превосходит X (с обычным ограничением). Или более кратко: для любой естественно возможной ситуации В-факты об этой ситуации выводимы из А-фактов. Эта модификация существенно не влияет на обсуждение в тексте, и поэтому я опускаю ее ради простоты. Но это обсуждение несложно переформулировать в терминах более строгой дефиниции путем простой замены по всему тексту соответствующих отсылок к «нашему миру» на отсылки ко «всем естественно возможным мирам», как правило, без каких бы то ни было потерь в плане убедительности высказываемых в этих местах тезисов.

Итог напоминает стандартную дефиницию «сильной» локальной супервентности (Kim 1984), в которой присутствует два модальных оператора. Согласно этой дефиниции, В-свойства супервентны на А-свойствах, если необходимо для всякого X и всякого В-свойства F , что если x имеет F , то существует такое А-свойство G , что x имеет G , и необходимо, что если какой-либо y имеет G , то он имеет F . (А-свойства G при необходимости могут мыслиться как комплексы более простых А-свойств.) Проблема с ангелами делает очевидным, что первый модальный оператор должен всегда пониматься в смысле естественной необходимости, даже если второй выражает логическую. Не столь гладко обстоит дело со стандартной дефиницией глобальной супервентности (А-неотличимые миры являются В-неотличимыми), и она нуждается в модификации в предложенном мной ключе. При необходимости можно дать и параллельную дефиницию «метафизической» супервентности. Разумеется, проблема с ангелами не возникает в случае естественной супервентности, так как нет основа-

- ний считать эктоплазму естественно возможной, и поэтому прямолинейная дефиниция естественной супервентности будет удовлетворительной.
17. Не исключено, что мы должны использовать более сильную дефиницию логической супервентности, и тогда материализм истинен, если все позитивные факты о всех естественно возможных мирах вытекают из физических фактов об этих мирах. Возьмем мир без эктоплазмы, в котором нефизическая эктоплазма является тем не менее естественной возможностью, — быть может, он возник бы при несколько иных случайных флуктуациях. Кажется, что мы вправе сказать, что в таком мире материализм ложен или, по крайней мере, истинен лишь в слабом смысле.
 18. Для этой эквивалентности мы нуждаемся в правдоподобном принципе, что если мир *A* есть собственная часть мира *B*, то ряд позитивных фактов, верных в *B*, неверны в *A*, то есть существует такой факт, который является верным в *B* и во всех больших, чем *B*, мирах, но при этом неверен в *A*.
 19. Сказанное также приблизительно соответствует дефинициям Хоргана (Horgan 1982) и Льюиса (Lewis 1983b), но, в отличие от последних, не опирается на несколько туманное понятие «чуждого свойства» для исключения эктоплазматических миров из списка релевантных возможных миров.
 20. В философской литературе множественная реализуемость нередко рассматривается в качестве главного препятствия на пути «редукции», однако, как доказывает Брукс (Brooks 1994), она, похоже, имеет весьма отдаленное отношение к тому, как редуктивные объяснения используются в науке. К примеру, биологические феномены, такие как крылья, могут быть реализованы множеством разных способов, но биологи все равно выдвигают редуктивные объяснения. Более того, как показали Уилсон (Wilson 1985) и Черчленд (Churchland 1986), многие физические феномены, нередко считающиеся парадигмами редуцируемости (к примеру, температура), в действительности являются множественно реализуемыми.
 21. Кто-то мог бы сказать, что мы не должны говорить, к чему отсыпала бы «вода», если бы XYZ-мир оказался актуальным, поскольку в этом случае слово, звучащее как «вода», было бы совершенно другим словом! Если это вызывает беспокойство, мы можем просто говорить о том, к чему отсыпало бы слово с таким же звучанием; или, еще лучше, можно мыслить эти сценарии в качестве эпистемических возможностей (в широком смысле), а кондиционалы — в качестве эпистемических кондиционалов, избавляясь тем самым от беспокойств по поводу сущностных свойств слов. В любом случае суть дела, а именно зависимость референции актуального мира от того, каким оказывается этот мир, остается незамутненной независимо от того, каким образом мы будем описывать данный сценарий. В тексте я по большей части проигнорирую эту тонкость.
 22. Не все убеждены в правильности тезиса Крипке и Патнэма о том, что вода с необходимостью есть H_2O и что XYZ — это не вода; сомнения на этот счет высказывает, к примеру, Льюис (Lewis 1994). Ясно, что наша лингвистическая практика, похоже, осталась бы по большей части неизменной, даже если бы мы использовали термин «вода» для указания на водянистую материю в контрафактических мирах. Тот факт, что этот вопрос несложно

развернуть как в ту, так и в другую сторону, наводит на мысль, что при объяснении таких феноменов, как вода, на самом деле не столь уж важны моменты, связанные с вторичным интенсионалом. Действительно, тот факт, что мы всегда можем заменить «воду» такими терминами, как «водянистая материя», указывает на то, что апостериорная необходимость едва ли может оказать серьезное влияние на рассмотрение вопросов, связанных с объяснением, физикализмом и т. п. Сиверт (Siewert 1994) обходит проблемы, порождаемые апостериорной необходимостью, именно в таком ключе. Я и сам подумывал о том, чтобы поступить так же, но в итоге решил, что двумерная модель представляет интерес и сама по себе.

23. Разумеется, могут существовать и пограничные случаи, в которых будет неопределенным, отсылало бы понятие к некоему объекту, если бы данный мир оказался актуальным. Это не проблема: мы можем допускать неопределенность в первичном интенсионале, подобно тому как мы иногда допускаем неопределенность референции в нашем собственном мире. Могут также существовать случаи, в которых имеется несколько в равной степени приемлемых кандидатов на роль референта понятия в каком-то мире (так, возможно, обстоит дело с «массой» в релятивистском мире — это слово могло бы отсылать либо к массе покоя, либо к относительной массе); так же как мы терпели бы разделенную референцию в подобных актуальных случаях, мы не должны были бы исключать ее наличие, говоря о значении первичного интенсионала.

В некоторых пограничных случаях могло бы оказываться даже, что отнесение объекта к экстенсионалу понятия зависит от разного рода случайных исторических факторов. Хороший повод для размышления на эту тему дает статья Уилсона (Wilson 1982), где обсуждаются подобные случаи, включая, к примеру, гипотетическую ситуацию с друидами, которые решили бы классифицировать самолеты в качестве «птиц», если бы они вначале увидели самолет в небе — но не тогда, когда они вначале нашли бы самолет, потерпевший крушение в джунглях. Можно было бы попробовать классифицировать два этих различных сценария в качестве различных вариантов реализации актуального мира, сохранив, таким образом, фиксированный, детализированный первичный интенсионал; или же можно было бы счесть подобные случаи неопределенными относительно ядра первичного интенсионала. Так или иначе, но небольшая размытость по краям первичного интенсионала полностью совместима с тем, каким образом я применяю данную модель.

24. См. (Field 1973). Думаю, что анализ в терминах первичных интенсионалов позволяет счесть «изменение значения» менее частым, чем обычно думают. Так, случай с относительностью не дает основания считать, что первичный интенсионал «массы» изменился за последний век, хотя наши представления об актуальном мире, несомненно, претерпели изменение. (По сходным причинам можно было бы попробовать использовать анализ в терминах первичных интенсионалов для противостояния идеи «холизма значения» мысли.) Любое «развитие» первичных интенсионалов в лучшем случае оказывается чем-то более тонким, как в примерах Уилсона (Wilson 1982), где ядро остается неизменным, но исторические случайности могут

изменять классификационные практики по краям. Впрочем, все это за-служивает гораздо более детального анализа.

25. Некоторые различия: (1) каплановское содержание очень близко вторичному интенсионалу, но он трактует характер как функцию от контекста к содержанию, тогда как первичный интенсионал есть функция от контекста к экстенсионалу. Однако при условии придания жесткости первичный интенсионал может быть напрямую извлечен из характера, и наоборот. Я говорю о первом по соображениям симметрии и простоты. (2) Каплан использует свою концепцию для решения проблем, связанных с индексикальными и демонстративными терминами, такими как «я» и «то», но не применяет ее к терминам естественных видов, таким как «вода», поскольку он считает, что «вода» указывает на H_2O во всех контекстах (одинаковое по звучанию слово на Двойнике Земли — это попросту другое слово), и он считает, что процесс фиксации референции в данном случае является частью не семантики, а «метасемантики». Как и прежде, для моих целей не столь уж важно, является ли он частью метасемантики или семантики; важно лишь то, что фиксация референции определенным образом зависит от того, каким именно оказывается актуальный мир.
26. Могло бы показаться, что первичный интенсионал хорошо определен лишь по отношению к возможным мирам, центрированным на индивидах, мыслящих надлежащие мысли или выражают их надлежащим образом. Думаю, однако, что первичный интенсионал может быть естественным путем распространен на более широкий класс миров: мы можем удерживать понятие из нашего собственного мира и рассматривать его применение к другим мирам, трактуемым в качестве актуальных (см. Chalmers 1994c), хотя его референция в некоторых из таких миров может быть неопределенной. Но это мало влияет на последующее изложение.
27. Отметим, что, строго говоря, первичный интенсионал указывает на жидкость в нашем историческом окружении: если я совершаю путешествие на Двойник Земли и говорю «вода», я по-прежнему отсылаю к H_2O .
28. Отношение между вторым и третьим соображениями в этом параграфе — то есть между куайновской эмпирической ревизуемостью и крипкеанской апостериорной необходимостью — сложно и интересно. Как отмечает Крипке, разрабатываемая им модель позволяет решить некоторые, но не все проблемы, поднятые Куайном. Анализ Крипке объясняет апостериорные пересмотры интенсионалов, а значит, и изменения определенных аспектов «значения». Двумерный анализ, однако, находится в согласии с одноинтенсиональным объяснением истинностных значений, приписываемых им в актуальном мире, и поэтому не объясняет куайновской возможности фальсификации ряда предполагаемых априорных концептуальных истин в актуальном мире перед лицом достаточных эмпирических данных. На мой взгляд, предполагаемые концептуальные истины такого рода попросту не являются концептуальными истинами, хотя и могут быть близки им.
29. В применении двумерной модели для установления содержания мысли есть одна тонкость, связанная с тем, что иногда мысль может признавать

центрированный мир в качестве потенциального окружения, даже если он не содержит копию самой этой мысли. К примеру, если я мыслю: «Я в коме», я признаю те центрированные миры, центральный индивид в которых находится в коме, независимо от того, содержат ли они мысли. Поэтому при определении первичных интенсионалов и первичных пропозиций мышления надо действовать осмотрительно; более осмотрительно, чем я действовал здесь в случае языка.

30. Миры должны трактоваться долингвистически, возможно, в качестве дистрибуций базовых качеств. По-видимому, лучше не рассматривать их в качестве совокупности тех или иных утверждений, поскольку подобные утверждения описывают мир, и мы видели, что они могут делать это разными способами. При рассмотрении мира в качестве совокупности утверждений мы утратили бы это различие. Быть может, миры могут рассматриваться в качестве совокупностей пропозиций (Adams 1974) при надлежащем понимании пропозиций, в качестве максимальных свойств (Stalnaker 1976), положений дел (Plantinga 1976), структурных универсалий (Forrest 1986) или же в качестве конкретных объектов, аналогичных тем, что есть в нашем мире (Lewis 1986a). Не исключено и то, что это понятие можно попросту признать элементарным. В любом случае, рассуждения о возможных мирах обоснованы не больше и не меньше, чем рассуждения о возможности и необходимости вообще. Подобно математическим понятиям, эти модальные понятия могут с пользой употребляться еще до осуществления по отношению к ним удовлетворительного онтологического анализа.

Я буду всегда рассматривать миры «качественно» и абстрагироваться от вопросов «этовости». Иными словами, я буду считать два качественно тождественных мира тождественными, и меня не будут интересовать вопросы о том, не могли ли индивиды в таких мирах иметь разные «идентичности» (некоторые авторы пытались доказать, что могли). Эти вопросы о трансмировой идентичности ставят множество интересных проблем, но они по большей части несущественны для моего использования моделей возможных миров.

31. В частности, кто-нибудь мог бы отрицать равенство значения и интенсионала в случае математических терминов. Часто утверждается, что такие положения, как «существует бесконечное множество простых чисел», не являются истинными по значению, несмотря на их истинность во всех возможных мирах; те, кто высказывает этот тезис, скорее всего, воспротивились бы приравниванию значения к интенсионалу.

Другие могли бы не согласиться с тезисом о том, что первичный интенсионал термина, такого как «вода», является частью его значения; возможно, они считают, что значение данного термина исчерпывается его референцией, полагая, что первичный интенсионал имеет отношение скорее к прагматике, чем к семантике. Третий могли бы сопротивляться тезису о том, что вторичный интенсионал является частью значения. В любом случае от использования термина «значение» ничего не зависит. Меня интересует истинность по интенсионалу независимо от того, будут ли интенсионалы значениями или нет.

32. Эта дефиниция представимости близка той, что дана Ябло (Yablo 1993): *P* представимо, если можно вообразить мир, который можно использовать для верификации *P*. Различие в том, что положение «который можно использовать для верификации» не исключает возможности неверного описания представленных ситуаций, так что эта версия представимости в лучшем случае является лишь небезупречным руководством в плане установления возможности. Моя дефиниция устраняет этот источник путаницы. Разумеется, он возвращается в виде большей пропасти между тем, что является на первый взгляд представимым, и тем, что является действительно представимым; поэтому приходится более осмотрительно судить о представимости.
33. Не исключено, что можно применить эту критику к аргументу Декарта: поскольку он может представить себя без тела, его бестелесность есть нечто возможное, и поэтому он нефизичен (так как любая физическая сущность с необходимостью телесна). «Я бестелесен» может быть представимым-1 и поэтому возможным-1, но из этого не следует, что это положение представимо-2 и возможно-2. Напротив, смысл, в котором положение «я телесен» было бы необходимым, если бы он был физическим объектом, связан с необходимостью-2, а не с необходимостью-1. (Первичный интенсионал понятия «я» указывает на индивида в центре любого мира; вторичный интенсионал указывает на Декарта в любом мире.)
34. Можно было бы сказать, что в метафизической необходимости нет ничего особо «метафизического». При таком взгляде она оказывается всего лишь разновидностью концептуальной необходимости с апостериорным семантическим уклоном, идущим от двумерной природы наших понятий. Подробнее о том, что апостериорная необходимость — предмет не только метафизики, но и соглашения, см. (Putnam 1983) и (Sidelle 1989, 1992).
35. Хорган (Horgan 1984c) развивает и обосновывает концепцию, согласно которой все высокоуровневые факты логически супервентны на микрофизических фактах. По его словам, высокоуровневые факты привязаны к микрофизическому «семантическим ограничениям», так что в мире есть лишь микрофизика и «космическая герменевтика». Нельзя, однако, не обратить внимания на то, что он избегает говорить о проблеме сознательного опыта. Среди других сторонников разного рода формулировок тезиса о логической супервентности — Джексон (Jackson 1993), Кирк (Kirk 1974) и Льюис (Lewis 1994).
36. Аргументы в пользу того, что факты об абстрактных сущностях логически супервентны на физическом, можно найти у Армстронга (Armstrong 1982).
37. Сознательный опыт, вероятно, вносит свой вклад в этот первичный интенсионал, если вода отчасти индивидуализируется той разновидностью опыта, который она порождает. Индексикальность точно делает это, что существует из слова «нашем» во фразе «прозрачная, пригодная для питья жидкость в нашем окружении». Эти факты не подрывают логической супервентности с поправкой на сознательный опыт и индексикальность.
38. Как в известном методе предложений Рамсея для применения теоретических терминов; см. (Lewis 1972).

39. Этот пример был предложен в качестве задачи Нэдом Блоком.
40. Сходный тезис о требовании анализируемости для супервентности высказывали Джексон (Jackson 1993) и Льюис (Lewis 1994).
41. Альтернативное правдоподобное объяснение предполагает, что для того чтобы нечто было красным, это нечто должно быть такой вещью, которая имеет тенденцию порождать суждения о красном. Это избавило бы от обсуждаемых здесь проблем, поскольку суждения, похоже, логически супервентны на физическом.
42. За исключением, возможно, того, что в то время, как у меня имеется убеждение о Билле Клинтоне, у моего двойника имеется убеждение о двойнике Билла Клинтона. Но опять-таки мы можем отвлечься от проблем, связанных с трансмировой идентичностью.
43. Ближе всего к такому аргументу аргумент, приводимый крипкеанским Витгенштейном (Kripke 1982), который, по сути, доказывает, что из физических и феноменальных фактов нельзя вывести интенциональные факты, поскольку такой вывод не может быть опосредован физическим, функциональным или феноменальным анализом интенциональных понятий. Соответствующие аргументы (особенно те, которые направлены против функционального анализа) спорны, но в любом случае, как отмечалось ранее, вывод данного аргумента состоит не в том, что интенциональные факты — это дополнительные факты, а в том, что они, строго говоря, вообще не факты.

Если согласиться с крипкеанско-витгенштейновским аргументом против выводимости, интенциональность оказывается в положении, напоминающем то, в котором оказывается мораль (см. ниже). В обоих случаях, (1) похоже, отсутствует концептуальная выводимость В-фактов из А-фактов, притом что (2) если в нашем мире есть В-факты, то они есть и в любом представимом А-неотличимом мире. Единственный осмысленный вывод заключается в том, что, строго говоря, В-фактов не существует и что В-атрибуции должны трактоваться в дефляционистском ключе. Возможность того, что В-факты есть фундаментальные дополнительные факты, исключается соображениями представимости, показывающими необходимость существования априорной связи А-фактов с В-фактами, если последние вообще имеют место.

44. Если существуют субъективные моральные факты, то моральные атрибуции имеют определенные условия истинности, зависимые, однако, от того, кто приписывает их. Если это так, то моральные понятия имеют индексикальный первичный интенсионал, и здесь имеется логическая супервентность с поправкой на индексикальность. За такой анализ ратуют сторонники «субъективистского морального реализма» (Sayre-McCord 1989), трактующие «добро» в духе «доброго для меня» или «доброго в моем сообществе». Поскольку здесь присутствует субъективность, этот реализм оказывается, однако, очень мягким. Из этой позиции, к примеру, следует, что два человека, спорящих о том, что есть добро, могли бы и вообще не иметь расхождений.

45. Аргументы Крипке (Kripke 1972), к примеру те, где говорится о референции имени «Гедель» в разных ситуациях, наводят на мысль, что первичный интенсионал, связанный с использованием имени, обычно нельзя выразить краткой дескрипцией. Они могут также указывать на то, что первичный интенсионал не может быть выражен какой-либо конечной дескрипцией, хотя я не так в этом уверен (они точно показывают, что любая подобная дескрипция должна включать металингвистический элемент и условие, требующее надлежащей каузальной связи с субъектом). Однако ничто в этих аргументах не указывает на то, что имя (как оно используется в любом наличном случае) вообще лишено первичного интенсионала. В самом деле, сами аргументы Крипке опираются на рассмотрение зависимости референции имени от того, каким оказывается актуальный мир, то есть на оценку первичного интенсионала имени в разных центрированных мирах.
46. Ясное обсуждение этого вопроса можно найти у Нагеля (Nagel 1986).
47. Как негативным фактам удается ускользнуть от приведенных выше аргументов в пользу логической супервентности? Аргумент от представимости не проходит, как показывает пример с ангелами. Аргумент от эпистемологии не проходит из-за явного наличия эпистемологической проблемы того, как мы можем познавать универсальные положения неограниченного объема (мы не можем быть уверены в том, что ангелов не существует). Аргумент от анализируемости не проходит, так как нельзя провести анализ этих негативных фактов в одних лишь терминах позитивных фактов (без использования второпорядкового факта «и это все»).
48. Как законам удается ускользнуть от приведенных выше аргументов в пользу логической супервентности? Аргумент от представимости не проходит, как показывает пример выше. Аргумент от эпистемологии не проходит из-за явного наличия проблем с эпистемологией законов и каузальности, засвидетельствованных скептическими выпадами Юма. Аргумент от анализа не проходит, так как закономерность требует поддерживающей контрафактической всеобщей регулярности, и такие контрафактуалы нельзя проанализировать в терминах конкретных фактов об истории мира (при всем уважении к Льюису (Lewis 1973)). Конкретные факты об истории мира в пространстве и времени совместимы с истинностью самых различных контрафактуалов.
49. Юмистские взгляды на законы и каузальность защищают Льюис (Lewis 1986b), Макки (Mackie 1974) и Скайрмс (Skyrms 1980). Аргументы против такого рода позиции можно найти у Армстронга (Armstrong 1982), Кэрролла (Carroll 1994), Дретске (Dretske 1977), Молнара (Molnar 1969) и Тули (Tooley 1977).
50. Среди тех, кто, напротив, похоже, считает, что логическая супервентность — это скорее правило, чем исключение, — Армстронг (Armstrong 1982), Хорган (1984c), Джексон (Jackson 1993), Льюис (Lewis 1994) и Нагель (Nagel 1974).
51. Подробнее на эту тему см. (Horgan and Timmons 1992b).

Глава 3

1. Кирк (Kirk 1974) живо описывает зомби и даже очерчивает ситуацию, в которой мы могли бы поверить, что некто в актуальном мире превратился в зомби, специфицируя соответствующие промежуточные случаи. В сходном ключе Кемпбелл (Campbell 1970) обсуждает «имитированного человека», физически идентичного нормальному, но полностью лишенного опытных переживаний.
2. Такую косвенную аргументацию в пользу логической возможности зомби приводит Кирк (Kirk 1974).
3. Якоби (Jacoby 1990) прекрасно показывает, что аргументы от представимости не более проблематичны для функционалистских концепций сознания, чем для материалистических концепций в целом. Он считает это доводом в пользу функционалистских концепций, я — доводом против материалистических концепций.
4. В действительности это привело бы к замене красного не синим, а желтым, так как оба этих цвета находятся в предельных позитивных точках своих осей. Эти детали, однако, несущественны. Любопытные характеристики, присущие цветовому пространству человека, на понятном уровне обсуждаются в (Hardin 1988).
5. Хардин (Hardin 1987, с. 138), кстати, признает это. Он говорит, что такая инверсия была бы «стренной», но не «концептуально некогерентной».
6. Гундерсон (Gunderson 1970) говорит нечто подобное, рассуждая об «исследовательской асимметрии» между высказываниями с позиции первого лица и с позиции третьего лица.
7. Томпсон (Thompson 1992) замечает, что в черно-белой комнате Мэри все равно не была бы лишена переживаний различных цветов — возникавших, когда она, к примеру, терла бы глаза. Чтобы обойти эту трудность, нам, быть может, пришлось бы допустить ее врожденную слепоту к цветам.
8. О витализме в этом контексте говорят Черчленд (Churchland 1995, с. 193) и Деннет (Dennett 1991, с. 281).
9. В конце второй части своей книги Деннет обещает, что в третьей части он покажет, почему его функциональная концепция может объяснить всё, что надо объяснить относительно сознания; однако соответствующие аргументы отыскать непросто. Большая часть его рассуждений состоит из наблюдений, касающихся когнитивных процессов, с которыми с радостью согласится тот, кто разделяет мои настроения. Вопрос состоит не в том, правильно ли это понимание когнитивных процессов, а в том, объясняет ли оно опыт. Решающий аргумент, похоже, находится в диалоге на с. 362–368, где Деннет (по сути) утверждает, что в объяснении нуждается кажимость вещей и что его теория объясняет ее. Но, как я доказываю в главе 5, это утверждение играет на двусмысленности слова «кажется», а именно на том, что оно может иметь как психологический, так и феноменологический смысл. Эта теория могла бы объяснить нашу предрасположенность выносить определенные суждения о стимулах, но они и не были чем-то загадочным, тем, что нуждается в объяснении.

