

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/369374579>

РОССИЙСКАЯ АКАДЕМИЯ НАУК СИБИРСКОЕ ОТДЕЛЕНИЕ НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ ИНСТИТУТ НЕЙРОНАУК И МЕДИЦИНЫ ЯВЛЯЕТСЯ ЛИ ПСИХИЧЕСКОЕ ЭПИФЕНОМЕНОМ АКТИВНОСТИ МОЗГА?

Book · March 2023

DOI: 10.53954/9785604788875

CITATIONS

0

READS

240

1 author:



Gennady G Knyazev

Institute of Physiology and Basic Medicine

181 PUBLICATIONS 4,836 CITATIONS

SEE PROFILE

РОССИЙСКАЯ АКАДЕМИЯ НАУК
СИБИРСКОЕ ОТДЕЛЕНИЕ
НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ ИНСТИТУТ
НЕЙРОНАУК И МЕДИЦИНЫ

Г.Г. Князев

**ЯВЛЯЕТСЯ ЛИ ПСИХИЧЕСКОЕ
ЭПИФЕНОМЕНОМ АКТИВНОСТИ МОЗГА?**

Философские и нейробиологические аспекты проблемы

НОВОСИБИРСК
2022

УДК 130.3 + 159.95
ББК 87.523 + 28.707
К54
DOI 10.53954/9785604788875

Рецензенты

доктор философских наук *А.Н. Савостьянов*
доктор биологических наук *О.М. Разумникова*
кандидат биологических наук *Е.Ю. Приводнова*

Князев Г.Г.

К54 Является ли психическое эпифеноменом активности мозга? Философские и нейробиологические аспекты проблемы / Г.Г. Князев; Рос. акад. наук, Сиб. отд-ние, НИИ нейронаук и медицины. – Новосибирск: СО РАН, 2022. – 316 с.
ISBN 978-5-6047888-7-5

В монографии обсуждаются философские и нейробиологические аспекты проблемы отношения психических процессов и активности мозга человека. Разбираются философские системы, рассматривающие субъективно переживаемые психические состояния в качестве эпифеномена активности мозга, доминирующим среди которых является редукционизм, постулирующий принципиальную сводимость психических процессов к процессам в мозге. Приводится систематический обзор эмпирических данных и наиболее влиятельных научных теорий о природе связи между нервными процессами и содержанием сознания, памяти, мышления и эмоций. Рассматриваются данные о нервных коррелятах оценочных суждений, включая мораль, чувство красоты и спиритуальность. Делается вывод о принципиальном различии между природой нервной активности и природой психических процессов. Анализ систем ценностных ориентаций показывает, что моральную, эстетическую или религиозную суть ментального состояния невозможно «расшифровать» из паттерна сопровождающей его активности мозга. Делается вывод, что на основе имеющихся в настоящее время данных редукционизм вряд ли можно рассматривать как жизнеспособную опцию.

Книга предназначена для психологов, нейробиологов, студентов медицинских, психологических и биологических факультетов, а также для всех читателей, интересующихся мировоззренческими проблемами нейронаук.

УДК 130.3 + 159.95
ББК 87.523 + 28.707

Утверждено к печати Ученым советом

Научно-исследовательского института нейронаук и медицины СО РАН

Издание осуществлено при поддержке Российского научного фонда
(проект № 22-15-00142)

ISBN 978-5-6047888-7-5

© Сибирское отделение РАН, 2022
© Князев Г.Г., 2022

ПРЕДИСЛОВИЕ

Отношения между философией и наукой всегда были двусторонними – философские доктрины опирались на научные знания, а развитие науки направлялось и регламентировалось доминирующей философской доктриной. Формально наука не должна руководствоваться идеологическими установками, но реалии жизни практически любого общества показывают, что это не так. Доминирующее в обществе мировоззрение достаточно жестко ограничивает рамки научной дозволенности и общее направление развития научных дисциплин. Это происходит и имплицитно – через мировоззрение самих ученых, сформированное под воздействием преобладающих в культуре ценностей, и эксплицитно – в форме деклараций научных сообществ и периодических изданий, а также запросов и финансовых потоков со стороны общества и государственных структур.

Не все научные дисциплины находятся под одинаковым прессингом доминирующей философии. Можно, наверное, всю жизнь заниматься изучением погоды или поведения муравьев, не задумываясь о философских проблемах. Однако, если речь идет, например, о возникновении и эволюции жизни, мировоззренческие и философские темы выходят на первый план. То же можно сказать и об изучении природы психических явлений, которое не может обойтись без принятия той или иной философской позиции. Например, изучение сознания вообще не имеет смысла, если принять точку зрения элиминативного материализма, отрицающего его существование [Ramsey, 2020]. Или же вместо того, чтобы объяснять, как возникает сознание, нужно на самом деле объяснить, как возникает иллюзия того, что у нас есть сознание [Dennett, 1991; Frankish, 2017].

Философия – это ристалище для нескончаемых дебатов, где невозможно выявить окончательного победителя. Любая философская позиция имеет сторонников и противников, которые обмениваются аргументами и контраргументами. Во многом похожая ситуация наблюдается и в науке, где любая теория, какой бы успешной и авторитетной она ни казалась, неизбежно в свое время уступит место более успешной или полной теории, а любые эмпирические данные могут опровергаться или интерпретироваться в свете альтернативных теорий. На бытовом уровне, в об-

щественном сознании эта неоднозначность, как правило, утрачивается и господствовавшая философская доктрина или научная теория воспринимаются как догмы.

Доминирующая в наши дни в общественном сознании материалистическая картина мира, по сути, вытекает из поверхностного понимания естественнонаучной картины, что свидетельствует о непререкаемом авторитете естественных наук. Все, наверное, слышали расхожую фразу: «наука пока не знает ответа на этот вопрос». Под этим выражением подразумевается, что если не знает наука, то не знает никто, но наука в будущем наверняка найдет нужный ответ. Безоговорочная вера во всемогущество науки появилась относительно недавно, и одним из определяющих ее факторов являются несомненные успехи естественных наук. На волне этих успехов представители естественных наук начинают проникать в области, ранее считавшиеся прерогативой философов и теологов.

В чем смысл жизни? Какова природа человеческого разума, любви, морали? На все подобные вопросы сейчас можно услышать естественнонаучные ответы и объяснения. Жизнь возникла из неживой природы в силу действия случайных физических и химических факторов, и не следует искать в этом какой-то смысл; человек – продукт естественного отбора, а разум, любовь и мораль – это результат химических и электрических процессов в мозге. Поскольку вопросы относятся к числу наиболее важных для человека, то, в силу непререкаемого авторитета естественных наук, предлагаемые ответы могут иметь и уже имеют большое влияние на все сферы человеческой жизни – от экономики и политики до психического здоровья и субъективного благополучия отдельного человека. Они, по сути, провозглашают бессмысленность человеческой жизни, которая является результатом действия слепых сил природы и случайных факторов, результатом борьбы за существование, победы в этой борьбе наиболее приспособленных и гибели менее приспособленных.

Единицей естественного отбора является ген, и отбираются лишь те гены, которые увеличивают вероятность выживания [Dawkins, 2016]. Все атрибуты человека, казавшиеся (и многим продолжающие казаться) наиболее важными – разум, сознание, свобода воли, мораль, – это только иллюзия, побочный продукт активности нейронов в мозге [Dennett, 1991; Свааб, 2019]. Делает ли человека счастливым такое восприятие мира и своего места в

нем? Социологические исследования однозначно говорят, что нет. Как показывают метаанализы большого количества работ, материалистическое мировоззрение отрицательно коррелирует с чувством удовлетворенности жизнью [Dittmar et al., 2014; Kasser, 2016]. Адепты этой картины мира утверждают, однако, что какой бы неприятной и противоречащей здравому смыслу она ни казалась, нужно иметь мужество ее принять, поскольку она верна, т. е. соответствует современным научным данным.