Ряд аргументов содержится и в главе 12: (1) аргумент против эмпирической возможности инвертированных квалиа (при неизменности поведения); но эмпирическая невозможность совместима здесь с нередуктивной позицией; (2) аргумент против джексоновского аргумента знания; я обсуждаю его в главе 4; (3) заявление о смехотворности эпифеноменализма квалиа; я обсуждаю эту проблему в главах 4 и 5. Деннет нередко смешивает вопросы о естественной и логической возможности. Так, он полагает, что «квалиафил» будет считать, что вычислительная машина лишена опытных переживаний, и поэтому уделяет немало места доказательству, что подобные машины будут обладать таким же сознанием, как и мы сами. Но это целиком совместимо с нередуктивной позицией. В последующих главах я сам доказываю это.

10. Позже Крик и Кох расширили свой поиск нейронной основы сознания за пределы колебаний в 40 Гц, но это не изменило его сути. Колебания хороши своей наглядностью.
11. Цитируется в (*Discover*, November 1992, с. 96). Крик (Crick 1994, с. 258) тоже не исключает, что наука не сможет объяснить квалиа, но он более осторожен в своих высказываниях.
12. Эдельман (Edelman 1989, с. 168) вполне определенно высказываеться на этот счет: «Достаточно создать модель, объясняющую их различие, вариацию и последствия. Как ученые мы можем не беспокоиться по поводу онтологических тайн: почему нечто вообще есть, или почему теплое ощущается как теплое». Он проводит аналогию с квантовой теорией поля, которая позволяет различать энергии и материальные состояния, но не говорит нам о том, почему вообще существует материя. Эта аналогия прекрасно совместима с нередуктивной концепцией, которую я развиваю в последующих главах.
13. В самом деле, если бы сознание было логически супервентным на физическом, то «коллапсические» интерпретации нельзя было бы даже выдвинуть, так как исчезли бы основания обосновывать сознание в законах.

Глава 4

1. Эдельман (Edelman 1992) дает похожий подзаголовок своей (как предполагается, материалистической) книге: «Как сознание возникает в мозге».
2. На мой взгляд, позицию Серла гораздо более естественно толковать как дуализм свойств, чем как материализм, несмотря на его собственные представления на этот счет. Тезис о том, что состояния мозга каузально вызывают феноменальные состояния, а также использование аргументов с зомби подкрепляют это толкование, как подкрепляет его и заявление о том, что «в мозге есть лишь нейрофизиологические процессы и сознание». Данную интерпретацию подтверждает и серловский аргумент об интенциональности в главе 8. Серл доказывает, что интенциональность реальна (с. 156), притом что интенциональные факты не могут быть конституированы нейрофизиологическими фактами (с. 157–158). Единственное решение этой загадки, доказывает он, состоит в том, что сознание должно частично

конституировать интенциональность, так как в онтологии мозга помимо мозга есть только сознание. Этот аргумент, похоже, предполагает дуализм свойств по отношению к сознанию.

Объясняя свою онтологию в главе 5, Серл доказывает, что сознание нередуцируемо, но это не имеет глубоких последствий. Он говорит, что такие феномены, как тепло, редуцируемы лишь потому, что мы переопределляем их так, чтобы элиминировать феноменологический аспект (тем способом, который я обсуждал в главе 2), но подобное переопределение тривиальным образом неприменимо к сознанию, состоящему исключительно в его субъективном аспекте. Это представляется верным. Как я утверждал в главе 2, такие феномены, как тепло, редуктивно объяснимы, только если мы не принимаем во внимание сознательный опыт. Он, однако, добавляет, что «это показывает, что нередуцируемость сознания является тривиальным следствием прагматики наших дефинитивных практик» (с. 122). Это, похоже, переворачивает всё с ног на голову. Скорее, такие практики являются следствиями нередуцируемости сознания: если бы мы не отвлекались от переживания тепла, мы вообще не смогли бы редуцировать тепло! Таким образом, нередуцируемость — это исток, а не следствие наших практик. Трудно понять, как что-то из этого тривиализирует нередуцируемость сознания.

3. Близкие аргументы, объясняющие то, почему материалист не может апеллировать к апостериорной необходимости, были выдвинуты Джексоном (Jackson 1980, 1994), Льюисом (Lewis 1994) и Уайтом (White 1986).
4. Сходный тезис высказывает Джексон (Jackson 1980), доказывающий, что, даже если апостериорные соображения и могут установить физичность свойства *больь*, материализм все равно столкнется с проблемой, связанной со свойством презентации боли.
5. Билер (Bealer 1994) тоже предлагает в качестве стратегии отталкиваться от физического термина, но он не доводит свое размышление до его естественного вывода.
6. Эта позиция редко артикулируется в публикациях. Большинство из тех, кто апеллирует к апостериорной необходимости, защищая материализм, обращается к крипкеанским соображениям (напр., Hill 1991; Lycan 1995; Тье 1995), и почти никто с этой целью четко не отстаивал более сильную версию метафизической необходимости. Кажется, однако, что подобные взгляды должны были бы неявно разделять Бигелоу и Паргеттер (Bigelow and Pargetter 1990), Бирн (Byrne 1993), Левин (Levine 1993) и Лор (Loar 1990). Бирн, Левин и Терри Хорган отстаивали эту позицию в частных беседах.
7. Иногда можно услышать, что метафизически, но не концептуально необходимыми являются математические истины. Это зависит от тонких моментов, связанных с анализом математических понятий и концептуальной необходимости, хотя и есть широкое согласие относительно того, что математические истины априорны (с небольшой оговоркой, отмеченной в следующем сегменте текста). Ключевым является то, что невозможно представить мир, в котором математические истины оказались бы

ложными. Поэтому эти истины не делают пространство возможных миров более узким, чем множество представимых миров.

Можно было бы предположить, что примером метафизической супервентности без априорной связи является моральная супервентность, но довод в пользу сильной метафизической необходимости в данном случае кажется даже более слабым, чем такой довод в случае опыта. Здесь имеются альтернативы (антиреализм, априорная связь) — гораздо более привлекательные, чем соответствующие альтернативы для сознательного опыта. Кроме того, похоже, нельзя представить мир, физически и ментально идентичный нашему, который, однако, отличался бы от нашего мира в моральном плане. Так что и здесь мы опять-таки видим, что моральная супервентность не накладывает дополнительных ограничений на пространство возможных миров.

8. Джексон (Jackson 1995) приводит простой аргумент, демонстрирующий странность физикалистских позиций, не признающих априорной связи физического и психологического (в равной степени он может быть приложен к позиции «сильной метафизической необходимости» и позиции «когнитивного ограничения»): «Едва ли имеются такие факты об очень простых организмах, которые нельзя было бы априори дедуцировать из достаточного количества информации об их физической природе и о том, как они взаимодействуют со своим окружением, охарактеризованным с его физической стороны. Физический рассказ об амебах и их взаимодействии с окружением — это всё, что можно рассказать об амебах... Но, согласно материализму, мы отличаемся от амеб, по сути, только сложностью наших составных частей и их организации. Трудно понять, как подобное различие могло бы сгенерировать важные факты о нас, в принципе неподвластные нашим дедуктивным способностям... Вспомним о схемах в классных комнатах по биологии, иллюстрирующих эволюционное развитие от одноклеточных организмов с левой стороны до высших обезьян и человека справа: где на этой шкале биолог сможет убедительно указать то место, где возникает невозможность априорного дедуцирования важных фактов, касающихся нас? Или, если на то пошло, какую стадию нашего индивидуального развития от зиготы материалист мог бы убедительно обозначить в качестве такой, на которой возникают такие важные факты о нас, которые нельзя было бы вывести из наших физических характеристик?»
9. Джон О'Лэри-Хоусон и Барри Левер независимо друг от друга в частных беседах провели аналогию между психологической супервентностью и сложными математическими истинами.
10. Отметим, что это рассуждение показывает отсутствие аналогии с математическими истинами, причем с ним должны согласиться даже те, кто признает сильный тезис о том, что существуют до такой степени глубокие математические истины, что никакой класс существ не может познать их априори.
11. Эта позиция странным образом очень близка той, которую занимают такие редукционисты, как Деннет. В конце концов, и те, и другие считают, что соответствующие интуиции возникают вследствие когнитивных

дефектов. Главное различие состоит в том, что, с точки зрения редукциониста, некоторые из нас могут преодолеть эти дефекты, тогда как тот критик, о котором мы говорим, полагает, что это невозможно. Тем не менее этот критик не может исключить, что кто-то другой (возможно, тот же Деннет) уже достиг просветления. В конце концов, человек с подобными дефектами не смог бы оценить решение просветленного!

12. Наиболее развернутая версия аргумента от логической возможности представлена Кирком (Kirk 1974). Его можно найти также в (Campbell 1970), (Nagel 1974), (Robinson 1976) и в других работах. Моя версия этого аргумента отличается от других главным образом использованием понятия супервентности для создания унифицированного контекста обсуждения и рассмотрением роли апостериорной необходимости. Близкий аргумент, отталкивающийся от возможности инвертированных спектров, представлен также в (Seager 1991).
13. Кажется, что мы вправе сказать, что в известном смысле фразы «вода мокрая» и «H₂O мокрая» выражают разные факты, так же как, кстати, и «вода есть H₂O» и «H₂O есть H₂O». В этом смысле мы индивидуализируем факты скорее с помощью первичных, чем вторичных интенсионалов выражавших их терминов.
14. Об этом, по сути, говорит Локвуд (Lockwood 1989, с. 136–137). Он полагает, что незнание кого-то о том, что каждому из модусов презентации соответствует один и тот же факт, должно быть связано с его незнанием какого-то существенного факта или фактов при любом модусе презентации. Сходный тезис высказывает Кони (Conee 1985a).
15. Лор высказывает предположение, что феноменальные понятия — это распознавательные понятия, и доказывает, что вполне оправданным было бы ожидание, что распознавательное понятие *R* будет «вводить» то же самое свойство, что и теоретически специфицированное свойство *P*. В качестве примера он говорит о том, что некто может распознавать какие-то кактусы в Калифорнийской пустыне, не имея теоретических познаний о них. Представляется, однако, что это не так: если субъект не может априори знать, что *R* есть *P*, то референция к *R* и *P* фиксируется по-разному и фиксирующие ее интенсионалы могут расходиться в ряде представимых ситуаций. Если мы не прибегнем к помощи дополнительных механизмов сильной метафизической необходимости, то различие первичных интенсионалов будет соответствовать различию фиксирующих референцию свойств.

В одном месте Лор допускает, что распознавательные понятия осуществляют референцию «напрямую», без содействия фиксирующих референцию свойств (первичных интенсионалов), но и это представляется неверным. Сам факт того, что понятие могло бы отсылать к чему-то другому (к примеру, к другому множеству кактусов) в иной представимой ситуации, говорит нам о том, что в этом случае задействован содержательный первичный интенсионал. Так что референция не может быть по-настоящему «прямой» в релевантном смысле этого слова.

16. Тот, кто увлечен проблемой индексикальности, мог бы сказать, что локализация центра центрированного мира может иметь онтологическое значение, или, быть может, что могло бы существовать различие индексикальных фактов между обычными возможными мирами (скажем, если бы существовало нечто подобное нагелевскому «объективному я»?). Эти вопросы, так же как и те, что связаны с онтологическим статусом индексикальности в целом, не вполне ясны для меня. Обсуждая эти темы, я исхожу из того, что мой оппонент признает, что индексикальность не ведет к онтологическому провалу, чтобы отметить, что даже в этом случае аналогия с провалом, имеющимся в области феноменального, не проходит.
17. Мысли об опыте действительно могут походить на индексикалы и еще в одном случае, а именно когда мы указываем на опыт как на «этот опыт». Когда кто-то отсылает к одному из двух качественно тождественных опытных переживаний (как в сценарии с двумя трубками в (Austin 1990)), представима ситуация, когда знание всех «объективных» фактов могло бы оставлять неопределенным вопрос о том, к какому именно переживанию мы отсылаем. (Отметим, что проблема здесь связана с отсылкой не к типам, а к конкретным случаям.) Если материалист захочет использовать этот пример как довод против аргумента знания, то я обращаю его внимание на то, что: (1) в данном случае эпистемически дополнительный факт независим от феноменальных фактов (даже знание всех феноменальных фактов не говорит мне о том, какой из них является этим опытным переживанием); (2) он не демонстрирует нам такую ситуацию, в которой имеется представимый нецентрированный мир, отличающийся от данного мира, и поэтому он не может быть использован для построения онтологического аргумента, подобного тому, который представлен в тексте; (3) этот дополнительный факт в лучшем случае может послужить локализации того, что происходит в центре мира, сообщая нам, какая сущность — эта, так же как индексикальные факты говорят нам о том, какая сущность — я.

Реально все это означает то, что в некоторых случаях нужно более информативно представлять центр мира: маркировать не только индивида и время (как «я» и «здесь»), но и опыт (как «этот»). Нечто подобное, похоже, справедливо и относительно ориентационных демонстративов, таких как «левое» и «правое» (знание объективных фактов о мире могло бы не говорить о том, какое из направлений является левым, какое — правым). Все это связано с эпистемическими провалами относительно неопасной индексикальной разновидности: ни в одном из этих случаев нет нецентрированных миров, где есть базовые, но нет добавочных фактов.

18. Выдвигая сходное возражение, Черчленд (Churchland 1985) высказывает предположение, что аргумент Джексона играет на двусмысленности слова «знает»: Мэри обладает совершенным пропозициональным или сентенциальным знанием о физических фактах, но ей недостает знания по знакомству с переживаниями красного. Ответ аналогичен. В той мере, в какой знание Мэри о переживании красного сужает способ данности мира, это фактическое знание, и аргумент проходит (в нем не требуется признания, что фактическое знание должно быть «сентенциальным»). Поэтому, как

- и Льюис с Немировым, Черчленд обрекает себя на неправдоподобный тезис, что знание Мэри о переживании красного ничего не говорит ей о том, каков мир.
19. Лайкан (Lycan 1995) приводит девять (!) аргументов в пользу того, что знание Мэри будет содержать новую информацию.
 20. Есть и множество других ответов на аргумент знания, которые я не обсуждал, но моя реакция на них должна быть предсказуема. Упомяну лишь один: Дретске (Dretske 1995) доказывает, что Мэри недостает знания о ее окружении. Если бы она больше знала об устройстве красных вещей, она знала бы, что именно репрезентируют переживания красного, и поэтому (согласно теории Дретске) знала бы, каковы переживания красного. Странно, но Дретске не рассматривает очевидное возражение: даже если Мэри знает всё об устройстве красных объектов, она по-прежнему не знает, каково это — видеть красное!
 21. Спасибо Фрэнку Джексону за обсуждение этого обстоятельства.
 22. Сам Кripке признает (Kripke 1972, прим. 74), что вывод об одном лишь отсутствии тождества может быть слабым. Он, однако, отмечает, что модельными аргументами можно атаковать и более общие виды материализма.
 23. Аналогичным образом аргументы от развоплощения могли бы устанавливать, что ментальные свойства не идентичны физическим свойствам, так как физические свойства могут быть реализованы только физическими объектами; подобный аргумент можно найти у Билера (Bealer 1994). Но вывод о подобной нетождественности опять-таки слаб: он по-прежнему совместим с логической супервентностью, а поэтому и с материализмом. В самом деле, подобный аргумент о нетождественности может быть выдвинут по отношению к чуть ли не любому высокоуровневому свойству.
 24. В ходе тщательного анализа Бойд (Boyd 1980, с. 98) отмечает, что возможность зомби, в отличие от возможности развоплощения, влечет ложность материализма. Поэтому он выдвигает отдельный аргумент против такой возможности, но этот аргумент оказывается схематичным и неубедительным. Бойд проводит аналогию с компьютером, исчисляющим какую-то конкретную функцию, доказывая, что (1) нам может показаться, что все электрические цепи, которые есть в компьютере, могли бы остаться неизменными, притом что он не исчислял бы эту функцию, хотя в действительности это невозможно, и что (2) кажущаяся возможность зомби аналогична этому случаю. Между тем аналогия не проходит. Ситуация с компьютером аналогична (очень трудноуловимой) «кажущейся возможности» того, что могла бы существовать моя физическая копия, которая не обучалась бы в ситуации, в которой я обучался бы, или не различала бы там, где я различал бы. Ничто в этой аналогии не может объяснить гораздо более наглядный характер кажущейся возможности такой копии, которая была бы лишена сознательного опыта.
 25. Замечание Кripке о том, что материалист должен показать, что «те вещи, которые мы можем вообразить, в действительности не являются вещами,

которые мы можем вообразить» (в предпоследнем абзаце (Kripke 1971)), также наводит на мысль об этой слабой трактовке. Крипке не исключает, что эта кажущаяся возможность могла бы быть устранина неким способом, совершенно не похожим на стандартные случаи типа вода/H₂O, но говорит, что «это должен был бы быть более глубокий и тонкий аргумент, чем я могу вообразить, и более тонкий, чем те, которые когда-либо приводились в известной мне материалистической литературе».

26. Хотя данный аргумент нередко трактуется в качестве приложения крипкеанской теории жесткой десигнации, вариация на его тему могла бы в принципе быть выдвинута за десять лет до разработки этой теории. Можно было бы спросить первых теоретиков тождества, почему физические факты об H₂O вынуждают то обстоятельство, что это вода (или нечто водянистое), тогда как физические факты о состояниях мозга, как кажется, не вынуждают то обстоятельство, что в данном случае здесь должна переживать какая-то боль.
27. Хорган (Horgan 1987), как и Бирн (Birn 1993), говорят в этом контексте о «метафизической супервентности». Но если я прав в том, что метафизическая возможность и логическая возможность миров совпадают, то из этого вытекает логическая супервентность.
28. Другие вариации на эту тему — см. (Blackburn 1990), (Feigl 1958), (Lockwood 1989), (Maxwell 1978) и (Robinson 1982).
29. Представление о мире как о чистом каузальном потоке выдвигает Шумейкер (Shoemaker 1980), доказывающий, что все свойства — это «силы», не базирующиеся ни на каких дополнительных свойствах. Аргумент Шумейкера в пользу этой концепции имеет по большей части верификационистский характер, и он напрямую не обсуждает проблемы, с которыми она сталкивается.

Шумейкер, кроме того, доказывает, что, поскольку силы, связанные с тем или иным свойством, существенны для него, законы природы с необходимостью должны иметь апостериорный характер. (Подобные аргументы выдвигает и Суоэр (Swoyer 1982), а Крипке (Kripke 1980) заигрывает с соответствующим выводом.) Двумерный анализ апостериорной необходимости наводит на мысль, что с этим предположением должно быть что-то не так или, по крайней мере, что оно имеет более ограниченный характер, чем кажется. В лучшем случае могло бы оказаться, что миры с другими законами неверно описываются (к примеру) как миры, содержащие электроны; но такие соображения не могут показать невозможность подобных миров. Кроме того, тезис о том, что все силы, связанные с электронами, конститтивны для электронности, кажется неубедительным. Более вероятно, что для того чтобы некая сущность считалась электроном, необходимы лишь некоторые из этих сил, и миры с несколько иными законами, содержащие электроны, возможны. Шумейкер пытается показать, что не существует способа отличия конститтивных сил от неконститтивных, но двумерный анализ указывает, что это различие вытекает из понятия электронности.

Необходимо разграничить ряд вопросов. (1) Реляционно ли фиксируется референция к физическим свойствам? (Шумейкер, Чалмерс: да.) (2) Тож-

- дественны ли физические свойства реляционным свойствам (во вторичном интенсионale)? (Ш: да; Ч: возможно, хотя семантические интуиции могут разниться.) (3) Существенны ли для физического свойства все его номические отношения? (Ш: да; Ч: нет.) (4) Лежат ли в основе этих реляционных свойств внутренние свойства? (Ш: нет; Ч: да)
30. В недавнее время этот взгляд отстаивали Локвуд (Lockwood 1989) и Максвелл (Maxwell 1978); оба они считали его неортодоксальной версией теории тождества. Мне пропагандировал его Грэгг Розенберг.
 31. Хотя см. (Lahav and Shanks 1992), где представлена противоположная позиция.
 32. Льюис (Lewis 1990) приходит к сходному выводу совершенно иным путем.
 33. На этот счет я иногда проводил опросы, выступая с докладами о сознании и по другим поводам. Результаты неизменно были 2:1 или 3:1 в пользу того, что здесь имеется что-то еще, что нуждается в объяснении. Разумеется, демократия — не лучший способ решения философских вопросов, но когда мы сталкиваемся с одной из проблем, которую нельзя решить аргументами, баланс изначальных интуиций имеет значение.
 34. *Биологический материализм*. Общая позиция (Hill 1991; Searle 1992) состоит в том, что сознание с необходимостью биологично. Согласно этой концепции, материализм верен, но лишенные сознания системы с той же самой функциональной организацией, что и системы, обладающие сознанием, логически, а вероятно даже и эмпирически, возможны. После того, однако, как мы признали логическую возможность моего лишенного сознания функционального изоморфа, мы должны наверняка допустить и логическую возможность моего лишенного сознания биологического изоморфа, так как концептуальная связь нейрофизиологии с сознательным опытом не более прочна, чем концептуальная связь кремния с сознательным опытом. Похоже, таким образом, что эту позицию лучше всего рассматривать как вариацию на тему дуализма свойств, где сознание оказывается дополнительным фактом, выходящим за пределы физических фактов. Если же нет, то в лучшем случае она должна сочетаться с апелляцией к сильной метафизической необходимости для подкрепления связи между биохимией и сознанием, наследуя все трудности такой концепции.
(Серл (Searle 1992) признает логическую возможность зомби и в действительности считает, что между микрофизическими процессами и сознательным опытом имеется лишь каузальная связь, так что его, по-видимому, лучше всего рассматривать как дуалиста свойств. Хилл (Hill 1991) пытается уйти от допущения возможности зомби, обращаясь к жестким десигнаторам, но мы видели, что эта стратегия не помогает.)
 35. *Физикалистский функционализм*. Согласно этому популярному воззрению (напр., Shoemaker 1982), свойство обладания сознательным опытом — функциональное свойство, но свойство обладания конкретным сознательным опытом (к примеру, ощущением красного) является нейрофизиологическим свойством. В соответствии с этим взглядом, инвертированные спектры между функциональными изоморфами логически, а быть может и эмпирически, возможны, но полностью лишенные сознания функцио-

нальные изоморфы — нет. Но опять-таки, если мы согласились с тем, что инвертированные функциональные изоморфы логически возможны, мы должны признавать и логическую возможность инвертированных физических изоморфов, так как концептуальная связь нейрофизиологии с конкретными переживаниями не более прочна, чем концептуальная связь кремния с конкретными переживаниями. Так что и здесь представляется, что физические факты не детерминируют все факты, и из этого вытекает некая разновидность дуализма свойств. И здесь тоже физикализм можно утвердить лишь на основе проблематичного понятия сильной метафизической необходимости.

Эта концепция нередко позиционируется в качестве апостериорного отождествления феноменальных свойств с нейрофизиологическими свойствами. В таком виде она сталкивается с типичными проблемами, связанными с апостериорной идентификацией (что здесь является первичным интенсионалом?), а также оказывается беззащитной перед приведенным выше аргументом. Критикуя ее в сходном ключе, Уайт (White 1986) отмечает, что сторонникам этой позиции лучше держаться общего функционализма.

36. *Психофункционализм*. Согласно этому воззрению, ментальные свойства отождествляются с функциональными свойствами апостериори, на основе ролей, играемых ими в развитой эмпирической психологии (см. Block 1980). В применении его к феноменальным свойствам получается, что феноменальные свойства обладали бы теми же вторичными интенсионалами, что и функциональные свойства, несмотря на различие первичного интенсионала. Проблемы, связанные с этой позицией, лучше всего анализировать в духе параграфа 2, то есть фокусируясь на первичных интенсионалах. Если первичный интенсионал феноменальных понятий сам функционален, то эта позиция в конечном счете подкрепляется неким аналитическим функционализмом, а если нет, то фокусировка на свойстве, вводимом этим интенсионалом, в итоге неизбежно приведет нас к какой-то разновидности дуализма. В любом случае это воззрение не оказывает никакой дополнительной помощи в деле спасения материализма.

Зашитники этой позиции нередко игнорировали роль понятий в фиксации референции посредством первичных интенсионалов. Даже при наличии научной теории с «убеждением» как теоретическим термином можно будет концептуально рассказать о том, почему такой тип состояний характеризуется как убеждение, а не как желание или что-то совершенно иное. Скорее всего, этот фиксирующий референцию интенсионал сам будет функциональным, указывая на что-то вроде состояния, играющего в этой теории роль, более всего подобающую убеждению, где «подобание убеждению» характеризуется в соответствии с заранее имеющимся у нас понятием. Какой бы ни была природа первичных интенсионалов для феноменальных свойств, мы не уйдем от этих проблем. Концентрация на вторичных интенсионалах означает, что мы лишь загоняем эти проблемы под ковер.

Другая проблема с психофункционализмом связана с тем, что из него вытекает своего рода шовинизм, так как здесь придается излишний вес человеческой психологии в решении того, что, к примеру, считать убежде-

нием. См. (Shoemaker 1981) для замечательной критики; впрочем, см. также (Clark 1986) для ответа на нее. Кажется более правдоподобным, что в большинстве ментальных понятий первичные и вторичные интенсионалы совпадают. В ином случае мы оказались бы в ситуациях, где мы и наши двойники на Двойнике Земли имели бы в виду разные вещи, говоря об «убеждении», несмотря на то что наши изначальные понятия не отличались бы друг от друга.

37. *Аномальный монизм*. Согласно этому воззрению, каждое ментальное состояние в индивидуальном плане тождественно какому-то физическому состоянию, хотя строгих психофизических законов не существует. Концепция аномального монизма была предложена Дэвидсоном (Davidson 1970) в качестве объяснения скорее интенциональных, чем феноменальных состояний, но ее все же можно считать уместной по двум соображениям: во-первых, она содержит априорный аргумент в пользу физикализма, основанный только на взаимодействии (пусть и одностороннем) физических и ментальных состояний, и, во-вторых, она отрицает психофизические законы, предполагаемые моей концепцией.

Чтобы убедиться в том, что моя позиция не затрагивается аргументами Дэвидсона, заметим, что ничто в этих аргументах не указывает на невозможность существования точечных законов такого вида: «Если система находится в максимально конкретном физическом состоянии Р, она находится в (максимально конкретном) ментальном состоянии М». В самом деле, Дэвидсон признает супервентность ментального на физическом, из чего при естественной интерпретации, как кажется, вытекает существование таких законов (см. (Kim 1985) для обсуждения). При наиболее благожелательной интерпретации Дэвидсона можно понять так, будто он отрицает не точечные законы, а более интересные типовые законы, связывающие ментальные состояния с физическими состояниями в контексте широких типов, таких как те, что встречаются в народной психологии. И очевидно, что это максимум того, что, как кажется, можно извлечь из его аргументов относительно холизма ментального. Если так, то естественной супервентности ничто не угрожает. Из этого также следует, что аргумент в пользу индивидуального тождества не может проходить. Этот аргумент основан на предположении об отсутствии строгих законов, поддерживающих каузальную связь между физическим и ментальным (и поэтому их заменяет тождество). Но даже строгого точечного закона достаточно для подкрепления того типа связи между физическими и феноменальными состояниями, который я признаю. Так что и дуализму ничто не угрожает.

38. *Репрезентационализм*. В недавние времена популярностью пользовалась позиция (напр., Dretske 1995; Hartman 1990; Lycan 1996; Tye 1995), согласно которой феноменальные свойства — это всего лишь репрезентативные свойства, так что желтые квалиа есть лишь перцептивные состояния, репрезентирующие желтые вещи, или что-то в этом духе. Разумеется, интерпретация этого предположения напрямую зависит от того, каким образом мы в свою очередь объясняем репрезентативные свойства. Чаще всего данное предположение комбинируется с редуктивным объяснением репрезентации (обычно функциональным или телеофункциональным);

в этом случае оно оказывается вариацией на тему редуктивного функционализма и сталкивается с типичными проблемами, связанными с последним. Нередуктивное объяснение репрезентации могло бы избежать этих проблем (хотя могло бы столкнуться с другими), но оно вело бы к нередуктивному объяснению опыта.

Каждая правдоподобность ряда репрезентационистских концепций может быть связана с тем, что им удается скользить между инфляционными и дефляционными прочтениями «репрезентации»; репрезентация во втором смысле — это сугубо функциональное (или телефункциональное) понятие, в первом — нет. Связь между феноменологией и репрезентацией кажется правдоподобной при первом прочтении, редукция репрезентации — при втором. Или же для связывания репрезентативных состояний и феноменальных состояний можно прибегнуть к сильной метафизической необходимости, но это приводит к проблемам, характерным для последней. (Среди современных репрезентационистов Дретске (Dretske 1995) и Харман (Harman 1990), судя по всему, одобряют выраженно редуктивную позицию А-типа, а Лайкан (Lycan 1986) и Тай (Tye 1995), похоже, занимают позицию В-типа, опирающуюся на апостериорную необходимость.)

Другой способ справиться с репрезентационизмом состоит в указании на то, что почти все согласны, что не все репрезентативные состояния являются феноменальными состояниями (не согласные с этим почти наверняка являются нередуктивистами относительно обоих), и поэтому можно спросить: что же делает какие-то из репрезентативных состояний феноменальными состояниями? Именно этот дополнительный критерий несет реальную нагрузку в репрезентативистской теории сознания. Часто этот критерий трактуется в духе требования надлежащей доступности репрезентативного состояния для центральных процессов; но тогда становится очевидно, что мы имеем дело с редуктивным функционалистским объяснением с типичными для него проблемами (почему это должно делать репрезентативное состояние феноменальным?). Альтернативный путь — выделять в качестве релевантных состояний именно те репрезентативные состояния, которые являются феноменальными, но это прямиком возвращает нас к дуализму свойств.

39. *Сознание как мышление более высокого уровня.* Сходным образом можно трактовать и идею, что сознательное состояние — это состояние, являющееся объектом мышления более высокого уровня (см., напр., (Rosenthal 1996) и др.). Если она комбинируется с редуктивным представлением о том, что значит обладать мышлением более высокого уровня, то это, по сути, редуктивное функционалистское воззрение с типичными для него проблемами. Если нет, то это приведет к нередуктивному представлению об опыте (В-типа или С-типа), и поэтому оно совместимо с рассматриваемым мной дуализмом свойств, хотя здесь могут быть свои трудности (обсуждаемые мной в главе 6).
40. *Редуктивный телефункционализм.* Стоит упомянуть позицию Дретске (Dretske 1995), согласно которой к критериям наличия опыта относится также телесный компонент: для наличия опытных пережива-

ний некая система не только должна функционировать определенным образом, но соответствующие процессы должны были быть надлежащим образом отобраны в ходе их истории. Утверждается, что эта концепция может уйти от некоторых проблем, характерных для стандартного функционализма, так как она, к примеру, допускает (и объясняет) возможность функционально идентичных зомби: это всего лишь системы с неверной историей. Она тем не менее сталкивается с собственными версиями главных проблем. Так, кажется в неменьшей степени логически возможным то, что функционально идентичная система с релевантной историей могла бы быть лишена сознания; и знание об опыте не вытекает из сочетания знания об организации с историей. Можно было бы сказать, что эта концепция «ходит» от проблем редуктивного функционализма неверным путем. В конечном счете оказывается, что она ближе редуктивному функционализму А-типа, чем позиции, всерьез принимающей сознание.

41. **Эмерджентная каузальность.** Многие хотели отвергнуть редуктивную концепцию сознания и при этом сохранить за ним ключевую каузальную роль. Популярными в этой связи были аргументы в пользу эмерджентной каузальности — существования новых видов причинности в сложных физических системах. К примеру, Сперри (Sperri 1969, 1992) доказывал, что сознание — это эмерджентное свойство сложных систем, которое само играет определенную каузальную роль; сходные идеи высказывали и британские эмерджентисты, такие как Александр (Alexander 1920; см. для обсуждения McLaughlin 1992). Селларс (Sellars 1981; см. также Meehl and Sellars 1956) тоже высказывал предположение, что в определенных системах — состоящих, к примеру, из протоплазмы или являющихся субстратом разумных субъектов — в игру могут вступать новые законы физической причинности. (Он именовал эту позицию «физикализмом₁» в отличие от «физикализма₂», согласно которому базовые физические принципы, находимые в неорганической материи, имеют универсальное значение.) Эти воззрения нельзя смешивать с «наивным» представлением об эмерджентной причинности, характерным для теории сложных систем, согласно которому низкоуровневые законы порождают качественно новое поведение в результате взаимодействия. В соответствии с более радикальным воззрением, здесь вступают в игру новые фундаментальные принципы, не являющиеся следствиями низкоуровневых законов.

С этой концепцией связаны две проблемы. Во-первых, нет свидетельств в пользу существования подобных эмерджентных принципов каузальности. Насколько нам известно, вся причинность вытекает из низкоуровневой физической причинности, и причинность «сверху вниз» никогда не конфликтует с низкоуровневыми процессами. Во-вторых — и это, быть может, еще более важно — при тщательном анализе выясняется, что эта концепция оставляет сознание столь же избыточным, как и раньше. Чтобы убедиться в этом, заметим, что ничто в этой картине эмерджентной каузальности не требует от нас каких-либо отсылок к феноменальным свойствам. Всю каузальную историю можно рассказать в терминах связи между конфигурациями физических свойств. И по-прежнему можно будет найти такой возможный мир, который будет физически идентичным на-

шему, но в котором будет полностью отсутствовать сознание. Из этого следует, что феноменальные свойства в лучшем случае скоррелированы с каузально действенными конфигурациями. Если при этом, исходя из данной позиции, все же можно показать действенность феноменальных свойств, то подобный маневр можно будет осуществить и в моей концепции. В действительности данную позицию лучше всего рассматривать в качестве одной из версий моей концепции, согласно которой сознание оказывается супервентным на физическом благодаря контингентной номической связи между ними. Она модифицируется добавлением новых законов эмерджентной физической причинности, но эти законы, не меняя ничего по сути, просто вносят в нее дополнительные усложнения.

42. *Мистерианство*. Те, кому несимпатичны редуктивные концепции сознания, нередко утверждают, что сознание может навеки оставаться тайной. Подобный подход был намечен Нагелем (Nagel 1974) и Джексоном (Jackson 1982) и проработан Макгинном (McGinn 1991). Согласно этому воззрению, сознание может настолько же превосходить наше понимание, как астрономия превосходит понимание морских слизней.

Такие суждения могут звучать заманчиво, но они все же поспешны. Утверждение о том, что не существует редуктивного объяснения сознания, не равнозначно утверждению, что сознание вообще необъяснимо. В частности, указание принципов, благодаря которым сознание оказывается естественно супервентным на физическом, могло бы привести к созданию проясняющей его теории, даже если мы придерживаемся нередуктивизма.

Макгинн (McGinn 1991) говорит о существовании необходимой связи между состояниями мозга и сознательными состояниями (в ином случае возникновение сознания было бы чудом), но доказывает, что мы никогда не сможем узнать, в чем состоит эта связь. Из его обсуждения можно сделать вывод, что он имеет в виду логическую или метафизическую необходимость; однако этот аргумент в лучшем случае позволяет говорить о естественной необходимости. Нет сомнений, что контингентная номическая связь между сознанием и физическим не более чудесна, чем любой контингентный закон, и на деле подобная связь кажется гораздо менее таинственной, чем логически или метафизически необходимая связь, превосходящая наше понимание. И неясно, почему мы не можем использовать наше знание о регулярностях, связывающих физические процессы и опыт, для вывода о существовании подобных законов. В следующих нескольких главах я попробую дать некоторые характеристики соответствующим законам. Подобным образом мы сможем увидеть, что нередуктивный взгляд на сознание не обязан приводить нас к пессимизму.

Глава 5

1. Элитцур говорит, что к этому обсуждению его подтолкнули размышления Пенроуза (Penrose 1987).
2. Здесь я оставляю в стороне религиозный опыт. Можно попробовать показать, что в действительности в объяснении здесь нуждается глубокий духовный опыт и благоговение.

3. Я не уверен, что эта линия четко представлена в литературе, но родственные аргументы отыскать можно. К примеру, Фосс (Foss 1989) отвечает на аргумент знания Джексона (Jackson 1982), отмечая, что Мэри могла бы знать всё, что субъект, обладающий цветным зрением, сказал бы о разных цветах, или даже что он мог бы сказать. Но, разумеется, этого совершенно недостаточно для того, чтобы знать всё, что надо знать.
4. На самом деле, конечно, это не помогло бы.
5. Дретске (Dretske 1995) выдвигает сходный аргумент, доказывая, что его теория объясняет то, какими кажутся вещи, и поэтому объясняет то, что нуждается в объяснении. Опять-таки здесь обыгрываются психологический и феноменальный смыслы слова «кажется». В целом «кажимость» — бедный термин для использования его при характеристике того, что нуждается в объяснении в теории сознания, именно из-за этой двусмысливности. Интересно, что обычно он используется только сторонниками редуктивных концепций.
6. Конечно, в литературе можно найти разнообразные попытки атаковать идею «данности» и «чувственных данных». Не думаю, однако, что какая-либо из этих попыток привела к опровержению идеи о том, что наличие опыта является источником обоснования того или иного убеждения об опыте. Они дают серьезные основания отрицать различные более сильные утверждения, такие как утверждение о том, что всякое знание производно от знания об опыте, а также тезисов о том, что мы воспринимаем мир, воспринимая чувственные данные, и что наличие опыта автоматически подводит этот опыт под какое-то понятие. Но я не высказываю подобных тезисов.

Селларс (Sellars 1956) убедительно критикует представление о том, что «чувствование чувственного содержания *s*» влечет невыводное знание об *s*. Он отмечает, что знание — это концептуальное состояние, так что подобное знание едва ли может быть чем-то примитивным; опыт, однако, кажется чем-то более примитивным. Наличие опыта — это, похоже, неконцептуальное состояние, и наше знакомство с опытом представляется собой неконцептуальное отношение (хотя это зависит от того, как мы определяем условия «концептуальности»). Остается, таким образом, вопрос, как неконцептуальное состояние может быть эвиденциальной основой для концептуального состояния. Это трудный вопрос, но с ним сталкивается не только тот, кто занимает нередукционистскую позицию по отношению к сознанию. Он возникает даже тогда, когда речь идет об обычном перцептивном знании мира, где даже редукционист должен признать, что обоснование убеждения отчасти основывается на неконцептуальном источнике, если только он не готов принять такие альтернативы, которые, как кажется, сталкиваются с еще большими трудностями. Думаю, что концепцию обоснования такого рода можно создать, но это отдельная объемная задача. Здесь я просто отмечаю, что нередукционизм относительно сознания не создает особых беспокойств в этой области.

7. Общую стратегию такого рода избирает Хилл (Hill 1991), детально отвечая на скептические аргументы, касающиеся опыта и основывающиеся

на возможности «эрзаца боли», в частности на аргумент Шумейкера (Shoemaker 1975a). Хилл высказывает ряд положений, конгениальных моей трактовке этих вопросов, хотя он защищает скорее биологический материализм В-типа, чем дуализм свойств. Среди них — сравнение скептических аргументов, касающихся опыта, и скептических аргументов, касающихся внешнего мира, а также аргументы против «условия различения», согласно которым субъект не может быть обоснованно убежден в том, что *P*, если во всякой ситуации, в которой у него отсутствуют свидетельства в пользу того, что *P*, он не был бы способен понять, что у него отсутствуют свидетельства в пользу того, что *P*.

8. Это предположение высказал во время дискуссии Джон О'Лэри-Хоусон.
9. Можно было бы подумать, что противоположный ход мыслей — согласно которому убеждение зомби «я обладаю сознанием» оказывается истинным, так как его понятие указывает на функциональное свойство, — мог бы оказаться полезным для разрешения эпистемологических проблем дуалиста свойств, так как при этом больше нельзя было бы заключить, что у меня есть убеждения, которые оказываются обоснованными там, где убеждение зомби необоснованно. Но какую бы стратегию мы здесь ни избрали, зомби все равно будет обладать какими-то ложными убеждениями, к примеру убеждением о том, что он наделен свойствами, выходящими за пределы его физических и функциональных свойств, и проблема обоснования присущих мне соответствующих убеждений вновь встанет перед нами в этом облике.
10. Об этом, к примеру, говорил Лайкан в частной беседе. Можно также вернуться к такому понятию, как «опытные переживания того типа, которые обычно вызываются (у большинства из нас) красными вещами», хотя здесь мы избегаем релятивизма ценой возможности систематического заблуждения относительно категории наших собственных переживаний.
11. Подобный релятивизм отсутствует в случае внешних понятий цвета, когда, к примеру, мы говорим о красноте как о свойстве скорее объектов, чем опытных переживаний. В первом приближении кажется, что референция к красным вещам фиксируется как референция к вещам, которые обычно вызывают (у большинства из нас) тот же тип переживаний цвета, что и некоторые эталонные примеры. Это более «публично» в двух отношениях: из-за отсылки к публичным эталонным примерам и из-за отсылки к опытным переживаниям целого сообщества. Таким образом, человек с инвертированным спектром использовал бы выражение «красные вещи» для обозначения тех же самых вещей, что и я, даже если его термин «переживания красного» указывает на нечто иное.

Можно было бы задаться вопросом об индивидах и иных границах их цветового пространства, к примеру о тех, кто считает, что морковь имеет тот же цвет, что роза и помидор. Наиболее естественным кажется сказать, что их высказывание «морковь красная» ложно вследствие того элемента понятия красного, который свойствен сообществу, хотя не исключено, что имеется и такой не столь связанный с сообществом смысл, в котором его можно признать истинным (где «красное» сводится к «красному для меня»). Впрочем, даже здесь остается не так много места для релятивизма,

так как для любого индивида этот термин по-прежнему будет привязан к внешним эталонам. Даже если опираться на этот релятивистский смысл, термин «красное» должен будет указывать для любого индивида на большое множество красных вещей; сходная картина наблюдается и с другими понятиями.

Можно также было бы устраниТЬ всякую зависимость от опыта из характеристик внешних понятий цвета, характеризуя вместо этого их в терминах суждений: тогда получалось бы, что красные вещи — это вещи, относительно цвета которых обычно высказывается то же суждение, что и относительно цвета эталонных объектов. Преимущество такого подхода состоит в том, что он позволяет зомби истинно высказываться о зеленых объектах, что может быть разумным. В конце концов, представляется, что для работоспособности референции терминов цвета требуется, скорее, не сходство опытных переживаний, а интерсубъективное сходство суждений.

12. Различие квазитативного и реляционного понятия «переживания красного» весьма близко различению, которое проводит Нида-Рюмелин (Nida-Rümelin 1995) между «феноменальной» и «нефеноменальной» интерпретацией приписывания убеждений, таких как «Марианна убеждена, что небо кажется Питеру синим». Согласно этому автору, «феноменальная» интерпретация приписывает убеждение при участии соответствующего квазитативного понятия того, кто приписывает убеждение; тогда как при нефеноменальной интерпретации убеждение приписывается при помощи реляционного понятия. (Пример релевантного реляционного понятия, приводимый Нида-Рюмелин, напоминает коммунальное понятие, упомянутое в прим. 10.)
13. Разумеется, я специально обозначаю понятие символом «*R*», чтобы вызывать реминисценцию с «*E*», которым пользовался Витгенштейн в своем аргументе относительно приватного языка. Здесь я не буду даже предпринимать попытки проанализировать его аргумент; эта задача оказалась бы особенно трудной в силу отсутствия общепринятой интерпретации того, в чем, собственно, состоит этот аргумент. Достаточно сказать, что все версии этого аргумента, которые я видел, или основываются на очень сомнительных посылках, или столь же применимы к обыденным понятиям, сколь и к приватным опытным понятиям, или и то, и другое.
14. В самом деле, даже небольших интроспективных усилий достаточно, чтобы обратить внимание на то примечательное обстоятельство, что даже «самим себе» мы можем мало что сказать относительно различия переживаний красного и зеленого, опять-таки если отвлечься от отсылок к реляционным свойствам — несмотря на нашу осведомленность относительно богатства их внутреннего различия. Это можно рассматривать в качестве дополнительного свидетельства в пользу того, что данные качества существуют в области феноменального и не являются чем-то, что находит прямое отображение в сфере психологического.
15. В чем-то этот ход мысли зеркален тому, который можно найти у Шумейкера (Shoemaker 1975): если инвертированные спектры возможны, то как

квалитативные состояния, так и квалитативные убеждения не могут получить функционального определения, хотя Шумейкер выражает всё это в терминах жесткой десигнации с реляционной фиксацией референции, так что в итоге он фокусируется здесь на вторичных интенсионалах.

16. Это наблюдение параллельно наблюдению Нида-Рюмелин (Nida-Rümelin 1995), согласно которому различие феноменальных и нефеноменальных убеждений не есть типичный пример различия *de re/de dicto*.
17. Отметим, что в данном случае речь не идет о простом индексикале «этот», первичный интенсионал которого не меняется в зависимости от того, опыт ли это *R* или *S*, и в связи с которым было бы бессодержательной тривиальностью сказать, что этот опыт есть опыт этого рода. Скорее, это содержательное «этот» с таким первичным интенсионалом, который указывает на переживания *S* во всяком центрированном мире.
18. Кони (Conee 1985b) опирается на подобное конститутивное отношение между квалиа и квалитативными убеждениями, отвечая на эпистемологический аргумент Шумейкера (Shoemaker 1975a).

Глава 6

1. Отметим, что для того, чтобы эти принципы поставляли психофизические законы, мы должны интерпретировать второпорядковые суждения, такие как «У меня имеется переживание красного», в реляционистском ключе, обсуждавшемся в последнем параграфе главы 5, а именно «У меня имеется переживание такого типа, который обычно вызывается красными объектами». Реляционистские элементы понятия «переживание красного», в отличие от внутренних квалитативных элементов, будут отражены в физических процессах: соответствующее убеждение, связанное с квалитативным понятием «переживание красного», не будет логически су-первентным на физическом, так что верность таких убеждений не будет давать психофизического закона. Я не особо пользуюсь этим, так как мое обсуждение будет сосредоточено на первопорядковых суждениях, по отношению к которым не возникает указанных проблем.
2. Содержательный анализ феноменологии, связанной с актуальными мыслями, см. в (Siewert 1994).
3. Ср. также наблюдение Нагеля (Nagel 1974) о том, что «структурные черты восприятия могли бы оказываться более доступными для объективного описания, даже если что-то и оставалось бы за бортом».
4. Это в близкой степени родственно гипотезе Джекендофа о вычислительной достаточности: «Всякое феноменологическое различие порождено/подкреплено/проецировано соответствующим вычислительным различием» (Jackendoff 1987, с. 24).
5. Обсуждение слепого зрения в таком ключе см. (Tye 1993, Block 1995 и особенно Dennett 1991).
6. Обсуждение опасностей смешения сознательного опыта и осознания опыта, а также превосходную критику высокоуровневых подходов в целом

- можно найти в (Siewert 1994). См. также (Dretske 1995) для параллельной критики с позиции редукционизма.
7. Различие первопорядковых и второпорядковых объяснений — отражение различия двух функциональных понятий сознания, C1 и C2, которое проводит Нилкин (Nelkin 1989).
 8. Подобное предложение рассматривает Керратерс (Carruthers 1992), доказывающий, что доступность для рефлексивного осмысления естественно необходима и достаточна для квалитативного чувства (похоже, правда, что Керратерс имеет в виду более сильную разновидность доступности, так как армстронговский невнимательный водитель грузовика не удовлетворяет его критерию). Поскольку он четко заявляет, что эта связь лишь естественно необходима, его предложение кажется нередуктивным, хотя при этом он характеризует этот взгляд как физикалистский. В частной беседе сходное объяснение выдвигал и Алвин Голдман, но не в качестве редуктивной концепции, а для характеристики сознательных состояний привычных систем.
 9. Различие первоуреневых регистраций и первоуреневых суждений параллельно различению Дретске (Dretske 1995) феноменальных и доксастических разновидностей когнитивных состояний. Вторая соответствует мнению системы о вещах, первая — способу презентации вещей системе. Конечно, между моей схемой и схемой Дретске есть существенное различие, и состоит это различие в том, что Дретске является, по сути, редуктивным функционалистом (на деле редуктивным телесофункционалистом), отождествляющим переживания с (телесо)функционально определенными первопорядковыми регистрациями. Я против этого отождествления по стандартным причинам, но тем не менее остается правдоподобным, что переживания соответствуют первопорядковым регистрациям. Моя схема попросту содержит дополнительное различие, допуская три различных типа состояний — суждения, регистрации и феноменальные состояния, причем регистрации и феноменальные состояния оказываются скоррелированными, но различными, тогда как Дретске признает два вида состояний: суждения и феноменальные состояния, не проводя даже концептуального различия между феноменальными состояниями и соответствующими им регистрациями.
 10. Существует громадное множество интересных вопросов относительно типа презентативного содержания, имеющегося в первопорядковых регистрациях, создающих осведомленность, и о типе содержания, параллельно имеющегося у соответствующих переживаний. Проблемы, связанные с содержанием, не имеют принципиального значения для моего обсуждения, так что здесь я лишь кратко коснусь их, хотя они относятся к числу наиболее глубоких и тонких вопросов об опыте и заслуживают гораздо более детального рассмотрения в другом месте.