Но так ли это? Действительно ли естественнонаучная картина мира позволяет сделать все приведенные выше далеко идущие выводы? Люди, профессионально занимающиеся наукой, знают, что она практически никогда не дает ответов на мировоззренческие вопросы, по определению лежащие вне сферы ее компетенции. С другой стороны, философские доктрины, рассматривающие мировоззренческие вопросы, как правило, не делают достаточно точных предсказаний, которые можно было бы проверить научными методами. Тем не менее между философией и наукой нет непреодолимой пропасти. Наука не может доказать или опровергнуть философскую теорию, но всегда можно оценить, в какой степени вытекающие из этой теории следствия совместимы с накопленными наукой данными. То есть философскую доктрину можно оценить не только по ее логичности и последовательности, но и по соответствию эмпирическим фактам тех выводов, которые из нее можно сделать. Если соответствие окажется натянутым или мало правдоподобным, это может подтолкнуть к ревизии доминирующей философской доктрины, что, в свою очередь, неизбежно отразится и на содержании научных теорий. Этот процесс, однако, вряд ли когда-нибудь приведет к окончательному ответу на мировоззренческие вопросы.

В данной книге мы попробуем пройти по узкой тропе между философией и наукой, оставаясь преимущественно на территории науки, но делая рейды и на другую сторону. Попытаемся «с высоты птичьего полета» обозреть существующие сведения о связи психических явлений с активностью мозга и понять, действительно ли эти сведения однозначно поддерживают доминирующий в настоящее время материалистический редукционистский взгляд на природу психических процессов.

Значение внутреннего мира для любого человека невозможно переоценить – в сумме это и составляет то, что человек имеет в виду, когда говорит: «Я». Это то, что является необходимым и до-

статочным свидетельством его существования и в наличии чего невозможно сомневаться. Во всяком случае, так оно видится с позиции нашего субъективного восприятия. Но что говорят об этом философы и ученые? И тут всякая однозначность сразу же исчезает, поскольку философы и ученые по сей день продолжают вести дебаты о природе сознания и разума. Размежевание между философскими системами во многом определяется тем, как они решают данный вопрос.

Слова «ум», «разум», «психика», «ментальность» неточно отражают то, о чем я хочу порассуждать в этой книге. Все они имеют оттенки смысла, ограничивающие область их применения. То, что я имею в виду, включает все явления и процессы нашего внутреннего мира – сознание (но также и неосознанное), мысли и чувства, восприятие внешнего мира и память. Точнее всего смысл передает английское слово «mind». В дальнейшем я буду использовать слова «разум», «психическое» и «ментальность» именно в таком широком смысле. Обозначенная здесь тема поистине необъятна и было бы неразумно пытаться всесторонне охватить ее в одной книге. Моя задача гораздо скромнее – рассмотреть некоторые, наиболее, на мой взгляд, значимые данные о природе взаимоотношений между активностью мозга и психическими процессами и попытаться понять, в какой степени эти данные согласуются с доминирующим сегодня в философии и науке представлением о психическом как об эпифеномене активности мозга.

Каждый человек понимает, что его внутренний мир не идентичен тому миру, который он воспринимает как внешний. Внешний мир подчиняется физическим законам. Мы не можем мгновенно переместиться в другую точку пространства, вернуться в прошлое или попасть в будущее, но легко можем это делать в наших мыслях. Поэтому до относительно недавнего времени и простые люди, и философы не сомневались в том, что наш мысленный мир (душа) и физический мир, к которому относится и наше тело, – две разные сущности. В философии такая позиция называется дуализмом, и в наиболее завершенной форме она была сформулирована в XVII в. французским философом Рене Декартом. В последующие столетия дуализм постепенно вышел из моды (хотя и не исчез полностью) и заместился разными формами монизма, признающего существование лишь какой-то одной субстанции. Материалистический монизм, или физикализм, отрицает существование чего-либо, кроме материальных физических объ-

ектов. Существуют и другие формы монизма (ментальный и нейтральный монизм), но физикализм сегодня, безусловно, является доминирующей философией.

Начнем с современных представлений о том, как этот феномен – человеческий разум – появился в процессе эволюции, а затем перейдем к анализу данных нейробиологии о природе и механизмах связи между активностью мозга и разными сферами ментальной жизни человека. Сначала рассмотрим философские аспекты проблемы – разные варианты объяснения природы психического с точки зрения физикализма – и попробуем оценить их последовательность и непротиворечивость, а также вытекающие из них следствия, которые потенциально можно проверить эмпирически. Последующие главы будут посвящены обзору научных теорий и эмпирических данных, относящихся к нейробиологии разных аспектов психической жизни, и обсуждению соответствия этих данных постулатам физикализма.

ЭВОЛЮЦИЯ ЧЕЛОВЕКА

С антропоцентрической точки зрения вершиной эволюции является человек. Можно думать, что человек, в особенности его мозг, – это наиболее сложно устроенный биологический объект и, соответственно, его следует считать вершиной эволюции. Это, однако, не соответствует видению биологов-эволюционистов. В «древо эволюции» человек (*Homo sapiens*) – лишь маленькая, совсем недавно появившаяся веточка рода *Homo* семейства гоминидов в отряде приматов. По критериям эволюционного успеха (ареал распространения, эффективный размер популяции, длительность периода существования в изменяющихся условиях среды) человек ни в какое сравнение не идет с многими видами беспозвоночных, одноклеточных и вирусов. Более того, в целом, как указывает Е.В. Кунин [2012], чем сложнее организм, тем меньший эффективный размер популяции он имеет и, по единственно разумному определению эволюционного успеха, тем менее успешным он является.

Должны ли мы, однако, ставить во главу угла критерий эволюционного успеха? *Homo sapiens* – это единственный вид, способный конструировать сложные искусственные системы и создавать теоретические модели устройства мира. Можно, конечно, сказать, что этот критерий имеет значение лишь для нас или для гипотетических похожих на нас разумных инопланетян. Однако любые точки зрения, включая мнение эволюционных биологов, имеют значение только для разумных существ. В рамках материалистической картины мира (она тоже является продуктом человеческого ума) никакая точка зрения по большому счету не важна, так как природа в целом бесцельна и бессмысленна. Поскольку в рамках материалистической картины мира большинство замечательных свойств человека, выделяющих его из ряда остальных биологических существ, связаны с активностью мозга, интересно вкратце рассмотреть данные о том, как человеческий мозг появился в процессе эволюции.

Подтриба (subtribe) *Hominina*, включающая из ныне живущих только человека, отделилась от подтрибы *Panina* (шимпанзе) 4–7 млн л.н. [Gibbons, 2012]. Ранние представители *Hominina* – ав-

стралопитеки. Считается, что они преимущественно передвигались на двух ногах, хотя и не все согласны с таким выводом [Oxnard, 1975; Shipman, 1994]. По другим характеристикам (включая размер мозга) они мало отличались от шимпанзе [Szpak, 2007]. Большинство специалистов предполагают, что род *Homo* появился в Африке около 2 млн л.н. и произошел от какого-то вида австралопитеков, хотя до сих пор нет согласия по поводу конкретного вида [Ibid]. Недавняя находка отпечатков ног древних гоминидов на о. Крит, датируемых 6 млн лет, если будет подтверждена, перевернет представление о появлении первых гоминидов в Африке [Gierliński et al., 2017].

Кроме *H. sapiens* род *Homo* включал *H. erectus* и *H. neanderthalensis* [Schrenk et al., 2007]. *H. erectus* широко распространились по Африке и Евразии. Пропорции тела и походка *H. erectus* мало отличались от человеческих. У них было плоское, как у человека, лицо, выступающий нос и редуцированный волосяной покров. Размер мозга, хотя и больший, чем у австралопитеков, был меньше, чем у *H. sapiens*. Тем не менее *H. erectus* были способны охотиться группами на средних и крупных животных, пользовались огнем и каменными орудиями. Они могли заботиться о больных или раненных соплеменниках. У них были начатки мореплавания и элементы искусства и, вероятно, примитивный язык, хотя анатомические особенности строения черепа указывают на ограниченные возможности продуцировать речь [Latimer, Ohman, 2001].

H. erectus существовали более миллиона лет и, как считается, около 500 тыс. л.н. постепенно дали начало другим видам, включая *H. sapiens* [Indriati et al., 2011]. Имя *H. habilis* было присвоено ископаемым останкам, обнаруженным в 1960 г. в Танзании. По предположению ряда ученых, эти существа жили 1.8 млн л.н., были прямоходящими и могли использовать каменные орудия. Эти выводы, однако, основаны на данных, валидность которых подвергается сомнению. Так, Лики и Левин пишут, что из нескольких дюжин образцов, которые в разные времена приписывались этому виду, по крайней мере 50 % к нему не относятся, но нет согласия по поводу того, какие 50 % должны быть исключены [Leakey, Lewin, 1992]. Известный антрополог Бернард Вуд считает, что обнаруженные останки *H. habilis* правильнее относить к австралопитекам [Wood, 1994] или выделить в особый род, а не рассматривать как промежуточное звено в эволюции человека [Wood, 2014].

Неандертальцы (*H. neanderthalensis*) – вымерший вид или подвид древних людей – жили в Евразии вплоть до 40 тыс. л.н. [Galván et al., 2014]. Причины их исчезновения не вполне понятны. В начале XX в. господствовало представление, согласно которому неандертальцы были примитивными, глупыми и жестокими существами [Drell, 2000]. Более поздние находки показывают, однако, что неандертальские технологии были довольно сложными и включали каменные орудия, использование огня, постройку каменного очага в пещерах, изготовление одежды, в том числе тканой, мореплавание в Средиземном море, использование лечебных растений и лечение тяжелых травм, а также готовку и консервирование пищи [Hardy et al., 2020]. Целый ряд образцов палеолитического искусства теперь приписывают неандертальцам, включая наскальную живопись и музыкальные инструменты [Turk et al., 2018]. Эти древние люди, вероятно, обладали речью [Dediu, Levinson, 2018]. Неандертальцы скрещивались с *H. sapiens* и около 2 % генома современных людей сохранилось от неандертальцев [Vernot, Akey, 2014]. Размер мозга у неандертальцев был даже больше, чем у современных людей, однако преобладали области, связанные с сенсорным восприятием (затылочная кора) и контролем движений, и пропорционально меньше были области ассоциативной коры, связанные с языком, вниманием, рабочей памятью, принятием решений и социальными способностями [Kochiyama et al., 2018].

Человек как биологический вид появился 300–200 тыс. л.н. в Африке [Callaway, 2017]; он отличается от родственных современных видов целым рядом морфологических, физиологических и поведенческих признаков, в том числе двуногость, снижение полового диморфизма, удлинённый период онтогенеза (включая беременность и период младенчества), увеличенный размер мозга, использование сложных искусственных орудий, языка и сложной социальной организации. Большинство этих признаков связаны друг с другом, так что одно является неизбежным следствием другого.

Например, хождение на двух ногах освобождало передние конечности от ходьбы, что позволяло использовать их для других целей, в том числе, при наличии необходимых когнитивных и двигательных способностей, для создания и использования искусственных орудий. Однако хождение на двух ногах требовало изменения формы костей скелета, включая таз. У двуногих гоминидов

более короткий чашеобразный таз и, соответственно, уменьшенный по сравнению с обезьянами, использующими для хождения передние конечности (например, гориллы), родовой канал.

Хотя некоторое расширение родового канала по сравнению с ранними австралопитеками имело место, позволяя проходить новорожденным с увеличенным размером черепа, оно было ограничено верхней частью таза, так как дальнейшее увеличение могло препятствовать нормальному двуногому движению [Strickberger, 2000]. Предполагается, что это стало лимитирующим фактором увеличения размера мозга у ранних людей и вызвало более короткий период беременности, что привело к относительной незрелости человеческого потомства, которое не может ходить раньше 12 месяцев и по сравнению с другими приматами является гораздо менее развитым в раннем возрасте [Curry, 2008].

Повышенный рост мозга после рождения и возрастающая зависимость детей от взрослых оказали большое влияние на репродуктивный цикл женщин и более частое появление у людей, по сравнению с другими гоминидами, аллопарентинга (участия в воспитании не прямых родителей) [Hrdy, 2011]. Задержка полового созревания человека привела также к развитию менопаузы, которая, по одной из интерпретаций, позволяет пожилым женщинам заботиться о потомстве своей дочери, вместо того, чтобы иметь больше собственных детей [Wayman, 2013]. Снижение полового диморфизма также интерпретируется как следствие тенденции к долговременным парным связям, способствующим участию обоих родителей в поддержании потомства на протяжении периода его зависимости от взрослых [Lovejoy, 2009].

1.1. Большой мозг

Увеличение размера мозга (в особенности некоторых его частей) считается самым значимым эволюционным достижением человека, которое сделало возможным появление типично человеческих когнитивных способностей. Поэтому очень важно понять, какие механизмы могли стоять за этим явлением. Хотя генетический дрейф может привести к случайной фиксации конкретных аллелей в одном месте, вероятность того, что эволюционная последовательность изменений во множестве точек генома прогрессивно ведет в определенном направлении, например, к увеличению размера мозга, чрезвычайно мала [Schoenemann, 2006]. Недавняя

находка в Марокко ископаемых останков *H. sapiens* отодвинуло предполагаемое время появления современного человека почти на 100 тыс. лет [Richter et al., 2017]. По этим новым представлениям, 300 тыс. л.н. люди, возможно, были уже широко распространены по всему Африканскому континенту. Форма черепа этих людей несколько отличалась от современной (была более продолговатой), но в остальном отличия были незначительными [Sample, 2017]. Они использовали огонь, обработанные каменные орудия, охотились на газелей, что, в общем-то, умели делать и *H. erectus*.

Интересно, что поведенческие признаки когнитивных способностей, характерных для современного человека, такие как символическое поведение [D’Errico et al., 2005] и использование огня для производства орудий [Brown et al., 2009], обнаруживаются не ранее 70 тыс. л.н. К сугубо человеческим особенностям поведения относят то, что можно найти во всех современных человеческих культурах, но что отсутствует или слабо выражено в поведении близких животных видов. Эти особенности включают способность абстрактного мышления, долговременное планирование, торговлю, кооперативный труд, украшение тела, контроль и использование огня и большую роль социального обучения [Nakahashi, 2013]. Археологические признаки, свидетельствующие о наличии этих особенностей, включают культовые захоронения, рыболовство, фигуративное искусство, применение пигментов и камней для украшения тела, использование костяных и композитных орудий, перенос ресурсов на большие расстояния, изготовление каменных лезвий (blade technology), построение очагов [Henshilwood, Marean, 2003].