Главной характеристикой содержательной стороны осведомленности и опыта является то, что содержание здесь, как правило, неконцептуально — это содержание, не требующее, чтобы агент обладал понятиями, которые могли бы использоваться для характеристики этого содержания. Правдоподобным, к примеру, кажется, что простая система — к примеру,

собака или мышь — могла бы иметь детализированные переживания цвета с соответствующими детализированными репрезентациями цветовых различий в когнитивной системе, обладая при этом лишь простейшей системой цветовых понятий. Сходным образом и у людей состояния сознания и осведомленности при восприятии музыки нередко оказываются наделены содержаниями, выходящими далеко за пределы имеющихся у них музыкальных понятий.

(Обсуждение неконцептуального содержания см. в (Стапе 1992, Cussins 1990, Evans 1982, Peacocke 1992.) В литературе, похоже, есть консенсус относительно того, что содержания опыта неконцептуальны. Исключением является позиция Макдауэлла (McDowell 1994), доказывающего на основании нашей способности реидентифицировать переживания с помощью таких понятий, как «этот оттенок», что всякое опытное содержание концептуально. Несомненно, нужна ли эта способность для обладания переживаниями: правдоподобным, к примеру, кажется, что некоторые тонкие аспекты музыкального опыта у некоторых субъектов (к примеру, небольшие изменения тональности) могли бы вообще не допускать концептуализации и реидентификации. Другим примером являются переживания животных. Макдауэлл, похоже, с радостью принимает вывод о том, что животные лишены опытных переживаний, но кому-то могло бы показаться, что в данном случае *modus tollens* выглядел бы по крайней мере не менее убедительным, чем *modus ponens*. И даже если бы пришлось принять точку зрения Макдауэлла, думаю, что некое подобие соответствующей дистинкции могло бы быть восстановлено под видом двух степеней концептуального содержания.)

Конечно, между понятиями и сознанием может иметься какое-то causalное отношение; концептуальные изменения нередко оказывают значительное влияние на характер опыта. Но подобные концептуальные ресурсы не кажутся тем, что требуется для сознательного опыта. Это верно и для осведомленности, в той мере, в какой она параллельна сознанию. Содержания, репрезентированные первопорядковыми регистрациями, соответствующими сознательным переживаниям, к примеру, визуальном восприятии не требуют соответствующих богатых концептуальных ресурсов. Так что содержания опыта и осведомленности в целом более прimitивны, чем содержания суждений, которые при наиболее естественном рассмотрении должны трактоваться как концептуальные.

Одним из наиболее интересных вопросов является вопрос, присущ ли репрезентационный контент опыту внутренним образом, или же это содержание каким-то образом порождается фундирующими опытом когнитивным состоянием. Вторая позиция может казаться привлекательной, но, похоже, не вполне верна: представляется, к примеру, что мой визуальный опыт в данный момент репрезентирует мир, в котором есть большой квадратный объект напротив меня, и он делает это просто благодаря его бытию в качестве такого опыта. Даже гипотетический бестелесный ум со сходным опытом имел бы похожее репрезентативное содержание. Сиверт (Siewert 1994) убедительно показывает, что по своей природе опыт информативен относительно состояния мира: визуальный опыт, к примеру, можно оце-

нивать в плане его точности (он может препрезентировать мир правильно и неправильно), причем в силу самой его природы визуального опыта. Утверждение, что опыт внутренне нагружен препрезентативным содержанием, может быть поэтому оправданным.

Могло бы возникнуть искушение занять обратную позицию и принять, что подлинное содержание имеется только в опыте, а содержание фундирующей первопорядковой регистрации само зависит от содержания связанного с ней опыта. В этом могло бы что-то быть, но это тоже не вполне удовлетворительно; ведь и о первопорядковых регистрациях зомби мы хотим сказать, что они в известном смысле препрезентируют мир существующим определенным образом. У нас точно есть содержательные состояния, не связанные с опытными переживаниями, и трудно принять, что все наши содержания каким-то образом зависят от содержания опыта. Промежуточная позиция, которую можно было бы занять, состоит в том, что (1) изначальное содержание в известном смысле находилось в опыте, но (2) мы развили у себя схему атрибутирования содержания когнитивным состояниям, опираясь, в частности, на когерентность содержаний связанных с ними переживаний, и что (3) после возникновения этой схемы она стала автономной, так что мы можем говорить о содержании когнитивных состояний даже при отсутствии опыта. Это означало бы, что опытные переживания и связанные с ними регистрации могли бы иметь самостоятельные содержания, при том, что здесь не было бы странной, случайной сверхдетерминации, посредством которой дважды конституировалось бы одно и то же содержание. Это весьма тонкие вопросы, которые скорее всего заместили бы их детальное рассмотрение.

Еще одна интересная проблема — является ли соответствующий тип содержания «широким» (зависящим от окружающих объектов) или «узким» (зависящим только от внутренних процессов)? Если опыт внутренне нагружен содержанием, и если опыт супервентен на организации субъекта, то релевантное содержание в данном случае должно быть узким. (Опытные переживания по-прежнему могли бы иметь широкое содержание, но оно не могло бы быть содержанием, фиксируемым одним только опытом.) Иногда утверждалось, что подлинно препрезентативным содержанием является лишь широкое содержание, но, на мой взгляд, существует естественный путь понимания узкого препрезентативного содержания (см. Chalmers 1994c). Можно было бы доработать подобную концепцию до объяснения узкого, неконцептуального содержания опыта и осведомленности как содержания, накладывающего ограничения на центрированные миры, являющиеся кандидатами на роль действительного мира данного субъекта. Другой вопрос, имеют ли препрезентативное содержание все опытные переживания? Многие или большинство, вероятно, имеют его; большинство перцептивных переживаний уж точно кажутся внутренне информативными относительно мира. Имеется, однако, ряд более запутанных случаев: как быть с оргазмами, тошнотой или некоторыми эмоциональными переживаниями (см. Block 1995 и Tye 1995)? Впрочем, даже в этих случаях можно было бы отыскать какое-то препрезентативное содержание, так как эти переживания зачастую содержательны относи-

тельно локализации (вот здесь, там внизу) или их качества (хорошие, плохие). Возможность опытных переживаний, полностью лишенных репрезентативного содержания, неочевидна; с другой стороны, неочевидно и обратное.

Некоторые философы полагали, что феноменальные свойства — это всего лишь репрезентативные свойства, так что опытные переживания исчерпываются их репрезентативным содержанием (напр., Dretske 1995; Нагман 1990; Lycan 1996; Тье 1992). Чаще всего эта позиция сопровождается редуктивным взглядом на репрезентативное содержание, так что она сводится к одной из разновидностей редуктивного функционализма и малоправдоподобна по обычным причинам. Другая вариация этой идеи могла бы совмещать ее с нередуктивным взглядом на репрезентативное содержание, к примеру, с тем, согласно которому подлинную репрезентацию можно найти только в опыте. Это было бы в большей степени совместимо с серьезным отношением к сознанию, но по-прежнему было бы не лишено трудностей. В частности, создавалось бы впечатление, что репрезентативное содержание могло бы оставаться неизменным у функциональных изоморфов с инвертированным спектром, и в этом случае феноменология выходила бы за пределы репрезентативного содержания. Упомянутые выше случаи также наводят на мысль, что даже если все опытные переживания имеют репрезентативное содержание, у них имеются и черты, не вмещающие их репрезентативным содержанием. Так что неясно, будет ли успешной даже нередуктивная версия рассматриваемой нами концепции.

11. В частном общении и в готовящихся к публикации работах.
12. Упомянутые мной принципы иногда четко обозначались в качестве одного из методологических компонентов эмпирического исследования ментального. Не удивительно, что чаще всего это случалось в той области мейнстримной психологии, которая более всего затрагивает сознательный опыт, а именно в психофизике. Ее задачу нередко видели в соотнесении свойств наших ощущений со свойствами связанных с ними физических стимулов. Типичными достижениями здесь являются закон Вебера—Фехнера и закон Стивенса (Stevens 1975), которые дают две формулы соотношения интенсивности стимула и интенсивности соответствующего ощущения. Хотя иногда утверждается, что в психофизике подлежат объяснению прежде всего данные от третьего лица, такие как субъективные отчеты, кажется очевидным, что к числу главных феноменов, которые пытаются объяснить психофизика, относятся характеристики опыта от первого лица, вроде переживаний определенных оптических иллюзий.

Хорст (Horst 1995) убедительно показывает, что исходными данными в этой области нередко оказываются переживания от первого лица различных феноменов, таких как иллюзии. К примеру, на конференциях исследователи придают особое значение возможности самостоятельно «увидеть» различные эффекты. Можно было бы также попробовать показать, что попытки выяснить, какой из подходов к измерению ощущения — Фехнера или Стивенса (см. Stevens 1975) — верен, имеют смысл лишь при допущении, что они стремятся измерить что-то одно, феноменаль-

ный опыт; в ином случае мы просто имеем неконфликтующие измерения разных функциональных феноменов).

В психофизике время от времени возникали дискуссии на тему того, каким образом эмпирические наблюдения могут содействовать объяснению субъективных ощущений. Некоторые исследователи пришли к формализации четких принципов, на которых основана подобная работа — известных под разными именами, как «психофизические связующие гипотезы» (Brindley 1960) или «общие связующие пропозиции» (Teller 1984). Хороший пример — «аксиомы психофизической корреляции», сформулированные Мюллером (Müller 1896; цитируется в (Boring 1942, с. 89)):

- (1) основанием любого состояния сознания является материальный процесс, так называемый психофизический процесс, протекание которого связано с присутствием сознательного состояния;
- (2) равенству, сходству или различию в устроении ощущений... соответствует равенство, сходство и различие в устроении психофизического процесса, и наоборот. Более того, большему или меньшему сходству ощущений также соответствует большее или меньшее сходство психофизического процесса, и наоборот;
- (3) если изменения ощущения имеют одну и ту же направленность, или если различия, существующие между рядами ощущений, имеют сходную направленность, то изменения психофизического процесса, или различия данного психофизического процесса, имеют сходную направленность. Более того, если ощущение может изменяться в *n* направлениях, то функционирующий его психофизический процесс тоже должен быть изменчив в *n* направлениях, и наоборот.

Ясно, что эти принципы близки принципу структурной когерентности. Сделав поправку на определенные различия в языке, мы можем увидеть, что все утверждения в духе тех, что были приведены выше, являются прямыми следствиями принципа когерентности; и в их совокупности они приближаются к его потенциальному. Так что мы опять-таки можем убедиться в том, что принцип структурной когерентности и его вариации играют ключевую роль в использовании эмпирических исследований для создания объяснятельных концепций различных характеристик опыта.

Не удивительно, что в психофизике обсуждался статус подобных принципов и это обсуждение было параллельно аналогичным дебатам в философии сознания (см., напр., Brindley 1960; Marks 1978; D. Teller 1984, 1990). Некоторые рассматривали их как эмпирические гипотезы, но они, похоже, не извлечены из эмпирических тестов и не фальсифицируются ими, по крайней мере если говорить о той разновидности последних, где речь идет о позиции третьего лица. Другие, особенно операционистского склада, считали их дефиниторными допущениями; это соответствует редуктивной функционалистской позиции в философии. Зачастую они трактовались просто как фоновые допущения, или посылки, касающиеся природы психофизического соединения. В любом случае науке удавалось достаточно успешно решать свои задачи без реального разрешения этих

вопросов. Для объяснительных целей важнее форма этой соединительной конструкции, чем ее метафизический статус.

В целом «философские» проблемы, связанные с соотношением физических процессов и опыта, составляют непосредственный фон многих теоретических дискуссий в психофизике. Насколько мне известно, это еще не стало предметом подробного обсуждения в философской литературе (хотя см. (Savage 1970) для философской критики методологии измерения ощущений). И похоже, эта тема заслуживает детального исследования.

Глава 7

1. Напр., (Lycan 1987).
2. Черчленды (Churchland and Churchland 1981) возражали аргументам «Китайской нации» на том основании, что подобная система должна была бы обрабатывать приблизительно с $10^{3000000}$ стимулов сетчатки, и еще более громадное количество внутренних состояний мозга. Симуляция с помощью народонаселения, требующая одного человека для одного стимула на входе и одного для состояния, нуждалась бы поэтому в таком количестве людей, которое было бы несопоставимо больше того, которое могло бы предоставить население.
- Это возражение игнорирует тот факт, что как данные на входе, так и внутренние состояния комбинаторно структурированы. Вместо того, чтобы представить каждый входящий паттерн (более 10^8 клеток) одним человеком, что потребовало бы 2^{10^8} людей, нам нужно лишь 10^8 людей для представления данных, рассматриваемых в качестве структурированного паттерна. Это же справедливо и для внутренних состояний. Количество требующихся нам людей, таким образом, не превышает числа клеток в мозге.
3. Боген (Bogen 1981) и Лайкан (Lycan 1987) высказывали предположение, что в подобных «случайных» ситуациях отсутствовали бы квалиа, так как квалиа нуждаются в телеологии. Странным последствием этого была бы зависимость наличия или отсутствия квалиа от истории системы. Думаю, лучше признать, что система такого рода имела бы квалиа, подчеркнув крайнюю маловероятность ее случайного возникновения.
4. Более подробно я обсуждаю это в (Chalmers 1994a).
5. Это число получается, если обратить внимание на то, что существует 10^{10^9} возможных вариаций консеквента каждого кондиционала, представляющего глобальное состояние, в которое переходит система. Шанс того, что данное глобальное состояние перейдет в нужное последующее состояние, равен поэтому 1 из 10^{10^9} . На деле он будет меньше, так как любое глобальное состояние будет реализуемо множеством различных «максимальных» состояний физической системы, каждое из которых должно осуществить надлежащий переход. Должно удовлетворяться 10^{10^9} таких кондиционалов, так что указанная пропорция уменьшается.

6. Некоторые сходные сценарии можно найти также в (Harrison 1981a, 1981b).
7. Быть может, можно избавиться от одного из компонентов этой ситуации, объяснив конституирование убеждения опытом, — как оно обсуждается в 7 параграфе главы 5: не исключено, что понятие «опыта красного», имеющееся у Джо, отсылает теперь, к примеру, к переживанию розового, так что нельзя сказать, что он совсем неправ. Впрочем, эта стратегия уж точно не поможет объяснить его ошибки в суждениях о различиях.
8. Када (Cuda 1985) утверждает, что описание систем с подобными ошибочными убеждениями не имеет смысла. Он не приводит никаких аргументов помимо тезиса о том, что если бы такое описание имело смысл, то не бессмысленным было бы полагать, что мы ошибаемся подобным образом, что (как он утверждает) явно не соответствует действительности. Однако это кажется ошибочным. Предположение о том, что я мог бы ошибаться подобным образом, не лишено смысла, так что в этом плане данная гипотеза когерентна, и дело лишь в том, что моя эпистемическая ситуация показывает мне, что эта гипотеза не является истинной в моем собственном случае, так как я напрямую переживаю ярко красные квалиа и т. п.
9. Это не значит, что в этой области вообще не может быть соритов; см., напр., (Tienson 1987).
10. Данная позиция ближе всего Шумейкеру (Shoemaker 1982), но ее защищали также Хорган (Horgan 1984a), Патнэм (Putnam 1981) и целый ряд других авторов.
11. Не исключено даже, что подобные случаи существуют в действительном мире. В интересной статье Нида-Рюмелин (Nida-Rümelin 1996) отмечает, что исследования в области цветного зрения приводят нас к допущению реальной возможности случаев «псевдонормального» цветного зрения, при котором R-колбочки сетчатки будут связаны с реактивным паттернами, обычно присущими G-колбочкам, а (2) G-колбочки будут связаны с реактивным паттернами, обычно присущими R-колбочкам. Если рассматривать эти ситуации, (1) и (2), по отдельности, то это будут типичные случаи слепоты к красному и зеленому. Теоретически, генетические отклонения, ответственные за (1) и (2), могут сосуществовать, и тогда возникнет субъект, который в плане своего поведения будет очень похож на обычного субъекта, но который, возможно, будет обладать инвертированными в сравнении с другими людьми переживаниями цвета.
12. Патнэм (Putnam 1981) и Шумейкер (Shoemaker 1982) используют этот пример как аргумент против функционалистских объяснений квалиа, но их аргументы в лучшем случае могут сработать против малодетализированного принципа инвариантности, согласно которому считается, что один и тот же тип опыта возникает, к примеру, из состояний, активируемых синими объектами и ведущими к отчетам о «синем». Высокодетализированный принцип остается незатронутым этой критикой. (Сходный тезис высказывает Левин (Levine 1988).)

В одной из разновидностей этого сценария субъекты оказываются включены в процесс адаптации, обучения и в конечном счете амнезии

(забывая о том, что вещи когда-то выглядели иначе), оказываясь в итоге поведенчески идентичными с их изначальным состоянием. Нет, однако, серьезных оснований считать, что они будут идентичными и в организационном плане, особенно если учесть, что их перепрошивка по-прежнему отражается на состоянии их мозга. Если вследствие протекания какого-то особого процесса их организация оказывается в итоге неотличимой от изначальной, то не показалось бы невероятным предположение, что и их переживания также вернулись бы к своему исходному виду (Коул (Cole 1990) и Рей (Rey 1992) отстаивают вариации на темы гипотезы возвращения к исходному состоянию в данном случае).

13. Сходный аргумент выдвигает Сигер (Seager 1991, с. 39–41), описывающий случай, когда клетки сетчатки «настраиваются» на более высокий диапазон оптического спектра. С этим аргументом можно работать так же, как с аргументом Блока.
14. Блок отвечает на сходное возражение, отмечая, что мы можем перенести «линзы» внутрь системы, перепрошивая ее, к примеру в зрительном нерве или в зрительной коре. Но, как и прежде, это не дает нам случаев организационных изоморфов с различными переживаниями. Действительно, описание Блоком подобных случаев кажется напрямую совместимым с позицией, в соответствии с которой квалиа зависят от организации «центральных» систем. Нас просят верить в неизменность переживаний в определенных случаях именно из-за незатронутости центральных процессов; и предполагается, что мы будем верить в различие переживаний при различии центральных процессов. Такие аргументы не могут опровергнуть принцип инвариантности.
15. Отдаленным поводом для формулировки аргумента в этом параграфе послужила деннетовская история «Где я?» (Dennett 1978d). Ситуацию, в чем-то напоминающую описываемую мной ниже, рассматривает Шумейкер (Shoemaker 1982). Еще ближе к делу дискуссия Сигера (Seager 1991, с. 43), хотя Сигер и не защищает принцип инвариантности. Когда эта книга уже уходила в печать, я обнаружил интересную недавнюю статью Арнольда Зубова (Zuboff 1994), который использует аргумент, по сути, являющийся аргументом от скачущих квалиа, для подкрепления одной из разновидностей редуктивного функционализма, доказывая, что скачущие квалиа являются априори невозможными.
16. Подобный тезис высказывает Уайт (White 1986), указывая, что если нефункциональные физические различия релевантны для квалиа, то на квалиа могли бы влиять даже небольшие изменения в ДНК.
17. Шумейкер (Shoemaker 1982) выдвигает комплексный критерий того, насколько специфичным должно быть физиологическое свойство для фиксации квалиа, или, как он выражается, «реализации квале». Мне, однако, кажется, что если то, что я здесь говорил, верно, то его критерий фактически будет выделять высокодетализированное функциональное свойство.
18. Эта идея была выдвинута Терри Хорганом в частной беседе.

Глава 8

1. О связи между сознанием и информацией говорили также Бом (Bohm 1980), Сейр (Sayre 1976) и Вилманс (Velmans 1991), хотя детали их концепций существенно отличаются от того, что говорю я. Сейровская идея «нейтрального монизма» информации, впрочем, весьма плодотворна (спасибо Стиву Хорсту за это указание). Некое подобие двуаспектного принципа обсуждается Локвудом (Lockwood 1989, гл. 11), хотя он и не формулирует свои тезисы в терминах информации.
2. В неопубликованной статье (Chalmers 1990) я фокусируюсь на этой стратегии для уяснения отношения между сознанием и суждениями о сознании и использую ее для формулировки базовой «теории» сознания (содержащей отсылку к паттерну и информации), предшественницы ряда идей этой главы. В этой статье я называю требование объяснительной когерентности «тестом на когерентность», который должна проходить любая теория сознания.
3. Хотя относительно наличия у термостатов убеждений и желаний — см. (McCarthy 1979).
4. Уилер уделяет особое внимание результатам измерений, или «ответам на да-нет вопросы» как основе всего, и в этом плане его подход может быть ближе идеализму, чем отстаиваемая здесь мной позиция.
5. См. также интересное изложение идей Фридкина в (Wright 1998) для детализации базовой метафизики.
6. Плодотворные обсуждения этих проблем расселовских воззрений см. в (Foster 1991, с. 119–130) и (Lockwood 1992).
7. Локвуд (Lockwood 1992) полагает, что расселовская позиция может отсыпать к элементарным законам в этих целях. Он не рассматривает возражение, состоящее в том, что введение дополнительных законов, похоже, лишает расселовский подход его изначальной привлекательности. Так, они поднимают проблемы эпифеноменализма, которых обещала избежать расселовская концепция, и они также требуют существенного расширения онтологии за пределы внутренних свойств, нужных для фундирования физики. (Должен отметить, что Локвуд не опирается на эти законы для решения проблемы детализации; главная его идея в связи с решением этой проблемы — интригующая обосновленная гипотеза, связанная с квантовой механикой.)

Глава 9

1. Материал этой главы во многом взят из (Chalmers 1994a).
2. Патнэм (Putnam 1988, с. 120–125) приводит свой аргумент в пользу того, что любая обычная открытая система имплементирует любой конечный автомат. Я детально анализирую этот аргумент в (Chalmers 1995a). По итогам этого анализа выясняется, что данный аргумент эффективен, похоже, лишь при допущении модальной мягкости кондиционалов перехода физических состояний в дефиниции имплементации.