По теории «палеолитической революции», анатомически современные люди не были когнитивно и поведенчески современными вплоть до 50 тыс. л.н., когда имела место их экспансия из Африки в Европу и Азию [Tattersall, 2009]. Авторы этой теории считают, что причиной «революции» послужило единичное явление в виде, например, генетической мутации. Другие авторы, однако, полагают, что «революция» началась раньше (порядка 100 тыс. л.н.), происходила постепенной и была связана не с единичным событием, а с периодом ускоренного развития, слишком быстрого для классического дарвиновского процесса, но достаточно постепенного [Henshilwood, Marean, 2003]. В любом случае, физиологическое обеспечение этого поведения (развитый мозг, не уступающий или мало уступающий мозгу современного человека)

возникло на 100–200 тыс. лет раньше, чем обнаруживаемые поведенческие последствия. Все это время, по дошедшим до нас сведениям, поведение *H. sapiens* мало отличалось от поведения *H. erectus*, несмотря на значительное превосходство в размерах и сложности мозга. Это ставит под вопрос непосредственное преимущество энцефализации для выживания и позволяет сомневаться в том, что человеческий мозг появился в результате сугубо дарвиновских процессов естественного отбора.

Действительно, обладание большим мозгом связано с большими эволюционными издержками, которые в сумме должны приводить к уменьшению количества потомства, производимого в единицу времени. Во-первых, мозг человека энергетически – очень затратный орган. Составляя лишь 2 % от веса тела, он потребляет 20 % энергетического бюджета [Clarke, Sokoloff, 1999]. Во-вторых, как обсуждалось выше, увеличение размера мозга у приматов коррелирует с более длительным периодом беременности, увеличенным периодом детской зависимости и задержкой деторождения, что в сумме снижает количество потомства, которое особь может произвести в единицу времени. В-третьих, у двуногих гоминидов существует очевидный компромисс между двигательной способностью и легкостью родов [Lovejoy, 2009]. В-четвертых, потенциально существует проблема охлаждения большого мозга [Falk, 1990].

Большой мозг, таким образом, – это не какой-нибудь аппендикс, который, хотя ни для чего и не нужен, но и, как правило, не очень мешает. Наличие такого энергозатратного органа может оправдываться только в том случае, если он резко увеличивает шансы выжить и произвести потомство, иначе, в условиях жесткой борьбы за существование, люди с меньшим размером мозга имели бы явное преимущество. Мозг – это не рога или копыта, функции которых и их значение для выживания лежат на поверхности. Сам по себе размер мозга не может быть признаком, подверженным естественному отбору. Отбираться могут лишь проявления когнитивных способностей, предположительно связанных с размером мозга. Однако очевидные и чрезвычайно полезные для жизни проявления когнитивных способностей, такие как развитие животноводства и земледелия, появились гораздо позже, чем сформировался не отличающийся от современного мозг. Предложено несколько гипотез для объяснения этого явного несоответствия фактов дарвиновской теории эволюции. Предполагается, в

частности, что связанные с энцефализацией поведенческие изменения первоначально были ограничены социальной сферой – увеличением размера социальных групп [Dunbar, 1992] и усилением эмпатии [Klein, 2014].

Наиболее популярна гипотеза Дунбара, согласно которой движущей силой эволюции мозга у гоминидов было увеличение размеров социальной группы [Dunbar, 1998]. Понятно, что выживать группой легче, чем поодиночке, но жизнь в группе тоже не проста, так как требует соблюдения баланса между эгоистическими драйвами каждого индивида и сохранением гармонии в группе. Нахождение оптимального баланса требует определенных когнитивных способностей, что, по теории Дунбара, и было движущей силой эволюции мозга человека. Основной эмпирический факт, на котором эта гипотеза основана, – корреляция между размером мозга и размером социальной группы у современных приматов. Дунбар предполагает, что размер мозга (по крайней мере, определенных его частей) и развитость соответствующих когнитивных способностей (память, способность манипулировать информацией) напрямую определяют количество соплеменников (число Дунбара), с которыми данная особь может поддерживать отношения. По его расчету, для современного человека это число составляет 150, т. е. в среднем современный человек способен (исходя из средних размеров человеческого мозга) поддерживать стабильные отношения со 150 другими людьми [Purves, 2008]. Соответственно, по теории Дунбара, именно размер социальной группы был параметром отбора в эволюции человека, что косвенно вело к отбору особей с большим размером мозга.

Несмотря на широкую популяризацию этой теории, ее выводы критиковались многими специалистами. Либерман, например, считает, что максимальный размер группы, которую можно было прокормить в палеолитические времена без наличия хотя бы rudimentарной агрикультуры, составлял 30–50 человек и, таким образом, группы в 150 особей не могли быть фактором отбора [Lieberman, 2013]. Множество видов с размером мозга гораздо меньше человеческого, при этом даже не относящихся к млекопитающим и позвоночным (например, социальные насекомые с достаточно примитивной нервной системой), способны поддерживать сложную иерархическую социальную структуру [Chittka, Niven, 2009; Attenborough, 2013]. Кроме того, компьютерные

программы, симулирующие виртуальных автономных агентов с простым набором функций, воспроизводят сложные социальные структуры, описанные приматологами [Pfeifer, Bongard, 2006]. Сравнение диет у разных видов приматов, описанных Дунбаром, показывает, что выявленная им корреляция между размером мозга и социальной группы легко объясняется различиями в диете [DeCasien et al., 2017]. Кроме того, корреляция не объясняет направления каузальных связей и естественней предположить, что высокий интеллект является причиной формирования сложных социальных групп, а не наоборот [Kluger, 2010].

Согласно другой популярной теории, причиной увеличения человеческого интеллекта (и мозга) в процессе эволюции стало снижение агрессивности (теория самоприручения) [Klein, 2014]. Последнее выделило человека из других родственных видов приматов и в конечном счете привело к появлению типично человеческих черт, таких как эмпатия, социальное познание и культура [de Waal, 1989]. Эта теория находит поддержку в экспериментах с приручением диких животных, у которых уже через несколько поколений наряду со снижением агрессии наблюдаются признаки «человеческих» особенностей социального поведения [Trut et al., 2004].

Этологи также отмечают, что животные с более мягкой и расслабленной манерой взаимодействий друг с другом (косопалые макаки, орангутанг и бонобо) демонстрируют более развитые социально-когнитивные способности, чем более агрессивные шимпанзе и бабуины [de Waal, 1989; Hare et al., 2012]. Критики данной теории считают нелогичным связывать появление человеческого вида со снижением агрессивности ввиду большого количества насильственных действий на протяжении всей истории человечества. Однако защитники теории указывают, что речь тут должна идти лишь о реактивной, а не проактивной агрессии, к которой относится подавляющее большинство исторических примеров агрессивных действий. По их мнению, увеличение плотности популяции у предков современного человека должно сочетаться с усилением толерантности к соплеменникам и снижением реактивной агрессии [Wrangham, 2018].

Как показывают недавние генетические данные, около 200 тыс. л.н., когда имели место наиболее драматические изменения развития мозга, популяция *H. sapiens* на самом деле снижа-

лась [Wrangham, 2019]. Лучше всего эмпирически подтвержденные проявления «самоприручения» у *H. sapiens* относят к гораздо более позднему периоду – между 40 и 25 тыс. л.н. – и объясняют результатом влияния культуры на процессы отбора половых партнеров [Bednarik, 2008].

Другая, наиболее популярная в настоящее время, теория связывает снижение агрессивности с развитием языка. Предполагается, что улучшение языковых способностей позволило усилить контроль над агрессивными членами сообществ охотников-собирателей. Те, кто пытался добиться господства над другими, подвергались наказанию, чему способствовала достигаемая с помощью языка договоренность. Язык позволял более слабым и менее агрессивным членам сообщества сотрудничать, координируя планы по подавлению попытки доминирования со стороны агрессора и устраняя его при необходимости. Со временем это привело к отбору против реактивной агрессии [Wrangham, 2019].

Иного рода попытки объяснить эволюцию интеллекта (и мозга) социальными процессами включают теорию социального обмена, согласно которой в процессе эволюции могли отбираться группы, способные заключать взаимовыгодные сделки, при которых две стороны обмениваются некоторыми продуктами, получая взаимную выгоду. Чтобы такие обмены были эффективными, должны существовать способности выявления обмана. Моделирование этих процессов на основе эволюционной теории игр показало, что они могут стабильно отбираться в эволюции, только если предусмотрены механизмы выявления и наказания обманщиков [Cosmides et al., 2010].

Одной из популярных теорий является теория полового отбора, выдвинутая еще Чарльзом Дарвином и в наши дни в отношении эволюции интеллекта развиваемая Джеффри Миллером [Miller, 2000]. По убеждению последнего, человеческий интеллект развивался непропорционально необходимым для выживания охотников-собирателей способностям. Такие проявления интеллекта, как язык, музыка и изобразительное искусство, появились не потому, что они были нужны для выживания, а потому, что интеллект расценивался как показатель приспособленности. Соответственно, отбирались особи с более высоким интеллектом по механизму положительной обратной связи, впервые описанному Фишером [Fisher, 1915]. У моногамных видов, к которым принадлежат большинство гоминид, ассортативное спаривание ведет к тому, что

особи, обладающие сексуально отбираемым свойством (в данном случае высокий интеллект), обычно сочетаются друг с другом, и, соответственно, особи, не обладающие этим свойством, могут найти только такого же, как они сами, партнера. В ряду поколений это уменьшает вероятность выживания последних и ведет к накоплению отбираемого качества.

Теорию полового отбора критиковали уже современники Дарвина. Так, по предположению Альфреда Валлиса, животные и птицы не выбирают партнеров по их красоте, а артистические способности человека говорят о его духовной природе и не связаны с механизмами естественного отбора [Fisher, 1915]. В противовес теории полового отбора была предложена теория, согласно которой определенные виды символического поведения, включая пение, танцы, раскраску тела и покррой одежды, выполняли скорее функцию запугивания хищников и соперников, и было бы ошибочно считать их результатами полового отбора [Jordania, 2011]. В целом, многие элементы полового поведения у людей не могут быть объяснены механизмами полового отбора и выполняют скорее социальную функцию [Roughgarden, 2004].

Согласно модели экологического доминирования и социальной конкуренции (ecological dominance-social competition), человеческий интеллект развился вследствие усиления доминирования в своей экологической нише и увеличения важности социальных взаимодействий. В результате главный прессинг естественного отбора в отношении интеллекта сместился с умения справляться с естественным окружением к умению доминировать над представителями собственного вида [Flinn et al., 2005]. В современном обществе прямое адаптивное преимущество большого мозга сомнительно, так как по генетическим данным гены, контролирующие размер головы, не находятся под сильным прессингом отбора в современной популяции [Ossorio, 2011].

Подводя итог краткому обзору данных по эволюции человека, нужно подчеркнуть их фрагментарность, что неизбежно в этой области исследований. Несмотря на длинную историю палеоантропологии и значительное количество накопленных данных, отсутствует общепринятое представление о хронологии эволюционных событий и «родословной» предков человека. Единичная находка может радикально изменить устоявшиеся представления и датировки. Так, находка в Марокко ископаемых останков *H. sapiens* отодвинула предполагаемое время их появления почти на

100 тыс. лет [Richter et al., 2017], а недавняя находка отпечатков ног древних гоминидов на о. Крит, датированных 6 млн лет, превращает представление о появлении первых гоминидов в Африке [Gierliński et al., 2017].

Другой особенностью исследований в данной области является тенденция интерпретировать каждую находку в качестве нового вида предков человека, что объяснимо горячим желанием найти «недостающие звенья» и выстроить родословную современного человека в свете теории Дарвина. Еще отцы неodarвинизма Эрнст Майер и Джордж Симпсон описывали современную им таксономию гоминидов как «сбивающее с толку разнообразие имен» [Mayr, 1950] и «хаос антропологической номенклатуры» [Simpson, 1963]. С тех пор ситуация только ухудшилась. Многие современные палеоантропологи продвигают каждую новую находку как свидетельство биоразнообразия древних гоминидов (например, [Tattersall, Schwartz, 2000]), хотя для этого часто нет оснований, и несходства, выявляемые между различными находками, фактически лежат в границах обычной внутривидовой вариации [White, 2003]. Судя по генетическим исследованиям, даже неандертальцы и *H. erectus* скрещивались с *H. sapiens*, что может поставить под сомнение их принадлежность к разным биологическим видам.

В целом, объяснение эволюции человека с позиций неodarвинизма сталкивается с трудностями, уже рассмотренными выше. Естественный отбор не может руководствоваться долговременными целями и выбор происходит только из уже существующих вариантов. Большой мозг в любом случае должен быть бременем, против которого должен работать естественный отбор. Если при отсутствии видимых преимуществ большого мозга обладающие им особи не элиминировались, это может говорить лишь о его слабости и неэффективности в силу, например, малой конкуренции и достаточности ресурсов [Кунин, 2012]. Кроме того, если положительный естественный отбор был причиной выживания существ с большим мозгом, то почему не исчезали конкурирующие особи с малым мозгом?

Действительно, данные палеонтологии показывают, что в течение большего времени существования человека, анатомически не отличимого от современного, параллельно с ним в этих же условиях продолжали существовать и предки шимпанзе, и австралопитеки, и *H. erectus*, и неандертальцы. Наконец, в ходе естественного отбора могли отбираться особи не с большим мозгом, а лишь

с поведенческими проявлениями высокого интеллекта, которые, по имеющимся данным, появились на 100–200 тыс. лет позже, чем сам большой мозг. Теории, рассматривающие социальность как движущий фактор эволюционного развития мозга, в большинстве своем спекулятивны и не могут объяснить, почему этот фактор действовал у человека, но не привел к появлению интеллекта у других социальных животных.

Предложено множество гипотез, связывающих увеличение размера мозга у предков человека с действием побочных факторов, таких как прямохождение, вызванное изменением климата, потеря волосяного покрова у живущих рядом с водой обезьян, хищничество (killer ape), увеличение размера мозга из-за хорошего питания, использование огня, язык и пр. Однако ни один из этих факторов при ближайшем рассмотрении не может рассцениваться как единственная причина. Прямохождение довольно часто встречается в животном мире, включая птиц, ящеров, грызунов и пр., и ни в одном случае оно не сочетается с достаточно большим мозгом и сравнимым интеллектом. Распространение в новые ареалы в силу изменения климата или по другой причине – частое явление в истории животного мира, но связанные с этим изменения интеллекта не отмечены. Безволосые, а также использующие достаточно богатый язык (например, некоторые насекомые) виды не выделяются высоким интеллектом. Хорошее питание неизбежно ведет к увеличению размеров популяции и иногда размеров тела, но никогда к появлению интеллекта, сравнимого с человеческим.