3. Я провожу эту линию понимания объяснительной роли вычисления в когнитивной науке в (Chalmers 1994b).
4. Сходные тезисы высказаны в (Korb 1991) и (Newton 1989). Они считают, что Китайская комната может быть хорошим аргументом если не против машинной интенциональности, то против машинного сознания.
5. Хоффстадтер (Hofstadter 1981) очерчивает сходный спектр промежуточных случаев между мозгом и Китайской комнатой.
6. Идея о том, что гомункул в Китайской комнате аналогичен демону, мечущемуся внутри черепной коробки, была высказана Хауглендом (Haugeland 1980).
7. Примечательно, что хотя Дрейфус (Dreyfus 1972) назвал свою книгу, в которой делается подобное возражение, «Чего не могут компьютеры», впоследствии он признал, что надлежащий тип вычислительной системы (такой как коннекционистские системы) не подпадал бы под эти возражения. На деле «возможности компьютера» отождествляется с тем, чего может достичь очень узкий класс вычислительных систем.
8. Насколько мне известно, это четкое возражение против геделевских аргументов впервые в печатных изданиях было высказано Патнэмом (Putnam 1960), и, насколько я знаю, оно никогда не было опровергнуто, несмотря на отчаянные усилия Лукаса и Пенроуза. Пенроуз (Penrose 1994, гл. 3.3) доказывает, что он может установить непротиворечивость формальной системы, ухватывающей его собственные рассуждения, поскольку он мог бы уверенно определить истинность аксиом и верность правил вывода. Это, похоже, зависит от изначального допущения, что вычислительная система является системой аксиом и правил, каковой она, однако, не обязана быть в общем случае (о чем свидетельствует нейронная симуляция мозга). Даже в случае системы с аксиомами и правилами мне не очевидно, что мы сможем определить значимость каждого правила, которым могла бы воспользоваться наша система, особенно если речь идет о таких правилах, применение которых выходит за пределы обычного вычисления и касается повторных геделизаций, относительно которых геделевские аргументы действительно имели бы реальный вес в случае, когда речь идет о людях.
9. Возможно, этим стоит заняться для учета случая, когда конкретная схема округления на уровне 10^{-10} дает поведенческий уклон при распределении. Для надежности — при допущении шума на уровне 10^{-10} — мы должны аппроксимировать систему на уровне 10^{-20} , аппроксимируя на этом уровне и распределение шума.
10. В некоторых случаях — как правило, только в философии сознания — такие термины, как «вычисление» используются для отсылки исключительно к классу символьных вычислений, или вычислений, производимых над представлениями (то есть системами, базовые синтаксические элементы в которых являются в то же время базовыми семантическими объектами). Разумеется, этот терминологический вопрос мало на что влияет: с позиции искусственного интеллекта важно то, чтобы в наличии имелась некая формальная система — такого рода, чтобы ее имплементация была достаточна для ментальности, — и неважно, считается ли

она «вычислением» согласно этому критерию. Следует, однако, отметить, что в любом случае использовать данный термин подобным образом — значит порывать связи с его истоками в теории вычисления. Даже большинство машин Тьюринга не будут считаться «вычислительными» в этом смысле, так как лишь некоторые из них могут быть интерпретированы в качестве таких, которые производят вычисления над концептуальными презентациями. По сходным причинам подобное ограничение класса «вычислений» приводит к утрате (черче-тьюринговой) универсальности вычисления, которая, собственно, является, возможно, одним из самых серьезных оснований для доверия к (функциональному) тезису ИИ.

Глава 10

1. По крайней мере, у таких частиц со спином $1/2$, как электрон. Я оставляю в стороне случаи, когда у спина имеются другие базовые характеристики.
2. Здесь, как и в других местах, я иду на упрощение картины. Никакое измерение не является абсолютно точным, так что в его результате никогда не возникает состояние, положение которого является в полном смысле определенным. На деле волновая функция коллапсирует в состояние, вся амплитуда которого сконцентрирована в очень узком диапазоне локаций. Проще говорить так, однако, будто коллапсированные положения являются в полной мере определенными.
3. Или плотность вероятности в случае континуума.
4. Алберт (Albert 1992) указывает, что «сознание» столь же неопределенno, как «измерение» и «макроскопическое»; но мне кажется, что привлекательность этого критерия отчасти связана с тем, что здесь, похоже, можно говорить о фактичности того, является ли та или иная система сознательной.
5. Мое обсуждение ЖРВ-интерпретации опирается на (Albert and Loewer 1990) и (Albert 1992).
6. Заметим, что это единственное место в данной главе, где мы касаемся теоремы Белла и результатов Эйнштейна—Подольского—Розена (ЭПР). Иногда эти результаты рассматриваются в качестве главного источника философских проблем, связанных с квантовой механикой, но, на мой взгляд, эти проблемы возникают еще до соображений ЭПР. Даже без ЭПР мы стояли бы перед трудным выбором между коллапсом, скрытыми параметрами и Эвереттом. ЭПР просто добавляют трудностей теориям скрытых параметров, показывая, что они (как и коллапс) должны быть нелокальными; и, вполне возможно, повышают привлекательность интерпретации Эверетта, единственной локальной интерпретации совместимой с данным результатом.
7. Интерпретация одного большого мира, похоже, является наиболее распространенным пониманием интерпретации Эверетта среди физиков (особенно среди квантовых космологов, постоянно использующих эту модель). Идея «расщепляющихся миров» — во многом продукт популяризации. Иногда даже сторонники концепции одного большого мира

говорят о «расщеплении», но это лишь фигура речи, отсылающая к факту появления суперпозиции волновой функции. Особого процесса расщепления миров не существует; в лучшем случае происходит некое локальное расщепление волновой функции. В любом случае, на мой взгляд, лучше избегать разговоров о «расщеплении», так как это неизбежно порождает путаницу.

8. Это возражение высказывалось Беллом (Bell 1981), Беллом и Хайли (Bell and Hiley 1993), Ходгсоном (Hodgson 1988) и многими другими.
9. Самой разумной стратегией, возможно, было бы сделать ставку через квантовое устройство, отвечающее «нет» с вероятностью 0,999, «да» — с вероятностью 0,001. Тогда, если теория Эверетта неверна, я почти на верняка буду прав, а если верна, то выживет хотя бы одно из сознаний, являющихся моими отпрысками.

Библиография

- Ackerman D. 1990. A Natural History of the Senses. New York: Random House.
- Adams R. M. 1974. Theories of actuality. *Nous* 8:211–31.
- Akins K. 1993. What is it like to be boring and myopic? In B. Dahlbom, ed., Dennett and His Critics. Oxford: Blackwell.
- Albert D. 1992. Quantum Mechanics and Experience. Cambridge, Mass: Harvard University Press.
- Albert D., and B. Loewer. 1988. Interpreting the many-worlds interpretation. *Synthese* 77:195–213.
- . 1989. Two no-collapse interpretations of quantum mechanics. *Nous* 23: 169–86.
- . 1990. Wanted dead or alive: Two attempts to solve Schrödinger's paradox. *PSA* 1990, vol. 1, pp. 277–85.
- Alexander S. 1920. Space, Time, and Deity. London: Macmillan.
- Armstrong D. M. 1968. A Materialist Theory of the Mind. London: Routledge and Kegan Paul.
- . 1973. Belief, Truth, and Knowledge. Cambridge: Cambridge University Press.
- . 1981. What is consciousness? In *The Nature of Mind*. Ithaca, N. Y.: Cornell University Press.
- . 1982. Metaphysics and supervenience. *Critica* 42:3–17.
- . 1983. What Is a Law of Nature? Cambridge: Cambridge University Press.
- . 1990. A Combinatorial Theory of Possibility. Cambridge: Cambridge University Press.
- Austin D. F. 1990. What's the Meaning of «This». Ithaca, N. Y.: Cornell University Press.
- Baars B. J. 1988. A Cognitive Theory of Consciousness. Cambridge: Cambridge University Press.
- Bacon J. 1986. Supervenience, necessary coextension, and reducibility. *Philosophical Studies* 49:163–76.
- Barwise J., and J. Perry. 1983. Situations and Attitudes. Cambridge, Mass.: MIT Press.
- Bateson G. 1972. Steps to an Ecology of Mind. San Francisco: Chandler.
- Bealer G. 1994. Mental properties. *Journal of Philosophy* 91:185–208.
- Bell J. S. 1964. On the Einstein—Podolsky—Rosen paradox. *Physics* 1:195–200. [Reprinted in Bell 1987b]
- . 1976. The measurement theory of Everett and de Broglie's pilot wave. In M. Flato, ed., Quantum Mechanics, Determinism, Causality, and Particles. Dordrecht: Reidel. [Reprinted in Bell 1987b]
- . 1981. Quantum mechanics for cosmologists. In C. Isham, R. Penrose, and D. Sciama, eds., Quantum Gravity. Vol. 2. Oxford: Oxford University Press. [Reprinted in Bell 1987b]
- . 1987a. Are there quantum jumps? In Schrödinger: Centenary of a Polymath. Cambridge: Cambridge University Press.
- . 1987b. Speakable and Unspeakable in Quantum Mechanics. Cambridge: Cambridge University Press.
- Bigelow J., and R. Pargetter. 1990. Acquaintance with qualia. *Theoria* 56:129–47.
- Bisiach E. 1988. The (haunted) brain and consciousness. In A. Marcel and E. Bisiach, eds., Consciousness in Contemporary Science. Oxford: Oxford University Press.

- Blackburn S.* 1971. Moral realism. In J. Casey, ed., *Morality and Moral Reasoning*. London: Methuen.
- . 1985. Supervenience revisited. In I. Hacking, ed., *Exercises in Analysis: Essays by Students of Casimir Lewy*. Cambridge: Cambridge University Press.
- . 1990. Filling in space. *Analysis* 50:62–65.
- Block N.* 1978. Troubles with functionalism. In C. W. Savage, ed., *Perception and Cognition: Issues in the Foundation of Psychology*. Minneapolis: University of Minnesota Press. [Reprinted in N. Block, ed., *Readings in the Philosophy of Psychology*. Vol. 1. Cambridge, Mass.: Harvard University Press, 1980]
- . 1980. What is functionalism? In N. Block, ed., *Readings in the Philosophy of Psychology*. Vol. 1. Cambridge, Mass.: Harvard University Press.
- . 1981. Psychologism and behaviorism. *Philosophical Review* 90:5–43.
- . 1990. Inverted earth. *Philosophical Perspectives* 4:53–79.
- . 1995. On a confusion about a function of consciousness. *Behavioral and Brain Sciences* 18:227–47.
- Boden M.* 1988. Escaping from the Chinese Room. In *Computer Models of Mind*. Cambridge: Cambridge University Press.
- Bogen J.* 1981. Agony in the schools. *Canadian Journal of Philosophy* 11:1–21.
- Bohm D.* 1952. A suggested interpretation of quantum mechanics in terms of «hidden variables», pts. 1 and 2. *Physical Review* 85:166–93.
- . 1980. Wholeness and the Implicate Order. London: Routledge.
- Bohm D., and B. Hiley.* 1993. *The Undivided Universe: An Ontological Interpretation of Quantum Theory*. London: Routledge.
- Boring E. G.* 1942. *Sensation and Perception in the History of Experimental Psychology*. New York: Appleton-Century-Crofts.
- Boyd R. N.* 1980. Materialism without reductionism: What physicalism does not entail. In N. Block, ed., *Readings in the Philosophy of Psychology*. Vol. 1. Cambridge, Mass.: Harvard University Press.
- . 1988. How to be a moral realist. In G. Sayre-McCord, ed., *Essays on Moral Realism*. Ithaca, N. Y.: Cornell University Press.
- Brindley G. S.* 1960. *Physiology of the Retina and Visual Pathway*. London: Edward Arnold.
- Brink D.* 1989. *Moral Realism and the Foundations of Ethics*. Cambridge: Cambridge University Press.
- Broad C. D.* 1925. *Mind and Its Place in Nature*. London: Routledge and Kegan Paul.
- Brooks D. H. M.* 1994. How to perform a reduction. *Philosophy and Phenomenological Research* 54:803–14.
- Byrne A.* 1993. The emergent mind. Ph. D. diss., Princeton University.
- Campbell K. K.* 1970. *Body and Mind*. New York: Doubleday.
- Carroll J. W.* 1990. The Humean tradition. *Philosophical Review* 99:185–219.
- . 1994. Laws of Nature. Cambridge: Cambridge University Press.
- Carruthers P.* 1992. Consciousness and concepts. *Proceedings of the Aristotelian Society*, suppl., 66:41–59.
- Chalmers D. J.* 1990. Consciousness and cognition. Technical Report 38, Center for Research on Concepts and Cognition, Indiana University.
- . 1994a. On implementing a computation. *Minds and Machines* 4:391–402.
- . 1994b. A computational foundation for the study of cognition. PNP Technical Report 94–03, Washington University.
- . 1994c. The components of content. PNP Technical Report 94–04, Washington University. [<http://www.artsci.wustl.edu/nphilos/pnp.html>

- . 1995a. Does a rock implement every finite state automaton? *Synthese*.
- . 1995b. Facing up to the problem of consciousness. *Journal of Consciousness Studies* 2:200–219. [Also in S. Hameroff, A. Kaszniak, and A. Scott, eds., *Toward a Science of Consciousness*. Cambridge, Mass.: MIT Press, 1996]
- . 1995c. Minds, machines, and mathematics. *PSYCHE* 2:1.
- . 1995d. The puzzle of conscious experience. *Scientific American* 273:80–86.
- Cheney D. L., and R. M. Seyfarth.* 1990. *How Monkeys See the World*. Chicago: University of Chicago Press.
- Chisholm R.* 1957. *Perceiving*. Ithaca, N. Y.: Cornell University Press.
- Churchland P. M.* 1985. Reduction, qualia and the direct introspection of brain states. *Journal of Philosophy* 82:8–28.
- . 1995. *The Engine of Reason, the Seat of the Soul: A Philosophical Journey into the Brain*. Cambridge, Mass.: MIT Press.
- Churchland P. M., and P. S. Churchland.* 1981. Functionalism, qualia and intentionality. *Philosophical Topics* 12:121–32.
- Churchland P. S.* 1986. *Neurophilosophy: Toward a Unified Science of the Mind-Brain*. Cambridge, Mass.: MIT Press.
- . 1988. The significance of neuroscience for philosophy. *Trends in the Neurosciences* 11:304–7.
- Clark A.* 1986. Psychofunctionalism and chauvinism. *Philosophy of Science* 53: 535–59.
- . 1993. *Sensory Qualities*. Oxford: Oxford University Press.
- Cole D.* 1990. Functionalism and inverted spectra. *Synthese* 82:207–22.
- Conee E.* 1985a. Physicalism and phenomenal properties. *Philosophical Quarterly* 35:296–302.
- . 1985b. The possibility of absent qualia. *Philosophical Review* 94:345–66.
- Cowey A., and P. Stoerig.* 1992. Reflections on blindsight. In D. Milner and M. Rugg, eds., *The Neuropsychology of Consciousness*. London: Academic Press.
- Crane T.* 1992. The nonconceptual content of experience. In T. Crane, ed., *The Contents of Experience*. Cambridge: Cambridge University Press.
- Crick F. H. C.* 1994. *The Astonishing Hypothesis: The Scientific Search for the Soul*. New York: Scribner.
- Crick F. H. C., and C. Koch.* 1990. Towards a neurobiological theory of consciousness. *Seminars in the Neurosciences* 2:263–75.
- Cuda T.* 1985. Against neural chauvinism. *Philosophical Studies* 48:111–27.
- Cussins A.* 1990. The connectionist construction of concepts. In M. Boden, ed., *The Philosophy of Artificial Intelligence*. Oxford: Oxford University Press.
- Daneri A., A. Loinger, and G. M. Prosperi.* 1962. Quantum theory of measurement and ergodicity conditions. *Nuclear Physics* 33:297–319. [Reprinted in Wheeler and Zurek 1983]
- Davidson D.* 1970. Mental events. In L. Foster and J. Swanson, eds., *Experience and Theory*. London: Duckworth.
- Davies M. K., and I. L. Humberstone.* 1980. Two notions of necessity. *Philosophical Studies* 38:1–30.
- Dennett D. C.* 1969. *Content and Consciousness*. London: Routledge and Kegan Paul.
- . 1978a. *Brainstorms*. Cambridge, Mass.: MIT Press.
- . 1978b. Are dreams experiences? In Dennett 1978a.
- . 1978c. Toward a cognitive theory of consciousness. In Dennett 1978a.
- . 1978d. Where am I? In Dennett 1978a.
- . 1979. On the absence of phenomenology. In D. Gustafson and B. Tapscott, eds., *Body, Mind, and Method*. Dordrecht: Kluwer.
- . 1987. *The Intentional Stance*. Cambridge, Mass.: MIT Press.

- . 1988. Quining qualia. In A. Marcel and E. Bisiach, eds., *Consciousness in Contemporary Science*. Oxford: Oxford University Press.
- . 1991. *Consciousness Explained*. Boston: Little, Brown.
- . 1993a. Back from the drawing board. In B. Dahlbom, ed., *Dennett and His Critics*. Oxford: Blackwell.
- . 1993b. The message is: There is no medium. *Philosophy and Phenomenological Research* 53:19–31.
- Descartes R.* 1984. *The Philosophical Writings of Descartes*. Translated by J. Cottingham, R. Stoothoff, and D. Murdoch. Cambridge: Cambridge University Press.
- DeWitt B. S.* 1970. Quantum mechanics and reality. *Physics Today* 23:30–35. [Reprinted in DeWitt and Graham 1973]
- . 1971. The many-universes interpretation of quantum mechanics. In B. d'Espagnat, ed., *Foundations of Quantum Mechanics*. New York: Academic Press. [Reprinted in DeWitt and Graham 1973]
- DeWitt B. S., and N. Graham, eds.* 1973. *The Many-Worlds Interpretation of Quantum Mechanics*. Princeton: Princeton University Press.
- Dretske F. I.* 1977. Laws of nature. *Philosophy of Science* 44:248–68.
- . 1981. *Knowledge and the Flow of Information*. Cambridge, Mass.: MIT Press.
- . 1995. *Naturalizing the Mind*. Cambridge, Mass.: MIT Press.
- Dreyfus H.* 1972. *What Computers Can't Do*. New York: Harper & Row.
- Eccles J. C.* 1986. Do mental events cause neural events analogously to the probability fields of quantum mechanics? *Proceedings of the Royal Society of London* B227:411–28.
- Edelman G.* 1989. *The Remembered Present: A Biological Theory of Consciousness*. New York: Basic Books.
- . 1992. *Bright Air, Brilliant Fire*. New York: Basic Books.
- Elitzur A.* 1989. Consciousness and the incompleteness of the physical explanation of behavior. *Journal of Mind and Behavior* 10:1–20.
- Evans G.* 1979. Reference and contingency. *The Monist* 62:161–89.
- . 1982. *The Varieties of Reference*. Oxford: Oxford University Press.
- Everett H.* 1957. «Relative-state» formulations of quantum mechanics. *Reviews of Modern Physics* 29:454–62. [Reprinted in Wheeler and Zurek 1983]
- . 1973. The theory of the universal wave function. In B. S. deWitt and N. Graham, eds., *The Many-Worlds Interpretation of Quantum Mechanics*. Princeton: Princeton University Press.
- Farah M.* 1994. Visual perception and visual awareness after brain damage: A tutorial overview. In C. Umiltà and M. Moscovitch, eds., *Conscious and Nonconscious Information Processing: Attention and Performance 15*. Cambridge, Mass.: MIT Press.
- Farrell B. A.* 1950. Experience. *Mind* 59:170–98.
- Feigl H.* 1958. The «mental» and the «physical». In H. Feigl, M. Scriven, and G. Maxwell, eds., *Concepts, Theories, and the Mind-Body Problem*. Minnesota Studies in the Philosophy of Science, vol. 2. Minneapolis: University of Minnesota Press.
- Feldman F.* 1974. Kripke on the identity theory. *Journal of Philosophy* 71:665–76. *Field H.* 1973. Theory change and the indeterminacy of reference. *Journal of Philosophy* 40:762–81.
- Flanagan O.* 1992. *Consciousness Reconsidered*. Cambridge, Mass.: MIT Press.
- Fodor J. A.* 1980. Searle on what only brains can do. *Behavioral and Brain Sciences* 3:431–32.
- . 1987. Psychosemantics: The Problem of Meaning in the Philosophy of Mind. Cambridge, Mass.: MIT Press.
- . 1992. The big idea: Can there be a science of mind? *Times Literary Supplement*, July 3, pp. 5–7.
- Forrest P.* 1986. Ways worlds could be. *Australasian Journal of Philosophy* 64:15–24.

- Foss J. 1989. On the logic of what it is like to be a conscious subject. *Australasian Journal of Philosophy* 67:305–20.
- Foster J. 1991. The Immaterial Self: A Defence of the Cartesian Dualism Conception of Mind. London: Routledge.
- Fredkin E. 1990. Digital mechanics. *Physica* D45:254–70.
- Geach P. 1957. Mental Acts. London: Routledge and Kegan Paul.
- Gell-Mann M., and J. B. Hartle. 1990. Quantum mechanics in the light of quantum cosmology. In W. Zurek, ed., Complexity, Entropy, and the Physics of Information. Redwood City, Calif.: Addison-Wesley.
- Gert B. 1965. Imagination and verifiability. *Philosophical Studies* 16:44–47.
- Ghirardi G. C., A. Rimini, and T. Weber. 1986. Unified dynamics for microscopic and macroscopic systems. *Physical Review D* 34:470.
- Goldman A. 1986. Epistemology and Cognition. Cambridge, Mass.: Harvard University Press.
- . 1993. The psychology of folk psychology. *Behavioral and Brain Sciences* 16:15–28.
- Gunderson K. 1970. Asymmetries and mind-body perplexities. In M. Radner and S. Winokur, eds., Analyses of Theories and Methods of Physics and Psychology. Minnesota Studies in the Philosophy of Science, vol. 4. Minneapolis: University of Minnesota Press.
- Hameroff S. R. 1994. Quantum coherence in microtubules: A neural basis for an emergent consciousness? *Journal of Consciousness Studies* 1:91–118.
- Hardin C. L. 1987. Qualia and materialism: Closing the explanatory gap. *Philosophy and Phenomenological Research* 48:281–98.
- . 1988. Color for Philosophers: Unweaving the Rainbow. Indianapolis: Hackett.
- Hare R. M. 1952. The Language of Morals. Oxford: Clarendon Press.
- . 1984. Supervenience. *Proceedings of the Aristotelian Society*, suppl., 58:1–16.
- Harman G. 1982. Conceptual role semantics. *Notre Dame Journal of Formal Logic* 28:242–56.
- . 1990. The intrinsic quality of experience. *Philosophical Perspectives* 4:31–52.
- Harnad S. 1989. Minds, machines and Searle. *Journal of Experimental and Theoretical Artificial Intelligence* 1:5–25.
- Harrison B. 1967. On describing colors. *Inquiry* 10:38–52. 1973. *Form and Content*. Oxford: Blackwell.
- . 1973. *Form and Content*. Oxford: Blackwell.
- Harrison J. 1981a. Three philosophical fairy stories. *Ratio* 23:63–67.
- . 1981b. Gulliver's adventures in Fairyland. *Ratio* 23:158–64.
- Haugeland J. 1980. Programs, causal powers, and intentionality. *Behavioral and Brain Sciences* 4:432–33.
- . 1982. Weak supervenience. *American Philosophical Quarterly* 19:93–103.
- Healey R. A. 1984. How many worlds? *Nous* 18:591–616.
- Heil J. 1992. The Nature of True Minds. Cambridge: Cambridge University Press.
- Hellman G., and F. Thompson. 1975. Physicalism: Ontology, determination and reduction. *Journal of Philosophy* 72:551–64.
- Hill C. S. 1991. Sensations: A Defense of Type Materialism. Cambridge: Cambridge University Press.
- Hodgson D. 1988. The Mind Matters: Consciousness and Choice in a Quantum World. Oxford: Oxford University Press.
- Hofstadter D. R. 1979. Gödel, Escher, Bach: an Eternal Golden Braid. New York: Basic Books.
- . 1981. Reflections on Searle. In D. R. Hofstadter and D. C. Dennett, eds., *The Mind's I*. New York: Basic Books.
- . 1985a. Who shoves whom around inside the careenium? In *Metamagical Themas*. New York: Basic Books.