Таким образом, все эти отличительные черты человека не связаны друг с другом (увеличение размеров мозга, например, у слонов и дельфинов не привело к появлению прямохождения) и их сочетанное появление у одного вида должно рассматриваться как серия счастливых случайностей. То есть так же, как и в эволюции всего живого мира [Кунин, 2017], счастливый случай должен был играть ключевую роль в эволюции человека.

Наиболее правдоподобный сценарий выглядит так: серия «счастливых случайностей» позволила первоначально появиться набору свойств, таких как прямохождение, освободившее руки, и некоторое увеличение интеллекта, повлекшее развитие примитивного языка и использование примитивных орудий и огня. Данные достижения для всех обладающих ими видов создали некую «подушку безопасности», обеспечившую выживание в течение последующих нескольких сотен тысяч лет. Поскольку указанные виды

имели явное преимущество по сравнению с видами, не обладавшими такими достижениями, конкуренция их друг с другом была относительно невысока, что привело к «случайному» появлению большого мозга у части из этих видов и выживание, несмотря на гандикап.

Параллельно нарабатывались механизмы (в первую очередь социальные, обусловленные совершенствованием языка), давшие в конечном счете возможность сделать рывок к поведенческим проявлениям, отличающим современного человека. На протяжении всего этого процесса структурные изменения возникали раньше, чем связанные с ними изменения поведения, полезные для выживания. Так же, как и в других случаях появления «нередуцируемой сложности», отдельные компоненты (такие как прямохождение, большой мозг, социальность) сами по себе либо не давали заметного преимущества, либо были умеренно вредны в конкуренции с ближайшими родственниками, и лишь по прошествии немалого времени, в течение которого все эти компоненты «притирались» друг к другу, произошел резкий скачок. Таким образом, в эволюции человека, так же как и во всех других случаях, появление сложности происходило не благодаря, а скорее вопреки механизмам естественного отбора и может быть объяснено только чередой счастливых случайных совпадений [Кунин, 2012].

ПСИХИЧЕСКОЕ И МОЗГ – ФИЛОСОФСКИЕ АСПЕКТЫ

На протяжении этой книги мы постоянно будем обращаться к философским аспектам соотношения между психическими процессами и активностью мозга, но вначале нужно очертить общие контуры проблемы. Декарт и многие другие философы прошлого рассматривали материальный и ментальный миры как две разные сущности. Эта дуалистическая позиция представляется крайне неудовлетворительной большинству современных философов, прежде всего потому, что в рамках науки сегодня трудно представить, какова может быть природа ментального мира и как он может взаимодействовать с материальным. Доминирует в современном мировоззрении, безусловно, материалистический монизм, отрицающий существование чего-либо, помимо материи. Один из аргументов в его пользу – самодостаточность физических явлений. Каждый физический эффект имеет достаточную физическую причину. Физическая вселенная содержит в себе все ресурсы для полного причинного объяснения любого из ее элементов и в этом смысле является «полной», т. е. для ее объяснения нет необходимости прибегать к чему-либо нефизическому [O'Connor, Churchill, 2010].

Вышеуказанная позиция хорошо укладывается в рамки классической физической картины мира, хотя открытия квантовой механики ставят под сомнение ее самоочевидность [Stapp, 2009]. Тем ни менее давайте обсудим, как такие явления, как сознание и разум, могут рассматриваться с позиции материализма в его современной редакции. Поскольку материализм отрицает существование нематериальных явлений, или считает, что они определяются движением материальных объектов [Smart, 2016], то с позиций материализма нужно либо отрицать существование мыслей, либо считать их побочным (и несущественным) продуктом материальных процессов. Поэтому в рамках материализма спектр возможных интерпретаций природы отмеченных явлений ограничен тремя вариантами. Первое направление получило название «элиминативный материализм», так как он отрицает реальность ментальных явлений [Ramsey, 2020]. Второе – редукционизм – утверждает, что данные явления можно «свести» к активности мозга

[Brigandt, Love, 2017; van Riel, van Gulick, 2019]. Представители третьего направления признают, что, хотя сознание и другие ментальные процессы есть результат активности мозга, их нельзя «свести» к этой активности, т. е. они являются эмерджентной сущностью. Приведенное разделение условно, и в трудах отдельных философов, таких как Деннетт [Dennett, 1991], разные варианты могут сочетаться друг с другом. Рассмотрим аргументы сторонников этих трех позиций.

2.1. Элиминативизм и иллюзионизм

По убеждениям элиминативистов, обыденные представления о ментальных процессах (в том числе их субъективное восприятие) не имеют нейробиологического базиса. На основании этого делается вывод, что данные явления просто не существуют, так как критерием реальности существования ментального явления выступает его редуцируемость до нейробиологического уровня. Так, Квайн, один из ранних элиминативистов, предлагал устранить описания ментальных состояний и заменить их описанием физиологических состояний, поскольку последние «по крайней мере существуют» [Quine, 1960, p. 264]. Более поздние элиминативисты утверждают, что концепции обыденной психологии, такие как вера или желание, недостаточно точно определены и им не будет найден нейробиологический базис. Следовательно, их нужно заменить более точными концепциями когнитивной науки, для которых существует нейробиологический базис [Churchland, 1986]. Эта более мягкая форма элиминативизма, по сути, предлагает заменить одни (ненаучные) концепции другими (научными). Вместе с тем признается, что концепции обыденной психологии, так же как и научные концепции, – это генерализации, но за ними не стоят реально существующие явления [Churchland, 1981].

Однако, если концепции когнитивной науки, как и концепции обыденной психологии, являются генерализациями, то можно лишь утверждать, что они лучше подходят для описания материальных процессов, таких, например, как инструментально регистрируемая активность мозга. Это не удивительно, поскольку концепции когнитивной науки изначально базировались на объективно регистрируемых инструментальными методами явлениях. Нет во всяком случае основания заявлять, что одни концепции отражают реальность, а другие нет. Просто они отражают разную реальность.

Реальность явлений, описываемых концепциями обыденной психологии, подтверждается безоговорочным согласием людей с тем, что эти явления существуют в их ментальном мире.

Вариантом элиминативизма можно считать направление, которое рассматривает ментальные явления как иллюзии. Так, по утверждению Деннетта и других авторов, феномены субъективного восприятия, такие как чувство боли, цвета и прочее, в корне ошибочны, поскольку не соответствуют реальным свойствам объектов и процессам в мозге и потому могут относиться к иллюзиям [Dennett, 1978, 1988; Hardcastle, 1999]. Другие исследователи делают заключение, что феномен сознания вообще – иллюзия [Rey, 1983, 1988; Frankish, 2016, 2017]. По данным теориям интроспекция – это что-то аналогичное сенсорным иллюзиям. Так же, как органы чувств могут создавать образы, которые радикально искажают природу внешнего мира, так и интроспекция дает представления, которые существенно искажают действительную природу нашего внутреннего опыта. Представление об иллюзорности сознания распространяется на иллюзорность и свободы воли [Wegner, 2002], и самости [Hood, 2012]. Если придерживаться позиций иллюзионизма, как называют это течение его апологеты, то научное изучение ментальных явлений фактически не имеет смысла. Можно лишь пытаться ответить на вопрос, почему у нас возникает иллюзия их существования [Frankish, 2016, p. 1]. Поскольку эта точка зрения становится все более популярной в последние годы и нередко появляется в научно-популярных статьях, имеет смысл рассмотреть ее подробнее.

Иллюзионисты утверждают, что феноменальное сознание кажется аномальным в рамках нашего устоявшегося научного мировоззрения, и у нас есть всего три варианта: принять его как реальное и попытаться объяснить либо в рамках современной науки (что кажется почти невозможным), либо предложить серьезные изменения в нашей науке, либо объявить, что «это явление иллюзорно, и приступить к исследованию того, как возникает иллюзия»* [Ibid]. Третий вариант кажется наиболее экономным с точки зрения усилий, необходимых для решения проблемы. Он также более жизнеспособен, чем строгий элиминативизм. Легче убедить кого-то в том, что его убеждения и желания – иллюзии, чем в том, что их просто не существует.

* Здесь и далее перевод мой.

Посмотрим на эту идею более внимательно. Чтобы ее проиллюстрировать, иллюзионисты приводят множество примеров и аналогий. Многие из них являются иллюзиями восприятия, когда мы «видим» какой-то объект, которого на самом деле нет. Деннетт сравнивает сознание с иконками компьютерного интерфейса, которые напрямую не соответствуют структурам внутри компьютера и, следовательно, они – иллюзии пользователя [Dennett, 1991]. Что такое иллюзия? Типичные определения иллюзии в словарях следующие: «идея или убеждение, которое не соответствует действительности» [<https://dictionary.cambridge.org/dictionary/english/illusion>]; «что-то, что вводит в заблуждение, создавая ложное впечатление о реальности» [<https://www.dictionary.com/browse/illusion>]; «обман чувств, нечто кажущееся, то есть искаженное восприятие реально существующего объекта или явления, допускающее неоднозначную интерпретацию» [<https://ru.wikipedia.org/wiki/Иллюзия>].

Какие выводы можно сделать из этих определений? Во-первых, иллюзии всегда субъективны – два человека, смотрящие на одну и ту же сцену (например, представление фокусника), могут воспринимать ее по-разному. У одного есть иллюзия, что уловка реальна, тогда как у другого, знающего правду, этой иллюзии нет. Во-вторых, они возникают в сознании, т. е. они – объекты ментального пространства. В-третьих, иллюзии контрастируют с другими объектами ментального пространства, которые не являются иллюзиями. Когда человек видит мираж оазиса в пустыне, только изображение оазиса есть иллюзия, а изображения песчаных дюн и безоблачного неба – нет.

Что подразумевается под утверждением, что сознание – это иллюзия? Во-первых, чья это иллюзия? Единственным возможным субъектом иллюзии в рамках предложенной парадигмы может быть мозг. Сознание – это «фокус, который мозг разыгрывает над самим собой» [Frankish, 2016, p. 5]. Итак, у мозга возникает иллюзия, что у него есть сознание. Где она возникает? Опять же, в рамках предложенной парадигмы, она может возникать только в мозгу. В каком виде она там представлена? Согласно иллюзионистам, сознание не существует как реальная сущность, но оно «кажется существующим», и эта кажущаяся реальность должна быть где-то в мозге «закодирована». То есть нужно не только выяснить, как мозг выполняет все невообразимо сложные когнитивные операции, порождающие иллюзию сознания, но также определить, как и где ее содержание «закодировано» в мозге.

Таким образом, иллюзионистский подход, который был предложен как самый простой способ избежать решения сложной проблемы сознания, на самом деле делает ее еще более сложной. Наконец, обычный способ определения иллюзии – противопоставить ее другим элементам сознания, которые не являются иллюзиями. Иллюзионистский подход не дает такой возможности, потому что он объявляет все содержание сознания иллюзией. В этом контексте термин «иллюзия» фактически теряет свой смысл, и его применение здесь можно считать примером категориальной ошибки. Поэтому термин «иллюзия» здесь примем лишь как метафору и попытаемся расшифровать, что на самом деле она означает.

Говоря о сознании как об иллюзии, иллюзионисты чаще всего имеют в виду явление, которое философы называют «феноменальным сознанием». Далее, в гл. 3, мы обсудим данный вопрос подробнее и увидим, что определение феноменальных свойств сознания действительно неоднозначно, поэтому можно по крайней мере частично согласиться с утверждением, что феноменальное сознание – иллюзия. Однако из этого не следует, что все содержание сознания (например, ментальные репрезентации внешних объектов и мысли) также является иллюзией. Можно апеллировать (как это часто делают и элиминативисты, и иллюзионисты) к следующему факту: ментальные репрезентации не соответствуют ни физическим объектам, которые они представляют, ни паттернам мозговой активности, сопровождающим восприятие данных объектов, и в этом смысле являются иллюзией [Ibid, p. 1].

В самом деле, хотя наши органы чувств оснащены «устройствами», чувствительными к некоторым физическим свойствам внешних объектов, они не позволяют понять «реальную» физическую природу этих объектов во всей ее глубине. Более того, наше восприятие неизбежно порождает некоторые искажения реальных физических объектов. Мы часто видим то, что ожидаем или хотим увидеть, а не то, что есть на самом деле. Однако это только количественно отличается от научного «восприятия» физических объектов, которое, безусловно, является более точным, но никогда не сможет описать «реальную» физическую природу данных объектов во всей ее глубине. Сенсорные иллюзии не доминируют в нашем восприятии внешнего мира. Они, безусловно, время от времени возникают (у некоторых людей или в некоторых состояниях чаще, чем в других случаях), но вряд ли мы могли бы существо-

вать, если бы органы чувств нас постоянно обманывали. В целом можно признать, что наши органы чувств адекватно отражают внешний мир.

Нет основания считать, что интроспекция, т. е. восприятие внутреннего мира, радикально отличается по надежности от восприятия внешнего мира. Для человека интроспекция является важным инструментом контроля собственных мыслей, ощущений, эмоций и в конечном счете поведения. Так же как искаженное восприятие внешнего мира не позволило бы в нем существовать, искаженное восприятие внутреннего мира не позволило бы человеку адекватно контролировать свое поведение. Кроме того, точно так же, как в отношении объектов внешнего мира мы можем сопоставить наше восприятие с восприятием других людей и проверить таким образом, не является ли наше восприятие иллюзией, каждый человек может сравнить восприятие собственного внутреннего мира с восприятием другими людьми их внутренних миров. Хотя конкретное ситуационное содержание ментальной реальности, безусловно, различно у разных людей и у одного и того же человека в разные моменты времени, общие законы и категории, описывающие эту реальность, настолько похожи, что это позволяет большинству людей надежно ориентироваться в социальном мире, предсказывать желания и мысли других людей в конкретных ситуациях по аналогии с тем, что сам человек думал бы и испытывал в данной ситуации.

Эмпатия, модель психического состояния другого человека (theory of mind), искусство и многие другие явления социальной действительности были бы невозможны, если бы интроспективно воспринимаемые явления ментального мира были бы ошибочны или радикально различны у разных людей. Можно думать поэтому, что они реальны и фундаментально одинаковы у разных людей. Кроме того, если распространить представление об иллюзорности на все ментальные процессы, включая мысли и суждения, то иллюзионизм, так же как элиминативизм, можно считать самоопровергающей теорией, так как и он тоже часть ментального пространства и, таким образом, сам должен быть признан иллюзией [Baker, 1987; Voghossian, 1990, 1991; Reppert, 1992].