- . 1985b. Heisenberg's Uncertainty Principle and the many-worlds interpretation of quantum mechanics. In *Metamagical Themas*. New York: Basic Books.
- Honderich T.* 1981. Psychophysical law-like connections and their problems. *Inquiry* 24:277–303.
- Horgan T.* 1978. Supervenient bridge laws. *Philosophy of Science* 45:227–49.
- . 1982. Supervenience and microphysics. *Pacific Philosophical Quarterly* 63: 29–43.
- . 1984a. Functionalism, qualia, and the inverted spectrum. *Philosophy and Phenomenological Research* 44:453–69.
- . 1984b. Jackson on physical information and qualia. *Philosophical Quarterly* 34:147–83.
- . 1984c. Supervenience and cosmic hermeneutics. *Southern Journal of Philosophy*, suppl., 22:19–38.
- . 1987. Supervenient qualia. *Philosophical Review* 96:491–520.
- . 1993. From supervenience to superdupervenience: Meeting the demands of a material world. *Mind* 102:555–86.
- Horgan T., and M. Timmons.* 1992a. Troubles for new wave moral semantics: The «open question argument» revived. *Philosophical Papers*.
- . 1992b. Trouble on moral twin earth: Moral queerness revived. *Synthese* 92:223–60.
- Horst S.* 1995. Phenomenology and psychophysics. Manuscript, Wesleyan University.
- Hughes R. I. G.* 1989. The Structure and Interpretation of Quantum Mechanics. Cambridge, Mass.: Harvard University Press.
- Humphrey N.* 1992. A History of the Mind: Evolution and the Birth of Consciousness. New York: Simon and Schuster.
- Huxley T.* 1874. On the hypothesis that animals are automata. In *Collected Essays*. London, 1893–1894.
- Jackendoff R.* 1987. Consciousness and the Computational Mind. Cambridge, Mass.: MIT Press.
- Jackson F.* 1977. Perception. Cambridge: Cambridge University Press.
- . 1980. A note on physicalism and heat. *Australasian Journal of Philosophy* 58:26–34.
- . 1982. Epiphenomenal qualia. *Philosophical Quarterly* 32:127–36.
- . 1993. Armchair metaphysics. In J. O'Leary-Hawthorne and M. Michael, eds., *Philosophy in Mind*. Dordrecht: Kluwer.
- . 1994. Finding the mind in the natural world. In R. Casati, B. Smith, and G. White, eds., *Philosophy and the Cognitive Sciences*, Vienna: Holder-Pichler-Tempsky.
- . 1995. Postscript to «What Mary didn't know.» In P. K. Moser and J. D. Trout, eds., *Contemporary Materialism*. London: Routledge.
- Jacoby H.* 1990. Empirical functionalism and conceivability arguments. *Philosophical Psychology* 2:271–82.
- Jaynes J.* 1976. The Origins of Consciousness in the Breakdown of the Bicameral Mind. Boston: Houghton Mifflin.
- Johnson-Laird P.* 1983. A computational analysis of consciousness. *Cognition and Brain Theory* 6:499–508.
- Kaplan D.* 1979. Dthat. In P. Cole, ed., *Syntax and Semantics*. New York: Academic Press.
- . 1989. Demonstratives. In J. Almog, J. Perry, and H. Wettstein, ed., *Themes from Kaplan*. New York: Oxford University Press.
- Kim J.* 1978. Supervenience and nomological incommensurables. *American Philosophical Quarterly* 15:149–56.
- . 1984. Concepts of supervenience. *Philosophy and Phenomenological Research* 45:153–76.
- . 1985. Psychophysical laws. In B. McLaughlin and E. LePore, eds., *Action and Events*. Oxford: Blackwell.
- . 1989. Mechanism, purpose, and explanatory exclusion. *Philosophical Perspectives* 3:77–108.

- . 1993. *Supervenience and Mind*. Cambridge: Cambridge University Press.
- Kirk R.* 1974. Zombies versus materialists. *Aristotelian Society* 48(suppl.):135–52.
- . 1979. From physical explicability to full-blooded materialism. *Philosophical Quarterly* 29:229–37.
- . 1992. Consciousness and concepts. *Proceedings of the Aristotelian Society* 66(suppl.):23–40.
- . 1994. *Raw Feeling: A Philosophical Account of the Essence of Consciousness*. Oxford: Oxford University Press.
- Korb K.* 1991. Searle's AI program. *Journal of Experimental and Theoretical Artificial Intelligence* 3:283–96.
- Kripke S. A.* 1971. Identity and necessity. In M. Munitz, ed., *Identity and Individuation*. New York: New York University Press.
- . 1972. Naming and necessity. In G. Harman and D. Davidson, eds., *The Semantics of Natural Language*. Dordrecht: Reidel. [Reprinted as Kripke 1980]
- . 1980. *Naming and Necessity*. Cambridge, Mass.: Harvard University Press.
- . 1982. *Wittgenstein on Rule-Following and Private Language*. Cambridge, Mass.: Harvard University Press.
- Lahav R., and N. Shanks.* 1992. How to be a scientifically respectable «property dualist.» *Journal of Mind and Behavior* 13:211–32.
- Langton C. G.* 1989. *Artificial Life: The Proceedings of an Interdisciplinary Workshop on the Synthesis and Simulation of Living Systems*. Redwood City, Calif.: Addison-Wesley.
- Leckey M.* 1993. The universe as a computer: A model for prespace metaphysics. Manuscript. Philosophy Department, Monash University.
- Levine J.* 1983. Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly* 64:354–61.
- . 1988. Absent and inverted qualia revisited. *Mind and Language* 3:271–87.
- . 1991. Cool red. *Philosophical Psychology* 4:27–40.
- . 1993. On leaving out what it's like. In M. Davies and G. Humphreys, eds., *Consciousness: Psychological and Philosophical Essays*. Oxford: Blackwell.
- Lewis D.* 1966. An argument for the identity theory. *Journal of Philosophy* 63:17–25.
- . 1972. Psychophysical and theoretical identifications. *Australasian Journal of Philosophy* 50:249–58.
- . 1973. Counterfactuals. Cambridge, Mass.: Harvard University Press.
- . 1974. Radical interpretation. *Synthese* 23:331–44.
- . 1979. Attitudes *de dicto* and *de se*. *Philosophical Review* 88:513–45.
- . 1983a. Extrinsic properties. *Philosophical Studies* 44:197–200
- . 1983b. New work for a theory of universals. *Australasian Journal of Philosophy* 61:343–77.
- . 1986a. *On the Plurality of Worlds*. Oxford: Blackwell.
- . 1986b. *Philosophical Papers*. Vol. 2. New York: Oxford University Press.
- . 1990. What experience teaches. In W. Lycan, ed., *Mind and Cognition*. Oxford: Blackwell.
- . 1994. Reduction of mind. In S. Guttenplan, ed., *A Companion to the Philosophy of Mind*. Oxford: Blackwell.
- Libet B.* 1993. The neural time factor in conscious and unconscious events. In *Experimental and Theoretical Studies of Consciousness*. Ciba Foundation Symposium 174. New York: Wiley.
- Loar B.* 1990. Phenomenal states. *Philosophical Perspectives* 4:81–108.
- Lockwood M.* 1989. *Mind, Brain, and the Quantum*. Oxford: Blackwell.
- . 1992. The grain problem. In H. Robinson, ed., *Objections to Physicalism*. Oxford: Oxford University Press.
- Logothetis N., and J. D. Schall.* 1989. Neuronal correlates of subjective visual perception. *Science* 245:761–63.

- London F., and E. Bauer.* 1939. The theory of observation in quantum mechanics (in French). *Actualités scientifiques et industrielles*, no. 775. [English translation in Wheeler and Zurek 1983]
- Lucas J. R.* 1961. Minds, machines and Gödel. *Philosophy* 36:112–27.
- Lycan W.G.* 1973. Inverted spectrum. *Ratio* 15:315–19
- . 1987. Consciousness. Cambridge, Mass.: MIT Press.
- . 1995. A limited defense of phenomenal information. In T. Metzinger, ed., *Conscious Experience*. Paderborn: Schöningh.
- . 1996. Consciousness and Experience. Cambridge, Mass.: MIT Press.
- Mackay D. M.* 1969. Information, Mechanism, and Meaning. Cambridge, Mass.: MIT Press.
- Mackie J. L.* 1974. The Cement of the Universe. Oxford: Oxford University Press.
- . 1977. Ethics: Inventing Right and Wrong. Harmondsworth: Penguin Books.
- Marks L. E.* 1978. The Unity of the Senses: Interrelations among the Modalities. New York: Academic Press.
- Matzke D.*, ed. 1993. Proceedings of the 1992 Workshop on Physics and Computation. Los Alamitos, Calif.: IEEE Computer Society Press.
- , ed. 1995. Proceedings of the 1994 Workshop on Physics and Computation. Los Alamitos, Calif.: IEEE Computer Society Press.
- Maxwell G.* 1978. Rigid designators and mind-brain identity. In C. W. Savage, ed., *Perception and Cognition: Issues in the Foundations of Psychology*. Minnesota Studies in the Philosophy of Science, vol. 9. Minneapolis: University of Minnesota Press.
- McCarthy J.* 1979. Ascribing mental qualities to machines. In M. Ringle, ed., *Philosophical Perspectives in Artificial Intelligence*. Atlantic Highlands, N. J.: Humanities Press.
- McDowell J.* 1994. Mind and World. Cambridge, Mass.: Harvard University Press.
- McGinn C.* 1977. Anomalous monism and Kripke's Cartesian intuitions. *Analysis* 2:78–80.
- . 1989. Can we solve the mind-body problem? *Mind* 98:349–66.
- McLaughlin B. P.* 1992. The rise and fall of the British emergentists. In A. Beckermann, H. Flohr, and J. Kim, eds., *Emergence or Reduction? Prospects for Nonreductive Physicalism*. Berlin: De Gruyter.
- . 1995. Varieties of supervenience. In E. E. Savellos and U. D. Yalcin, eds., *Supervenience: New Essays*. Cambridge: Cambridge University Press.
- McMullen C.* 1985. «Knowing what it's like» and the essential indexical. *Philosophical Studies* 48:211–33.
- Meehl P. E., and W. Sellars.* 1956. The concept of emergence. In H. Feigl and M. Scriven, eds., *The Foundations of Science and the Concept of Psychology and Psychoanalysis*. Minnesota Studies in the Philosophy of Science, vol. 1. Minneapolis: University of Minnesota Press.
- Molnar G.* 1969. Kneale's argument revisited. *Philosophical Review* 78:79–89.
- Moore G. E.* 1922. Philosophical Studies. London: Routledge and Kegan Paul.
- Müller, G. E.* 1896. Zur Psychophysik der Gesichtsempfindungen. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane* 10:1–82.
- Nagel T.* 1970. Armstrong on the mind. *Philosophical Review* 79:394–403.
- . 1974. What is it like to be a bat? *Philosophical Review* 4:435–50.
- . 1983. The objective self. In C. Ginet and S. Shoemaker, eds., *Knowledge and Mind: Philosophical Essays*. New York: Oxford University Press.
- . 1986. The View from Nowhere. New York: Oxford University Press.
- Natsoulas T.* 1978. Consciousness. *American Psychologist* 33:906–14.
- Nelkin N.* 1989. Unconscious sensations. *Philosophical Psychology* 2:129–41.
- . 1993. What is consciousness? *Philosophy of Science* 60:419–34.
- Nemirov L.* 1990. Physicalism and the cognitive role of acquaintance. In W. Lycan, ed., *Mind and Cognition*. Oxford: Blackwell.

- Newell A. 1992. SOAR as a unified theory of cognition: Issues and explanations. *Behavioral and Brain Sciences* 15:464–92.
- Newton N. 1989. Machine understanding and the Chinese Room. *Philosophical Psychology* 2:207–15.
- Nida-Rümelin, M. 1995. What Mary couldn't know: Belief about phenomenal states. In T. Metzinger, ed., *Conscious Experience*. Paderborn: Schöningh.
- . 1996. Pseudonormal vision: An actual case of qualia inversion? *Philosophical Studies*.
- Papineau D. 1993. Philosophical Naturalism. Oxford: Blackwell.
- Parfit D. 1984. Reasons and Persons. Oxford: Oxford University Press.
- Peacocke C. 1992. Scenarios, concepts, and perception. In T. Crane, ed., *The Contents of Experience*. Cambridge: Cambridge University Press.
- Penrose R. 1987. Quantum physics and conscious thought. In B. Hiley and Peat, eds., *Quantum Implications: Essays in Honor of David Bohm*. New York: Methuen.
- . 1989. The Emperor's New Mind. Oxford: Oxford University Press. (Рус. пер.: Пенроуз Р. Новый ум короля: О компьютерах, мышлении и законах физики. М.: Издательство ЛКИ/URSS, 2011.)
- . 1994. Shadows of the Mind. Oxford: Oxford University Press.
- Perry J. 1979. The problem of the essential indexical. *Nous* 13:3–21.
- Perrie B. 1987. Global supervenience and reduction. *Philosophy and Phenomenological Research* 48:119–30.
- Place U. T. 1956. Is consciousness a brain process? *British Journal of Psychology* 47:44–50.
- Plantinga A. 1976. Actualism and possible worlds. *Theoria* 42:139–60.
- Putnam H. 1960. Minds and machines. In S. Hook, ed., *Dimensions of Mind*. New York: New York University Press.
- . 1975. The meaning of «meaning.» In K. Gunderson, ed., *Language, Mind, and Knowledge*. Minneapolis: University of Minnesota Press.
- . 1981. Reason, Truth, and History. Cambridge: Cambridge University Press.
- . 1983. Possibility and necessity. In *Philosophical Papers*. Vol. 3. Cambridge: Cambridge University Press.
- . 1988. *Representation and Reality*. Cambridge, Mass.: MIT Press.
- Pylyshyn Z. 1980. The «causal power» of machines. *Behavioral and Brain Sciences* 3:442–44.
- Quine W. V. 1951. Two dogmas of empiricism. *Philosophical Review* 60:20–43.
- . 1969. Propositional objects. In *Ontological Relativity and Other Essays*. New York: Columbia University Press.
- Rensink R. A., J. K. O'Regan, and J. J. Clark. 1995. Image flicker is as good as saccades in making large scene changes invisible. *Perception* 24 (suppl.):26–27.
- Rey G. 1982. A reason for doubting the existence of consciousness. In R. Davidson, S. Schwartz, and D. Shapiro, eds., *Consciousness and Self-Regulation*. Vol. 3. New York: Plenum.
- . 1986. What's really going on in Searle's «Chinese Room.» *Philosophical Studies* 50:169–85.
- . 1992. Sensational sentences reversed. *Philosophical Studies* 68:289–319.
- Reynolds C. 1987. Flocks, herds, and schools: A distributed behavioral model. *Computer Graphics* 21:25–34.
- Robinson H. 1976. The mind-body problem in contemporary philosophy. *Zygon* 11:346–60.
- . 1982. Matter and Sense. Cambridge: Cambridge University Press.
- Robinson W. S. 1988. *Brains and People: An Essay on Mentality and Its Causal Conditions*. Philadelphia: Temple University Press.
- Rosenberg G. H. 1996. Consciousness and causation: Clues toward a double-aspect theory. Manuscript, Indiana University.

- Rosenthal D. M. 1996. A theory of consciousness. In N. Block, O. Flanagan, and G. Güzeldere, eds., *The Nature of Consciousness*. Cambridge, Mass.: MIT Press.
- Russell B. 1927. *The Analysis of Matter*. London: Kegan Paul.
- Ryle G. 1949. *The Concept of Mind*. London: Hutchinson.
- Savage C. W. 1970. *The Measurement of Sensation*. Berkeley: University of California Press.
- Savitt S. 1982. Searle's demon and the brain simulator reply. *Behavioral and Brain Sciences* 5:342–43.
- Sayre K. M. 1976. *Cybernetics and the Philosophy of Mind*. Atlantic Highlands, N. J.: Humanities Press.
- Sayre-McCord G. 1988. Introduction: The many moral realisms. In G. Sayre-McCord, ed., *Essays on Moral Realism*. Ithaca, N. Y.: Cornell University Press.
- Schacter D. L. 1989. On the relation between memory and consciousness: Dissociable interactions and conscious experience. In H. Roediger and F. Craik, eds., *Varieties of Memory and Consciousness: Essays in Honor of Endel Tulving*. Hillsdale, N. J.: Erlbaum.
- Schlick M. 1932. Positivism and Realism. *Erkenntnis* 3.
- . 1938. Form and content: An introduction to philosophical thinking. In *Gesammelte Aufsätze 1926–1936*. Vienna: Gerold [Reprinted in H. L. Mulder and B. van de Velde-Schlick, eds., *Philosophical Papers*. Dordrecht: Reidel, 1979]
- Seager W. E. 1988. Weak supervenience and materialism. *Philosophy and Phenomenological Research* 48:697–709.
- . 1991. *Metaphysics of Consciousness*. London: Routledge.
- Searle J. R. 1980. Minds, brains, and programs. *Behavioral and Brain Sciences* 3: 417–24.
- . 1984. *Minds, Brains and Science*. Cambridge, Mass.: Harvard University Press.
- . 1990a. Consciousness, explanatory inversion and cognitive science. *Behavioral and Brain Sciences* 13:585–642.
- . 1990b. Is the brain a digital computer? *Proceedings and Addresses of the American Philosophical Association* 64:21–37.
- . 1992. *The Rediscovery of the Mind*. Cambridge, Mass.: MIT Press.
- Sellars W. 1956. Empiricism and the philosophy of mind. In H. Feigl and M. Scriven, eds., *The Foundations of Science and the Concepts of Psychology and Psychoanalysis*. Minnesota Studies in the Philosophy of Science, vol. 1. Minneapolis: University of Minnesota Press.
- . 1965. The identity approach to the mind-body problem. *Review of Metaphysics* 18:430–51.
- . 1981. Is consciousness physical? *Monist* 64:66–90.
- Shalllice T. 1972. Dual functions of consciousness. *Psychological Review* 79:383–93.
- . 1988a. Information-processing models of consciousness: Possibilities and problems. In A. Marcel and E. Bisiach, eds., *Consciousness in Contemporary Science*. Oxford: Oxford University Press.
- . 1988b. From Neuropsychology to Mental Structure. Cambridge: Cambridge University Press.
- Shannon C. E. 1948. A mathematical theory of communication. *Bell Systems Technical Journal* 27:379–423. [Reprinted in C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*. Urbana: University of Illinois Press, 1949]
- Shepard R. N. 1993. On the physical basis, linguistic representation, and conscious experience of colors. In G. Harman, ed., *Conceptions of the Human Mind: Essays in Honor of George A. Miller*. Hillsdale, N. J.: Erlbaum.
- Shoemaker S. 1975a. Functionalism and qualia. *Philosophical Studies* 27:291–315.
- . 1975b. Phenomenal similarity. *Critica* 7:3–37.
- . 1980. Causality and properties. In P. van Inwagen, ed., *Time and Cause*. Dordrecht: Reidel.
- . 1981. Some varieties of functionalism. *Philosophical Topics* 12:93–119.

- . 1982. The inverted spectrum. *Journal of Philosophy* 79:357–81.
- Sidelle A.* 1989. Necessity, essence, and individuation. Ithaca, N. Y.: Cornell University Press.
- . 1992. Rigidity, ontology, and semantic structure. *Journal of Philosophy* 89:410–30.
- Siewert C.* 1993. What Dennett can't imagine and why. *Inquiry* 36:93–112.
- . 1994. Understanding consciousness. Ph. D. diss. University of California, Berkeley. [Forthcoming as a book from Princeton University Press]
- Skyrms B.* 1980. Causal Necessity. New Haven: Yale University Press.
- Smart J. J. C.* 1959. Sensations and brain processes. *Philosophical Review* 68:141–56.
- Sperling G.* 1960. The information available in brief visual presentations. *Psychological Monographs* 74.
- Sperry R. W.* 1969. A modified concept of consciousness. *Psychological Review* 76:532–36.
- . 1992. Turnabout on consciousness: A mentalist view. *Journal of Mind and Behavior* 13:259–80.
- Sprigge T. L. S.* 1971. Final causes. *Proceedings of the Aristotelian Society* 45 (suppl.):149–70.
- . 1994. Consciousness. *Synthese* 98:73–93.
- Squires E.* 1990. Conscious Mind in the Physical World. Bristol: Hilger.
- Stalnaker R.* 1976. Possible worlds. *Nous* 10:65–75.
- . 1978. Assertion. In P. Cole, ed., Syntax and Semantics: Pragmatics. Vol. 9. New York: Academic Press.
- Stapp H. P.* 1993. Mind, Matter, and Quantum Mechanics. Berlin: Springer-Verlag.
- Stevens S. S.* 1975. Psychophysics: Introduction to Its Perceptual, Neural, and Social Prospects. New York: Wiley.
- Sutherland N. S.*, ed. 1989. The International Dictionary of Psychology. New York: Continuum.
- Swoyer C.* 1982. The nature of natural laws. *Australasian Journal of Philosophy* 60:203–23.
- Teller D. Y.* 1984. Linking propositions. *Vision Research* 24:1233–1246.
- . 1990. The domain of visual science. In L. Spillman, and J. S. Werner, eds., Visual Perception: The Neurophysiological Foundations. New York: Academic Press.
- Teller P.* 1984. A poor man's guide to supervenience and determination. *Southern Journal of Philosophy*, suppl., 22:137–62.
- . 1992. A contemporary look at emergence. In A. Beckermann, H. Flohr, and J. Kim, eds., Emergence or Reduction? Prospects for Nonreductive Physicalism. Berlin: De Gruyter.
- Thagard P.* 1986. The emergence of meaning: An escape from Searle's Chinese Room. *Behaviorism* 14:139–46.
- Thompson E.* 1992. Novel colours. *Philosophical Studies* 68:321–49.
- Tienson J. L.* 1987. Brains are not conscious. *Philosophical Papers* 16:187–93.
- Tooley M.* 1977. The nature of laws. *Canadian Journal of Philosophy* 7:667–98.
- . 1987. Causation: A Realist Approach. Oxford: Oxford University Press.
- Tye M.* 1986. The subjective qualities of experience. *Mind* 95:1–17.
- . 1992. Visual qualia and visual content. In T. Crane, ed., The Contents of Experience. Cambridge: Cambridge University Press.
- . 1993. Blindsight, the absent qualia hypothesis, and the mystery of consciousness. In C. Hookway, ed., Philosophy and the Cognitive Sciences. Cambridge: Cambridge University Press.
- . 1995. Ten Problems of Consciousness. Cambridge, Mass.: MIT Press.
- van Cleve J.* 1990. Mind-dust or magic? Panpsychism versus emergence. *Philosophical Perspectives* 4:215–26.
- van Gulick R.* 1988. A functionalist plea for self-consciousness. *Philosophical Review* 97:149–88.

- . 1992. Nonreductive materialism and the nature of intertheoretical constraint. In A. Beckermann, H. Flohr, and J. Kim, eds., *Emergence or Reduction? Prospects for Nonreductive Physicalism*. Berlin: De Gruyter.
- . 1993. Understanding the phenomenal mind: Are we all just armadillos? In M. Davies and G. Humphreys, eds., *Consciousness: A Mind and Language Reader*. Oxford: Blackwell.
- Velmans M.* 1991. Is human information processing conscious? *Behavioral and Brain Sciences* 14:651–69.
- Weinberg S.* 1992. *Dreams of a Final Theory*. New York: Pantheon Books. (Рус. пер.: *Вайнберг С. Мечты об окончательной теории: Физика в поисках самых фундаментальных законов природы*. М.: Издательство ЛКИ/URSS, 2008.)
- Weiskrantz L.* 1986. *Blindsight: A Case Study and Implications*. Oxford: Oxford University Press.
- . 1992. Introduction: Dissociated issues. In D. Milner and M. Rugg, eds., *The Neuropsychology of Consciousness*. London: Academic Press.
- Wheeler J. A.* 1990. Information, physics, quantum: The search for links. In W. H. Zurek, ed., *Complexity, Entropy, and the Physics of Information*. Redwood City, Calif.: Addison-Wesley.
- . 1994. It from bit. In *At Home in the Universe*. Woodbury, N. Y.: American Institute of Physics Press.
- Wheeler J. A., and W. H. Zurek.* 1983. *Quantum Theory and Measurement*. Princeton: Princeton University Press.
- White S. L.* 1986. Curse of the qualia. *Synthese* 68:333–68.
- Wigner E. P.* 1961. Remarks on the mind-body question. In I. J. Good, ed., *The Scientist Speculates*. New York: Basic Books.
- Wilkes K. V.* 1984. Is consciousness important? *British Journal for the Philosophy of Science* 35:223–43.
- Wilson M.* 1982. Predicate meets property. *Philosophical Review* 91:549–89.
- . 1985. What is this thing called «pain»? The philosophy of science behind the contemporary debate. *Pacific Philosophical Quarterly* 66:227–67.
- Winograd T.* 1972. *Understanding Natural Language*. New York: Academic Press.
- Wittgenstein L.* 1953. *Philosophical Investigations*. London: Macmillan.
- . 1965. Notes for lectures on «private experience» and «sense data». *Philosophical Review* 77.
- Wright R.* 1988. *Three Scientists and Their Gods*. New York: Times Books.
- Yablo S.* 1993. Is conceivability a guide to possibility? *Philosophy and Phenomenological Research* 53:1–42.
- Zuboff A.* 1994. What is a mind? In P. A. French, T. E. Uehling, and H. K. Wettstein, eds., *Philosophical Naturalism*. Midwest Studies in Philosophy, vol. 19. Notre Dame, Ind.: University of Notre Dame Press.
- Zurek W. H.* 1990. Complexity, Entropy, and the Physics of Information. Redwood City, Calif.: Addison-Wesley.

Именной указатель

Альберт Д. (Albert, D.) 427, 430, 438, 485
Александер С. (Alexander, S.) 75, 469
Армстронг Д. М. (Armstrong, D. M.) 32,
33, 51, 213, 444, 454, 456

Баарс Б. Дж. (Baars, B. J.) 148, 149, 298,
301, 302
Байсия Э. (Bisiach, E.) 445
Барвайс Дж. (Barwise, J.) 349
Бауэр Э. (Bauer, E.) 420
Бейтсон Г. (Bateson, G.) 351
Белл Дж. С. (Bell, J. S.) 157, 426, 434,
485, 486
Беркли Дж. (Berkeley, G.) 199
Бигелоу Дж. (Bigelow, J.) 459
Билер Дж. (Bealer, G.) 459, 463
Бирн А. (Bürgé, A.) 7, 214, 459, 464
Блок Н. (Block, N.) 7, 50, 128, 285, 286,
312, 314, 316, 327, 331, 455, 482
Блэкберн С. (Blackburn, S.) 116, 123
Боген Дж. (Bogen, J.) 480
Бойд Р. Н. (Boyd, R. N.) 191, 463
Бом Д. (Bohm, D.) 426, 427, 440,
441, 483
Бор Н. (Bohr, N.) 424
Броуд Ч. Д. (Broad, C. D.) 75, 167
Брукс Д. Г. М. (Brooks, D. H. M.) 450

Вайнберг С. (Weinberg, S.) 163
Ван Клив Дж. (Van Cleve, J.) 52, 294
Вебер Т. (Weber, T.) 425, 478
Вигнер Е. П. (Wigner, E. P.) 420, 440, 441
Вилманс М. (Velmans, M.) 483
Виноград Т. (Winograd, T.) 237
Витгенштейн Л. (Wittgenstein, L.) 42, 79,
455, 473

Гексли Т. (Huxley, T.) 193
Гелл-Манн М. (Gell-Mann, M.) 436

Голдман А. (Goldman, A.) 475
Гундерсон К. (Gunderson, K.) 457

Декарт Р. (Descartes, R.) 30, 169, 249,
444, 454
Деннет Д. К. (Dennett, D. C.) 7, 8, 51,
116, 149, 150, 187, 213, 219, 242–244,
287, 288, 300, 445, 457, 458, 460, 461
Джейнис Дж. (Jaynes, J.) 52
Джекендофф Р. (Jackendoff, R.) 45,
445, 474
Джексон Ф. (Jackson, F.) 7, 66, 137, 138,
180, 181, 192, 214, 448, 454–456,
459, 460, 462, 463, 470, 471
Джемс У. (James, W.) 30
Джонсон-Лэйрд Ф.
(Johnson-Laird, P.) 151
Дрейфус Х. (Dreyfus, H.) 484
Дретске Ф. И. (Dretske, F. I.) 116, 213,
290, 291, 349, 456, 463, 468, 471, 475
Дэвидсон Д. (Davidson, D.) 446, 467
Дэвис М. К. (Davies, M. K.) 83, 91

Журек В. Х. (Zurek, W. H.) 377

Зубов А. (Zuboff, A.) 482

Када Т. (Cuda, T.) 317, 481
Каплан Д. (Kaplan, D.) 83, 87, 88,
92, 452
Керратерс П. (Carruthers, P.) 475
Ким Дж. (Kim, J.) 7, 229, 446
Кирк Р. (Kirk, R.) 7, 193, 290, 454,
457, 461
Кларк Э. (Clark, A.) 7, 181, 182, 294
Кони Э. (Conee, E.) 461, 474
Коул Д. (Cole, D.) 482

- Кох К. (Koch, C.) 7, 152, 295, 298, 299, 458
 Крик Ф. (Crick, F.) 7, 152, 193, 295, 298, 299, 458
 Крипке С. А. (Kripke, S. A.) 7, 61, 69, 79, 82–84, 86, 111, 161, 170, 172, 173, 180, 181, 188–192, 211, 444, 450, 452, 456, 463, 464
 Куайн У. (Quine, W. V.) 79, 81, 88, 122, 452
 Кэрролл Дж. У. (Carroll, J. W.) 456

- Лайкан У. Г. (Lycan, W. G.) 7, 181, 187, 214, 288, 463, 468, 472, 480
 Левер Б. (Loewer, B.) 7, 427, 430, 438, 460
 Левин Дж. (Levine, J.) 7, 72, 134, 142, 214, 459, 481
 Либет Б. (Libet, B.) 298, 299
 Лики М. (Leckey, M.) 7
 Логотетис Н. (Logothetis, N.) 297
 Локвуд М. (Lockwood, M.) 7, 175, 214, 430, 444, 461, 465, 483
 Локк Дж. (Locke, J.) 328, 329
 Лондон Ф. (London, F.) 420
 Лор Б. (Loar, B.) 183, 184, 186, 214, 459, 461
 Льюис Д. (Lewis, D.) 7, 32, 33, 83, 88, 115, 138, 160, 186, 187, 213, 448, 450, 454–456, 459, 463, 465

- Макгинн К. (McGinn, C.) 8, 179, 190, 470
 Макдауэлл Дж. (McDowell, J.) 476
 Макдонаелл Дж. (McDonnell, J.) 7
 Маккей Д. М. (Mackay, D. M.) 353
 Макки Дж. Л. (Mackie, J. L.) 7, 456
 Маклафлин (McLaughlin, B. P.) 7, 447
 Макмаллен К. (McMullen, C.) 181
 Maxwell Г. (Maxwell, G.) 165, 175, 214, 217, 465
 Метцке Д. (Matzke, D.) 377
 Мил П. Э. (Meehl, P. E.) 469
 Молнар Дж. (Molnar, G.) 456
 Мур Дж. Э. (Moore, G. E.) 116, 117, 446
 Мюллер Г. Э. (Müller, G. E.) 479

- Нагель Т. (Nagel, T.) 7, 20, 115, 137, 383, 444, 456, 470, 474

- Немиров Л. (Nemirov, L.) 138, 186, 187, 463
 Нетсоулас Т. (Natsoulas, T.) 445
 Нида-Рюмелин М. (Nida-Rümelin, M.) 473, 474
 Нилкин Н. (Nelkin, N.) 7, 445, 475
 Ницше Ф. (Nietzsche, F.) 232
 Ньютон И. (Newton, I.) 219, 307
 Ньюэлл А. (Newell, A.) 50

О'Лери-Хоусон Дж. (O'Leary-Hawthorne, J.) 7

- Папино Д. (Papineau, D.) 214
 Паргеттер Р. (Pargetter, R.) 459
 Парфит Д. (Parfit, D.) 437
 Патнэм Х. (Putnam, H.) 83, 84, 86, 450, 481, 483, 484
 Пенроуз Р. (Penrose, R.) 156, 390, 409, 410, 431, 470, 484
 Перри Дж. (Perry, J.) 349
 Пилишин З. (Pylyshyn, Z.) 317
 Плейс У. Т. (Place, U. T.) 188, 191, 444, 445

- Райл Г. (Ryle, G.) 31–33, 42, 213, 444
 Рассел Б. (Russell, B.) 197, 198, 214, 380
 Рей Дж. (Rey, G.) 7, 13, 88, 213, 482
 Римини А. (Rimini, A.) 425
 Робинсон Г. (Robinson, H.) 214
 Робинсон У. (Robinson, W.) 214
 Розенберг Г. Г. (Rosenberg, G. H.) 8, 196, 465
 Розенталь Д. М. (Rosenthal, D. M.) 51, 213, 288, 445

- Севидж К. У. (Savage, C. W.) 7
 Сейр К. М. (Sayre, K. M.) 483
 Сейфарт Р. М. (Seyfarth, R. M.) 295
 Селларс У. (Sellars, W.) 382, 469, 471
 Серл Дж. Р. (Searle, J. R.) 8, 38, 115, 168, 211, 322, 323, 390, 391, 393, 397, 400–402, 404–406, 444, 458, 459, 465
 Сиверт Ч. (Siewert, C.) 252, 451, 476
 Сигер У. Э. (Seager, W. E.) 193, 447, 448, 482
 Скайрмс Б. (Skyrms, B.) 456
 Смарт Дж. Дж. К. (Smart, J. J. C.) 188, 191, 213, 444, 445

- Сперри Р. У. (Sperry, R. W.) 469
 Стивенс С. С. (Stevens, S. S.) 478
 Столнейкер Р. (Stalnaker, R.) 83, 92
 Суоэр К. (Swoyer, C.) 464
- Т**ай М. (Tye, M.) 181, 214, 290, 468
 Тимmons М. (Timmons, M.) 117
 Томас Л. (Thomas, L.) 7, 8, 20, 443
 Томпсон Э. (Thompson, E.) 446, 457
 Тули М. (Tooley, M.) 196, 456
- У**айт С. Л. (White, S. L.) 213, 459,
 466, 482
 Уиллер Дж. А. (Wheeler, J. A.) 377, 483
 Уилкис К. В. (Wilkes, K. V.) 213
 Уилсон М. (Wilson, M.) 450, 451
- Ф**арах М. (Farah, M.) 298, 299, 302
 Фаррел Б. А. (Farrell, B. A.) 443
 Фейгл Г. (Feigl, H.) 214
 Фелдман Ф. (Feldman, F.) 190
 Фехнер Г. (Fechner, G.) 478
 Фланаган О. (Flanagan, O.) 7, 214
 Фодор Дж. (Fodor, J.) 116
 Фосс Дж. (Foss, J.) 471
 Фреге Г. (Frege, G.) 83
 Фрейд З. (Freud, S.) 30, 31
 Фридкин Э. (Fredkin, E.) 377, 378, 483
- Х**айли Б. (Hiley, B.) 486
 Хамберстоун Я. Л.
 (Humberstone, I. L.) 83, 91
 Хамерофф С. Р. (Hameroff, S. R.) 156
 Хардин К. Л. (Hardin, C. L.) 457
 Харман Г. (Harman, G.) 468
 Харрис Р. (Harris, R.) 118, 437
 Хартл Дж. Б. (Hartle, J. B.) 436
 Хаугленд Дж. (Haugeland, J.) 484
 Хейр Р. М. (Hare, R. M.) 116, 446
 Хелман Дж. (Hellman, G.) 446
 Хендрich Т. (Honderich, T.) 214
 Хили Р. А. (Healey, R. A.) 7, 214, 435

- Хилл К. С. (Hill, C. S.) 465, 471, 472
 Ходгсон Д. (Hodgson, D.) 486
 Хорган Т. (Horgan, T.) 7, 117, 123, 143,
 181, 193, 214, 446, 448, 450, 454,
 456, 459, 464, 481, 482
 Хорст С. (Horst, S.) 7, 478, 483
 Хоффштадтер Д. Р. (Hofstadter, D. R.) 6,
 52, 237, 238, 409, 436, 484
 Хьюз Р. И. Г. (Hughes, R. I. G.) 435
- Ч**алмерс Д. Дж. (Chalmers, D. J.) 88,
 119, 305, 437, 464
 Чейни Д. Л. (Cheney, D. L.) 295
 Черчленд (Churchland, P. M.) 151, 181,
 218, 457, 462, 463, 480
 Черчленд П. С. (Churchland, P. S.) 218,
 450, 480
- Ш**актер Д. Л. (Schacter, D. L.) 302
 Шеллис Т. (Shallice, T.) 151
 Шенон К. Э. (Shannon, C. E.) 347, 350,
 353, 354
 Шепард Р. Н. (Shepard, R. N.) 227
 Шлик М. (Schlick, M.) 281
 Шумейкер С. (Shoemaker, S.) 8, 134,
 135, 247, 464, 472–474, 481, 482
- Э**ванс Г. (Evans, G.) 83, 91, 93
 Эверетт Х. (Everett, H.) 428–435,
 437–442, 485, 486
 Эдельман Дж. (Edelman, G.) 153,
 154, 458
 Экерман Д. (Ackerman, D.) 25, 443
 Экинс К. (Akins, K.) 295
 Экклз Дж. К. (Eccles, J. C.) 203
 Элитцур А. (Elitzur, A.) 7, 227, 234, 470
- Ю**м Д. (Hume, D.) 106, 219, 456
- Я**бло С. (Yablo, S.) 454
 Якоби Г. (Jacoby, H.) 457

Предметный указатель

Абстрактные сущности 210

актуальные мысли 27, 278

анализ см. концептуальный анализ

аналитическая истина

см. концептуальная истина

аномальный монизм 216

апостериорная необходимость 79, 126,
131, 132, 167, 169, 170, 172–174,
176, 184, 190, 192, 209, 211

— —, нерелевантность для
рекурентного объяснения 172, 176,
192, 211

априорная истина 263

Базовые законы

см. фундаментальные законы

бессознательные ментальные
состояния 37

биологический материализм 216

бихевиоризм 31, 213

блекнувшие квалиа 309, 313, 317,
319–324, 327, 328, 332, 336, 338,
342, 400

Бог 65, 161, 176, 178, 191, 240, 427, 442

— при характеристике логической
возможности 57

бодрствование 22, 46

боль 26, 188–194, 223, 226, 274, 277

бритва Оккама 217

Великое Разделение 215

витализм 144

вкусовые переживания 25

внимание 47, 48, 52, 147, 148, 273–275,
331, 335

внутренние свойства 197–200, 203, 210,
214, 232, 260, 332, 379–381, 386,
387, 406

водянистая материя 170–173, 188, 189

— —, содержание одних в других 64, 66

— —, характеристика 57, 177, 178,
180, 189

волновая функция 155–157, 414–420

все из бита 377, 379

вторичный интенсионал 132, 170–173,
175–177, 182, 183, 192, 262

— —, апостериорная детерминация 170,
171, 175

второпорядковое суждение 274

выводимость 57, 78, 126, 131, 139, 141,
170, 210

высказывания о сознании 222, 227, 228,
230, 231, 233–236, 238, 259, 260
см. также феноменальные
суждения

высокодетализированная организация
310, 311, 331, 342, 398, 399, 408

высококачественные презентации
298, 299, 302

высокоуровневое мышление 51

высшее существо 75

вычисление 318, 390–397, 402–404, 409,
412, 418, 424, 433, 439

Галлюцинация 57, 248, 276

геометрия опыта 281, 295

гипотеза Гольдбаха 179

глобальная доступность см. доступность
глобальное рабочее пространство 148,
298, 299, 301, 302

гомогенность 272, 275, 283, 382

гомункул 130, 218, 306, 316, 370, 374
гравитация 146, 219, 410

Данность 142, 276, 291, 392
Двойник Земли 261
двуаспектная теория 434, 435
двуаспектный принцип 355, 359, 371,
 374–376, 382, 383, 434
декомпозиция 430, 431, 435, 436
Демон Лапласа 58
дефляционное конструирование
 см. убеждение
дополнительный вопрос 72, 141,
 145, 147
— факт 65
доступность 140, 148, 276, 297–300
дуализм 160–163, 165, 166, 168, 169,
 171–176, 178, 180, 191, 194, 199,
 201, 203, 208–220, 227, 234, 235,
 253, 258, 266, 311, 322, 330, 373,
 374, 377, 384, 420, 442

Естественная возможность 236, 313,
 314, 316, 321, 330, 336, 391
— —, характеристика 59, 313, 321
— необходимость см. естественная
 возможность
— супервентность 193–195, 200, 206,
 207, 219, 373, 439
— —, введение 57, 59–62

Жесткая десигнация 193
животное сознание 47, 49, 283
жизнь 144–145, 370
— , размытость понятия 79

Зависимость от истории 338
заключение к наилучшему
 объяснению 270
закон газового состояния 58
законы природы 55, 59–61, 156, 162,
 166, 178, 206, 207, 214, 220
— —, юмистский взгляд 162, 166, 178
замена нейронов 310, 317, 318, 323, 324,
 337, 374, 402

знание 47, 244–253, 426
— о сознании см. эпистемология
значение 78, 81, 442
зомби 127–133, 135, 145, 158, 170, 172,
 179, 181, 188–191, 201, 204, 210,
 214, 224, 228, 231, 232, 243,
 245–247, 250, 254, 259, 260, 262,
 265, 286, 317, 361, 390
зомбийный мир 127, 128, 132, 133, 160,
 161, 166, 169–172, 174, 175,
 177–179, 184, 189, 197, 199, 208
— — и эпифеноменализм 184
зрительная кора 152, 301, 339, 356, 362
зрительные переживания 23

«И это все» факт 66
идеализм 176, 199, 424
измерительное устройство 420
иллюзия 57, 188, 231, 291, 385, 437
имена 329
имплементация 392, 393, 395–399, 402,
 404–407, 412, 433
инвертированная Земля 331
инвертированные квалиа 312–314, 317,
 328–333, 337, 340, 342, 343
— —, реальные случаи 312–314, 317,
 330, 332, 337
инвертированный спектр
 см. инвертированные квалиа
индексикальность 185, 261, 436, 437,
 439, 440
индукция, проблема 440
интенсивность 281, 295, 352, 375
интенсионал 81, 84
 см. также первичный
 интенсионал, вторичный
 интенсионал
интенсиональность 181, 183
интенциональность 38, 401, 406
— , связь с сознанием 39
интерпретация множественных миров
 430, 434 см. также интерпретация
 Эверетта
— относительности состояния 430
— расщепляющихся миров 429–431, 435
— Эверетта 430, 434

интроспекция 30, 47, 48, 51–53, 127, 242, 243, 247, 289, 367
 интуиция 129, 131, 146, 175, 179, 180, 205, 206, 437
 инфляционистское конструирование
 см. убеждение
 информационное пространство 347–360
 — состояние 276, 347, 349–352, 354–357, 364–366, 376–381, 383, 384
 — — в когнитивных процессах 350–352, 354, 357, 377, 383
 — —, фундирование 352
 информация 347–360
 искусственная жизнь 159, 167
 искусственный интеллект 150, 317, 343, 389–391, 398–400, 407, 408, 411, 412

Кажимость 243, 244

картизанский дуализм 162, 234
 каузальная действенность см. сознание, каузальная действенность
 — организация 130, 135, 228, 309, 310, 312, 395–398, 404, 407, 412
 — сверхдетерминация 195, 196
 — теория знания 246, 247, 252, 255
 — — референции 256, 257, 265
 каузальные пути 194, 202, 218, 351–353, 375, 386
 каузальный поток 197, 199, 379, 386
 квалиа 20, 154, 227, 228, 230, 295, 309, 312–314, 316, 317, 319–324, 326–333, 335–344, 399, 400, 402, 403
 —, дефиниция 20

квантовая механика 155–157, 159, 201, 202, 271, 313, 413, 414, 416, 418, 419, 421, 423, 428, 430, 431, 434, 440–442

— — при объяснении сознания 413, 419
 китайская комната 390, 400–402, 404, 405

— нация 306

клеточный автомат 378, 393, 394, 396

когнитивная модель 52, 71, 147–151, 201, 301, 409

— наука 28, 50, 71, 147, 150, 159, 166, 218, 227, 236, 294, 361, 397, 398, 409

когнитивные ограничения 179

коллапс волновой функции 156, 157, 202, 414, 419, 425, 441
 комбинаторная и реляционная структура 349, 355, 385
 комбинаторный автомат (КоА) 394
 компьютер 138, 139, 227, 230, 236, 312, 378, 389–392, 399, 405, 407
 конечный автомат 392–394
 коннекционизм 71
 концептуальная истина 78–101, 143, 158, 241, 287, 300
 — — и отсутствие дефиниций 79
 концептуальный анализ, критика 260
 концепции А-типа 213
 — В-типа 213
 копенгагенская интерпретация 424
 кошка Шредингера 313, 421, 424
 креативность 408
 кремниевый изоморф 130, 324, 325, 333, 337
 крипкеанская необходимость
 см. апостериорная необходимость
 куйновская критика 79, 81

Летучие мыши 137, 138, 295, 367

личность 127, 271, 305, 326, 339, 375, 436, 439, 440

логика 57, 78, 409

логическая возможность 126–133, 208, 313

— —, аргументы за логическую возможность 127, 133, 176, 179, 231, 336

— необходимость 126, 173, 342

— — (утверждений) 177

— — —, априорная vs. апостериорной разновидности 177

— — — и априорная истинна 177

— — —, широкая vs. узкой 57, 78

— супервентность 126, 127, 133, 135–137, 141, 145, 160, 161, 167, 181, 186–188, 193, 194, 204, 205, 211, 213, 214, 216, 229, 230, 236, 315, 343

— — —, vs. естественной супервентности 60

— — —, дефиниция 57, 64

— — и априорная выводимость 161, 211
 — — — материализм 65–66
 — — — редуктивное объяснение 73

Макроскопическая феноменология 381–383

максимальное состояние 432, 433
 математические истины 179
 материализм 138, 145, 160–163,
 166–168, 170, 176, 178–181,
 183–188, 191–193, 207–211, 213,
 214, 216–219, 306, 329, 330, 382
 —, дефиниция 65
 машина Тьюринга 392–394, 396, 411
 ментальные образы 26
 — понятия, двойная жизнь
 терминов 35–41
 —, разделение на феноменальные
 и психологические понятия 28–41
 мера на пространстве наблюдателей 438
 метафизическая необходимость 61, 158,
 176–178, 182, 184, 192, 201, 209,
 211, 213, 232, 330
 микросуждение 291
 микротрубочки 156
 микрофеноменальные свойства 381, 383
 мир ангелов 62
 миры см. возможные миры
 мистерианство 216
 множественная реализуемость 67
 модуль 150, 302, 404, 405
 модусы презентации 181–183
 мозги в бочке 249, 257, 427
 монизм 167, 168, 176, 199, 200, 216
 музыкальный опыт 24, 225
 мысленный эксперимент 142, 307, 312,
 313, 316, 319, 322, 324, 325, 332,
 334, 337

Небезразличное различие 351, 360, 364, 365, 377, 381, 382, 384

неверное описание миров 173
 невыразимость 263
 невычислимость 325, 410
 негативные факты и свойства 64

нейронаука 137, 150–152, 166, 181, 218,
 294–296, 361
 неконцептуальное содержание 292
 нелинейная динамика 159, 202, 319, 320
 нелокальность 156, 157, 426
 необходимая истина см. логическая
 необходимость
 необходимость-1 95, 98, 126
 необходимость-2 95
 необходимые и достаточные
 условия 79, 140
 неопределенность 79, 80, 162, 193, 202,
 308, 371, 423, 440
 нередуктивное объяснение 143, 159,
 214, 267, 287, 294, 360
 ничего не понимающий
 материализм 209
 НЛО 234, 240
 новая физика 155, 208–210
 nomическая возможность
 см. естественная возможность
 — супервентность см. естественная
 супервентность
 nomологические бездельники 206

Обонятельный опыт 25

обучение
 — как пример редуктивного
 объяснения 67, 77
 — — функциональное понятие 29, 37
 объяснительная когерентность 294,
 360, 364
 — нерелевантность 201, 203, 204, 210,
 213, 227, 229, 231–233, 235, 244,
 245, 252, 255, 256, 263, 266
 — —, аргументы против 229, 232, 244
 — —, проблемы с 229
 объяснительное исключение 229
 онтология 65, 357, 358, 376–378,
 380–385
 опыт; переживания см. сознание
 опытометр 147, 300
 оргазм 26
 организационная инвариантность 309,
 311, 327, 331, 342, 344–346, 359,
 392, 398, 408, 434

- организационные свойства 311, 343, 407, 408
- осведомленность 140, 148, 151, 154, 212, 225, 237, 242, 275–287, 292–308, 345, 359
- , дефиниция 49
 - , когерентность с сознанием 49
 - , объяснение 50–53
- осцилляции 337
- относительность 155, 156, 172, 421, 426, 430, 441
- отрицание слепоты 274, 325
- отсутствующие квалиа 309, 312–314, 316, 317, 324, 328, 330, 337, 342, 343
- —, аргументы за 342
 - —, — против 316, 317, 328, 330, 342, 343
 - —, логическая возможность
см. китайская нация; зомби
- ощущение 11, 20–28, 36, 233
- Память** 27, 136, 147, 150, 152, 154, 229, 244, 255, 256, 272, 277, 287, 298, 399, 437
- панпсихизм 196, 198, 271, 371, 373, 374, 381, 393, 442
- первичный интенсионал 126, 132, 145, 170–173, 175, 177, 182–184, 186, 192, 257–259, 261–265
- — и априорная концептуальная
истина 170, 173, 182, 184, 192, 257
- первопорядковая регистрация 290–292
- первопорядковое суждение 225, 226, 233, 275–277, 288–291
- —, vs. первопорядковой
регистрации 277
 - —, содержание 277, 290, 291
- передаваемость 354
- переживание глубины 23, 279
- переживания, вызванные
наркотиками 28
- цвета 23, 251, 280, 294, 295, 320, 339, 354, 355, 385
- переход от состояния к состоянию
393–397, 399, 402
- перцептивное многообразия 367, 368
- поведенческая инвариантность 327
- поведенческие диспозиции 32
- подпространства 348–350, 353, 356
- подсостояние 394, 395, 433
- позитивные факты и свойства 64, 65
- познание 221
- понятие каузальной роли 69
- понятия цвета 362
- последняя теорема Ферма 132
- постулат измерения 416–419, 422, 425, 426, 428, 432, 434, 437, 438
- поясняющее объяснение 74
- правдоподобные ограничения 271–273
- представимость 57, 126, 131–133, 135, 145, 169–171, 177, 179, 180, 188, 201, 286, 330
- представимость-1 97, 98
- представимость-2 97, 98
- привидение в машине 162, 166, 238
- принцип детектируемости 274
- достоверности 274
 - инвариантности см. организационная
инвариантность
 - принципы когерентности 273–275, 279, 282, 284–286, 292–297, 299, 303, 304, 308, 345, 346, 359, 364
 - — между сознанием
и осведомленностью 346, 359
 - — — и познанием 359
- проблема грануляции 382
- ментальное-ментальное 45
 - остановки 411
 - сознание—тело 20, 44–46
 - соотношения ментального
и физического см. проблема
сознание-тело
- проводил в объяснении 72, 157, 294, 413
- прроверяемость 270
- программа 392–394, 400–406
- пропозиция 38, 223
- простота 74, 217, 269, 271, 275, 307, 371, 387, 427, 428, 432, 438
- пространство-время 163, 378, 379
- протофеноменальные свойства 164, 175, 176, 198, 199, 214, 216, 372, 373, 381

прямая доступность *см.* доступность
 психологические состояния *см.* фено-
 менальные/психологические
 состояния
 психология 30, 52, 148, 223, 224, 227,
 273, 282, 296, 336, 367, 368
 психоны 203, 204
 психофизика 295
 психофизические законы 164, 166, 200,
 206, 219–221, 267–270, 273, 303,
 304, 306, 307, 344–346, 357, 358,
 360, 373, 385, 393, 432, 438
 — —, эпистемология 164, 269
 психофункционализм 171, 216

Развоплощение 189–192

ревизуемость 82
 регистрация 233, 276, 290–292
 редуктивное объяснение 66–77, 125,
 126, 132, 135, 138, 140–142,
 144–148, 151, 154, 156–159, 178,
 207, 208, 211, 214, 226–228, 231,
 232, 234, 238, 240–242, 264, 287,
 290, 293, 294, 360, 361, 364
 — — в когнитивной науке 71–73, 158,
 159, 228
 — — посредством функционального
 анализа 68–71
 — — поясняющее vs. устраниющее
 тайну объяснение 74
 редукция 67, 212
 религиозная вера 239
 реляционные свойства 200, 263
 repidентионализм 216
 робот 130, 312, 317–320, 323
 рок 320

Самосознание 19, 47, 48, 51, 52, 154, 289, 367

связующие принципы 142, 143, 158,
 178, 211, 296, 313, 345, 357
 связывание 152
 симуляция 167, 202, 314, 341, 390, 399,
 405, 407–410, 412
 скакущие квадри 309, 313, 314, 317, 332,
 336–342, 344, 399, 400, 402, 403

скрытые параметры 425, 426, 440
 следование правилам 408
 слепое зрение 243, 283, 284, 286,
 316, 363
 сложность 277, 349, 355, 366–369,
 371, 441
 слуховой опыт 24, 321
 сновидения 285
 сознание
 —, дефиниция 19, 20
 — доступа 50, 285, 286
 —, единство 28
 —, знание *см.* эпистемология
 —, крах редуктивного объяснения 72
 —, локальная супервентность 57
 —, неожиданный характер 21
 —, повсеместность *см.* панпсихизм
 —, поток 27
 —, различные значения термина 46–48
 —, специфический характер 21
 —, феноменальная и психологическая
 разновидности 46–53
 сознательный опыт *см.* сознание
 солипсизм 270
 сон 46, 322
 сориты 326
 спин 163, 164, 414, 415, 417, 418, 421
 способность давать отчет 47, 49, 51, 131,
 139, 148, 150, 151, 154, 215, 243,
 259, 271–300
 структура различия 142, 155, 279, 280,
 347–349, 352, 354, 355, 359, 364,
 365, 377
 структурная когерентность 279, 282,
 293–295, 303, 304, 345, 346, 359
 структурные свойства 141, 209, 210, 231
 субъективный опыт *см.* сознание
 суждение 205, 221–228, 230–236,
 238–248, 250, 251, 253, 259, 264,
 265, 274–277, 288–291, 305, 321,
 323, 327, 344, 360–365, 387, 426
 суждения о сознании
 см. феноменальные суждения
 супервентностные законы 164
 — кондиционалы 79–82

супервентность 13, 15, 16, 54, 125–127, 133, 135–137, 141, 143, 145, 160, 161, 166, 167, 169, 172, 176, 178, 181, 183, 186–188, 190, 193–195, 200, 205–207, 211, 213, 214, 216, 219, 229, 230, 236, 268, 315, 343, 373, 439
 —, дефиниция 54, 55
 —, различие глобальной/локальной 56–57, 62, 77
 —, — логической/естественной 56–62
 — сильная и слабая 61 *см. также*
 логическая супервентность
 естественная супервентность
 суперпозиция 414–426, 428–438, 440, 441
 существенные свойства 172, 174, 175, 199, 210

Таблица соответствий 318, 327, 404
 тактильный опыт 24
 телеология 68
 телеофункционализм 216
 телесное ощущение 26
 теорема Геделя 179, 390, 409
 теоретические сущности 204, 208
 теория всего 155, 163, 165, 239, 273, 361, 426
 термодинамика 235, 269, 423
 термостат 351, 365–374, 376
 тождество 16, 55, 169, 182, 188–192, 201, 339, 411, 436, 437, 439
 топологическая структура 348, 349
 трудная проблема 384, 413, 435

Убеждение 136, 139, 144, 170, 222–225, 228, 234, 235, 245–256, 258–260, 263–265, 291, 308, 321–323, 325, 328, 335, 336, 341, 365
 —, инфляционные/дефляционные конструкции 39
 —, о сознании *см. феноминальные суждения*
 —, реляционный элемент 40
 —, содержание 38
 уравнение Шредингера 416–418, 422, 425, 426, 428, 429, 431–434, 438, 441

условия истинности 82, 259
 устраниющее тайну объяснение 74

Фальсифицируемость
см. проверяемость

феноменальное качество *см. квалиа*
 — многообразие 242, 367, 368
 феноменальные понятия 170, 171, 175, 180, 183, 243, 245, 258, 259, 261, 264, 265 *см. также*
 феноменальные суждения
 — свойства 162, 164, 171, 173–176, 183, 197–200, 203, 204, 210, 212–214, 216, 232, 233, 260, 369, 372, 377, 378, 380, 381, 383, 386, 408
 — суждения 175, 222–224, 275, 276, 293
 — —, второпорядковые 274
 — —, достоверность 274
 — —, принципы когерентности 274
 — —, содержание 258
 феноменальные/психологические состояния, как то, что исчерпывает ментальное 35–41
 — —, проблемы 44, 45
 — —, различие 36, 41, 43
 — —, сопутствование 35, 36, 41, 42
 физика 55, 127, 137, 146, 155–157, 159, 163, 165, 166, 176, 184, 197, 199, 201, 202, 204, 208–210, 218, 219, 221, 227, 229, 235, 268, 346, 358, 377–381, 390, 408, 422–424
 —, автономия и простота 74
 — как фундаментальная теория 267
 физикализм 199, 214
 физикалистский функционализм 216
 физические свойства 141, 143, 157, 159, 162, 164, 166–168, 171, 174–176, 184, 197–199, 201, 203, 207, 210, 214, 253, 260, 304, 309, 346, 377, 378, 380, 381, 386, 387, 407, 413, 420
 — —, внутренняя/внешняя природа 162, 210
 — —, дефиниция 55, 162, 164, 167, 171, 175, 198, 199, 210, 309
 фиксация референции 173 *см. также*
 первичный интенсионал

- функциональная теория 152, 157, 163, 165, 267, 345, 346, 358–360, 377, 386
 функциональные законы 163–166, 168, 178, 200, 219, 267, 269, 272, 273, 303, 304, 307, 320, 344–346, 358, 360, 366, 375, 377, 419, 422, 425, 439
 — —, простота 163, 164, 269, 304, 346, 419
 — —, свойства 157, 162, 164–166, 184, 198, 304, 358, 371–373, 380–382
 функционализм (нередуктивный) 210, 311, 342, 343
 — (редуктивный) 208, 211–213, 215, 216, 247, 313
 функциональная организация 130, 135, 285, 294, 295, 304, 306, 309–318, 323, 328, 329, 331, 332, 334–338, 341–344, 360, 392, 398, 399, 403, 407, 408
 — — и вычисление 311, 331, 332
 функциональные свойства 141, 143, 155, 158, 171, 184, 208, 210, 230, 231, 259, 260, 331, 336
 функциональный анализ 68, 139–141, 180, 211, 283
 — изоморф 130, 215, 311, 317, 318, 320, 321, 323–325, 327–330, 333, 334, 337, 338, 360
- Ц**ветовое пространство 134, 293, 294, 330, 332, 355, 356, 385
 центрированные миры 172, 185, 186, 257–260, 262, 263
 церебральная селебрити 287, 300

- Ч**ерно-белая комната 137
 чувствительная зависимость от изначальных условий 410
- Э**волюция 158, 220, 271, 288, 336, 375, 421, 441
 экспериментальные исследования 271
 экстроспекция 243
 эктоплазма 62–64, 203
 эмерджентность 75, 167, 168
 эмоции 27, 37, 233, 277, 281, 385
 эпистемическая асимметрия 135–137, 143
 эпистемический рычаг 293, 296, 297
 эпистемология 126, 252, 254, 255, 264–266, 365
 эпифеноменализм 193, 194, 196, 200, 201, 203–206, 213
- Я**зык 142, 150, 154, 263, 282, 364, 396, 398, 400–402, 404
 — и сознание 49
 —, менталистский см. феноменальные понятия
- S**HRLU 237
- X**-фактор 305, 307
XYZ 173, 174, 177, 188

Представляем Вам следующие книги:



Серия «Маркс, Энгельс — творцы научного социализма»

- ✓ *Маркс К. Капитал: Квинтэссенция ВСЕХ ТОМОВ «Капитала».*
- ✓ *Маркс К. Капитал: Отредактированные Марксом фрагменты для французского издания, отсутствующие в общепринятой версии.*
- ✓ *Маркс К. Экономические рукописи 1857–1861 гг. Первоначальный вариант «Капитала». В 2 ч.*
- ✓ *Маркс К. Ницета философии.*
- ✓ *Маркс К. К критике политической экономии.*
- ✓ *Маркс К. Труды по эпикурейской философии.*
- ✓ *Энгельс Ф. Конспект первого тома «Капитала» К. Маркса.*
- ✓ *Энгельс Ф. Происхождение семьи, частной собственности и государства.*
- ✓ *Энгельс Ф. К жилищному вопросу.*
- ✓ *Энгельс Ф. Роль насилия в истории.*

Философия. История философии

- ✓ *Асмус В. Ф. Платон.*
- ✓ *Асмус В. Ф. Проблема интуиции в философии и математике.*
- ✓ *Асмус В. Ф. Немецкая эстетика XVIII века.*
- ✓ *Могилевский Б. М. Платон и сицилийские тираны. Мудрец и власть.*
- ✓ *Зубов В. П. Аристотель: Человек. Наука. Судьба наследия.*
- ✓ *Соколов В. В. От философии Античности к философии Нового Времени.*
- ✓ *Соколов В. В. Средневековая философия.*
- ✓ *Попкова Н. В. История философских идей: Введение в историю философии.*
- ✓ *Чухлеб С. Н. Археология идей: История философии в доступном изложении.*
- ✓ *Купарашвили М. Д. Лекции по современной философии: Модернизм.*

Серия «Bibliotheca Scholastica»

- ✓ *Фома Аквинский. Сочинения.*
- ✓ *Бозэй Дакийский. Сочинения.*
- ✓ *Уильям Окам. Избранное.*
- ✓ *Роберт Гроссестест. Сочинения.*
- ✓ *Блаженный Иоанн Дунс Скот. Трактат о первоначале.*
- ✓ *Апполонов А. В. Латинский аверроизм XIII века.*

Серия «Из наследия мировой философской мысли: история философии»

- ✓ *Гегель Г. В. Ф. Введение в философию: Философская пропедевтика.*
- ✓ *Кант И. Антропология с прагматической точки зрения.*
- ✓ *Шульц И. Разъясняющее изложение «Критики чистого разума».*
- ✓ *Форлендер К. История философии: Античность. Средние века.*
- ✓ *Швеглер А. История философии: От досократиков до средневековой схоластики.*
- ✓ *Оболенский Л. Е. История мысли: Опыт критической истории философии.*
- ✓ *Вебер А. История европейской философии.*
- ✓ *Штёклер А. История средневековой философии.*
- ✓ *Бэймкер К. Европейская философия Средневековья.*

Представляем Вам следующие книги:



URSS

Серия «Классики науки»

- ✓ Ньютон И. Математические начала натуральной философии.
- ✓ Ньютон И. Математические работы.
- ✓ Евклид. Начала.
- ✓ Фреге Г. Логика и логическая семантика.
- ✓ Гюйгенс Х. Трактат о свете.
- ✓ Гейзенберг В. Избранные труды.
- ✓ Max Э. Механика: Историко-критический очерк ее развития.
- ✓ Лукреций. О природе вещей: Билинга.
- ✓ Бозий. «Утешение Философии» и другие трактаты.
- ✓ Гельмольц Г. Учение о слуховых ощущениях как физиологическая основа для теории музыки.
- ✓ Бюффон Ж. Всеобщая и частная естественная история: История и теория Земли.
- ✓ Дарвин Ч. Пangenезис.

- ✓ Гайденко П. П. Эволюция понятия науки: Становление и развитие первых научных программ.
- ✓ Гайденко П. П. История новоевропейской философии в ее связи с наукой.
- ✓ Шишков И. З. Современная западная философия: Очерки истории.
- ✓ Шишков И. З. В поисках новой рациональности: Философия критического разума.
- ✓ Розин В. М. Традиционная и современная философия.
- ✓ Розин В. М. (ред.) Приобщение к философии: Новый педагогический опыт.
- ✓ Розин В. М. Типы и дискурсы научного мышления.
- ✓ Розин В. М. Метаморфозы российского менталитета: Философские этюды.
- ✓ Розин В. М. Введение в схемологию: Схемы в философии, культуре, науке, проектировании.
- ✓ Розин В. М. Демаркация науки и религии. Анализ учения и творчества Эмануэля Сведенборга.
- ✓ Аристотель. Физика.
- ✓ Семенов Ю. И. Философия истории.

Наши книги можно приобрести в магазинах:

- «НАУКУ – ВСЕМ!» (м. Профсоюзная, Нахимовский пр-т, 56. Тел. (499) 724-2545)
- «Библио-Глобус» (м. Лубянка, ул. Мясницкая, 6. Тел. (495) 625-2457)
- «Московский дом книги» (м. Арбатская, ул. Новый Арбат, 8. Тел. (495) 203-8242)
- «Молодая гвардия» (м. Полянка, ул. Б. Полянка, 28. Тел. (495) 238-5001, (495) 780-3370)
- «Дом научно-технической книги» (Ленинский пр-т, 40. Тел. (495) 137-6019)
- «Дом книги на Ладожской» (м. Бауманская, ул. Ладожская, 8, стр. 1. Тел. (495) 267-0302)
- «Санкт-Петербургский Дом книги» (Невский пр., 28. Тел. (812) 448-2355)
- «Книжный бутик» (г. Киев, книжный рынок «Петровка», ряд 62, место 8 (павильон «АкадемКнига»). Тел. +38 (067) 273-5010)
- Сеть магазинов «Дом книги» (г. Екатеринбург, ул. Антона Валека, 12. Тел. (343) 253-5010)

Тел./факс:

+7 (499) 724-25-45
(многоканальный)

E-mail:

URSS@URSS.ru
<http://URSS.ru>

Уважаемые читатели! Уважаемые авторы!

Наше издательство специализируется на выпуске научной и учебной литературы, в том числе монографий, журналов, трудов ученых Российской академии наук, научно-исследовательских институтов и учебных заведений. Мы предлагаем авторам свои услуги на выгодных экономических условиях. При этом мы берем на себя всю работу по подготовке издания — от набора, редактирования и верстки до тиражирования и распространения.



Среди вышедших и готовящихся к изданию книг мы предлагаем Вам следующие:

- ✓ *Пенроуз Р. Новый ум короля. О компьютерах, мышлении и законах физики.*
- ✓ *Майнцер К. Сложносистемное мышление: Материя, разум, человечество. Новый синтез.*
- ✓ *Фома Аквинский. Сумма теологии. Билингва (латынь—русский). Т. 2–4.*
- ✓ *Миль Дж. Ст. Система логики силлогистической и индуктивной: Изложение принципов доказательства в связи с методами научного исследования.*
- ✓ *Поппер К. Р. Знание и психофизическая проблема: В защиту взаимодействия.*
- ✓ *Поппер К. Р. Объективное знание. Эволюционный подход.*
- ✓ *Поппер К. Р. Все люди — философы.*
- ✓ *Волков Д. Б. Бостонский зомби: Д. Деннет и его теория сознания.*
- ✓ *Иванов Д. В. Природа феноменального сознания.*
- ✓ *Гусейнов А. А. Античная этика.*
- ✓ *Балагушкин Е. Г. Мистицизм в современной России: Теория. Основные представители.*
- ✓ *Яковлев В. А. От креативов метафизики к философии творчества: Универсум принципов современной науки.*
- ✓ *Гусейнов А. А., Белкина Г. Л. (ред.) Человек в единстве социальных и биологических качеств.*
- ✓ *Белкина Г. Л. (ред.) Человек и его будущее: Новые технологии и возможности человека.*
- ✓ *Майоров Г. Г. Философия как искание Абсолюта: Опыты теоретические и исторические.*
- ✓ *Майоров Г. Г. Формирование средневековой философии: Латинская патристика.*
- ✓ *Майданов А. С. Методология научного творчества.*
- ✓ *Новиков А. С. Философия научного поиска.*
- ✓ *Крушинов А. А., Мамчур Е. А. (ред.) Будущее фундаментальной науки.*
- ✓ *Яо ИН Философия гуманизма И. Т. Фролова: Взгляд из Китая.*
- ✓ *Руднев В. П. Божественный Людвиг. Витгенштейн: Формы жизни.*
- ✓ *Саврей В. Я. Александрийская школа в истории философско-богословской мысли.*
- ✓ *Васильев В. В. (ред.) 100 этюдов о Канте.*
- ✓ *Гайденко П. П. Эволюция понятия науки (XVII–XVIII вв.): Формирование научных программ нового времени.*

По всем вопросам Вы можете обратиться к нам:
тел. +7 (499) 724–25–45 (многоканальный)
или электронной почтой URSS@URSS.ru
Полный каталог изданий представлен
в интернет-магазине: <http://URSS.ru>

Научная и учебная
литература



ДЭВИД ДЖОН ЧАЛМЕРС

Профессор кафедры философии Нью-Йоркского университета и Австралийского национального университета. Родился в Австралии в 1966 г. Изучал математику в Университете Аделаиды и в Оксфордском университете. Защитил диссертацию по философии в Индианском университете под руководством Д. Хофтадтера в 1993 г. Приобрел мировую известность после выступления на Первой Туссанской конференции в 1994 г. с докладом о «трудной» и «легких» проблемах сознания. Автор трудов по философии сознания, языка, эпистемологии и метафизике. В числе недавних работ – книги «Характер сознания» (2010) и «Конструируя мир» (2012) – первая часть задуманной Д. Чалмерсом масштабной трилогии.



...выдающийся вклад в наше понимание сознания».

Стивен Пинкер



Один из лучших текстов о сознании как для продвинутых читателей, так и для новичков. Кажется, будто Чалмерс извлек все что можно из этой самой неподъемной из проблем».

Колин Макгинн



...исключительно амбициозная и исключительно успешная работа – лучшая книга о сознании за многие годы».

Дэвид Льюис

Наше издательство предлагает следующие книги:



10443 ID 157324



9 785397 037785 >

Издательская группа
URSS

Каталог изданий
в Интернете:
<http://URSS.ru>
E-mail: URSS@URSS.ru

117335, Москва, Телефон / факс
Нахимовский (многоканальный)
проспект, 56 +7 (499) 724 25 45

Отзывы о настоящем издании, а также обнаруженные
ошибки присылайте по адресу URSS@URSS.ru.
Ваши замечания и предложения будут учтены
и отражены на web-странице этой книги на сайте
<http://URSS.ru>