Что же касается несоответствия между содержанием сознания и активностью мозга, то в этом и заключается суть трудной проблемы сознания. Утверждение, что, во-первых, реальна только ак-

тивность мозга, и, во-вторых, что содержание сознания является иллюзией, потому что не соответствует активности мозга, – это пример кругового рассуждения.

2.2. Эпифеноменализм

Эпифеноменалисты не отрицают существования ментальных феноменов, но считают их эпифеноменом активности мозга. В обычном употреблении слова «эпифеномен» подразумевается причинная связь между явлениями. В отношении психических явлений суть эпифеноменализма состоит в установлении причинно-следственных связей между мозгом, психикой и поведением. Подразумевается, что мозг эволюционировал как продукт естественного отбора наиболее эффективных поведенческих паттернов. Следовательно, то, что действительно имеет значение, – это двунаправленная причинно-следственная связь между мозгом и поведением. Выбранные естественным отбором модели поведения формировали мозг на протяжении поколений, а мозг формирует поведение конкретного субъекта. В данной картине нет места психике. Значит, психика – это просто побочный продукт мозговой деятельности, и ее можно легко убрать из картины, поскольку она не имеет никакой причинной силы. Состояние мозга порождает одновременно поведение и психическое состояние, причем последнее влияет и на поведение, и на состояние мозга «не более чем тень влияет на шаги путешественника, которого она сопровождает» [James, 1879, цит. по: Robinson, 2019, p. 1], или «свист пара, который сопровождает работу локомотива, влияет на его движение» [Huxley, 1874, цит. по: Robinson, 2019, p. 1].

Ключевой загадкой с эпифеноменологической точки зрения является кажущаяся причинная эффективность ментального содержания. Для каждого человека эта эффективность настолько очевидна, что занимает центральное место в его субъективной картине мира [Robb, Neil, 2021]. Научное обоснование эпифеноменологической концепции должно объяснить, как эта очевидная эффективность ментального содержания опосредуется мозгом. В рамках эпифеноменализма может быть два варианта такого объяснения – редукционизм и эмерджентизм. Первый утверждает, что все психические события можно свести к физическим событиям в мозге, тогда как второй признает, что психические события не сводятся напрямую к активности мозга, а являются эмерджентными состояниями [O'Connor, 2020]. Чтобы быть эпифеноменом, эмердж-

жентные состояния не должны обладать причинной силой, т. е. это должна быть так называемая слабая эмерджентность (в отличие от сильной эмерджентности, обладающей нисходящей причинной силой) [Chalmers, 2006]. Если выйти за пределы эпифеноменализма, то можно утверждать, что рассмотрение сознания как примера сильной эмерджентности в принципе совместимо с материализмом, однако здесь уже определенно пограничный вариант [Bedau, 1997]. Все другие возможные варианты явно несовместимы с ортодоксальным материализмом и требуют принятия других философских позиций, таких как дуализм [Robinson, 2020], нейтральный монизм [Stubenberg, 2018] или панпсихизм [Goff et al., 2020].

2.2.1. Редукционизм

Редукционизм, в отличие от элиминативизма, признает, что ментальные явления (по крайней мере некоторые) реально существуют, но они могут быть «сведены» к физическим процессам в мозге. Это, пожалуй, господствующее в настоящее время представление, в том числе (и особенно) среди специалистов, занимающихся изучением мозга, и далее мы неоднократно будем возвращаться к аргументам «за» и «против» редукционизма. Ключевым тут, конечно, является определение того, что понимается под данной «сводимостью». Рассмотрим, как это определяют философы. «Основной вопрос редукции заключается в том, могут ли свойства, концепции, объяснения или методы из одной научной области (обычно на более высоких уровнях организации) быть выведены или объяснены с помощью свойств, концепций, объяснений или методов из другой области науки (обычно на более низких уровнях организации)?» [Brigandt, Love, 2017, p. 1]. «Утверждение, что x сводится к y , обычно означает, что x есть не что иное, как y ... Если кто-то утверждает, что ментальное сводится к физическому... он подразумевает, что... ментальное может быть возвращено к физическому» [van Riel, van Gulick, 2019, p. 1].

Таким образом, наличия корреляций между ментальными явлениями и активностью мозга недостаточно – эти корреляции можно объяснить даже в рамках дуализма. Какова природа этих корреляций? Как содержание сознания «возникает» из работы мозга? Редукционистский подход предполагает, что кажущееся казуально эффективным ментальное содержание может быть «сведено» к состояниям мозга, т. е. оно должно быть каким-то образом

«закодировано» в мозговой деятельности, и нейробиологический прогресс должен в конечном итоге позволить взломать код (по аналогии с генетическим кодом) и «читать» содержание мыслей прямо по активности мозга. Надежда в конечном итоге взломать код и прочесть содержание сознания по активности мозга жизненно важна для редукционизма, поскольку только в этом случае мы можем с уверенностью утверждать, что соответствующие состояния мозга управляют поведением и одновременно производят эпифеноменальное содержание сознания.

На протяжении всех последующих глав мы будем неоднократно возвращаться к вопросу о редуцируемости ментального содержания к активности мозга. Здесь лишь отметим, что по мере накопления данных о природе мозговой активности надежда «взломать код» и «прочитать мысли» по активности мозга становится все более призрачной; по открытому признанию некоторых авторов, обозначенная проблема является, по-видимому, неразрешимой для большинства ментальных репрезентаций [Hutto, Myin, 2012]. Можно по аналогии с устройством компьютера утверждать, что анализ активности мозга – это вообще не тот уровень анализа, который может позволить выявить содержание сознания [Pylyshyn, 1984]. Действительно, если представить, что, имея в своем распоряжении все необходимые приборы для измерения физических показателей работы компьютера, мы попытались бы на основе этих показателей понять, какую программу выполняет компьютер, то (в общем случае) потерпели бы фиаско.

В каком-то смысле представления редукционистов и элиминативистов о соотношении ментального и физического можно считать противоположными. По убеждению редукционистов, есть соответствие между ментальными процессами и активностью мозга. Именно это соответствие и позволяет объяснить и в конечном счете «свести» ментальные процессы к активности мозга. Элиминативисты же отрицают такое соответствие. Они указывают на отсутствие в известных науке процессах в мозге чего-либо, что напоминало бы содержание наших ощущений и мыслей [Ramsey et al., 1990; Beer, 2000; Chemero, 2009; Hutto, Myin, 2012]. На самом деле позиции редукционистов и элиминативистов часто совпадают, так как, по мнению редукционистов, все ментальные явления, которым нет соответствия в активности мозга, можно считать иллюзией [Dennett, 1991].

2.2.2. Эмерджентизм

Третья, потенциально совместимая с материализмом концепция состоит в том, что сознание – это эмерджентная сущность [Chalmers, 1995, 1996, 2006]. Философская концепция эмерджентной сущности была представлена Льюисом в 1875 г. [Lewes, 1875] и многократно использовалась теоретиками для объяснения свойств ряда явлений. Разные теоретики вкладывают в нее разный смысл, но общим во всех определениях является то, что эмерджентная сущность «возникает» из более фундаментальных сущностей, но не «сводится» к ним [O'Connor, 2020]. Эти идеи прикладывались, например, к объяснению свойств живых объектов и имели целью найти «средний путь» между редукционизмом, согласно которому свойства живых объектов полностью «выводятся» из физических и химических процессов, и витализмом, постулирующим существование энтелехии, т. е. некоей нематериальной сущности, определяющей специфику биологических объектов [Ibid].

Эмерджентизм существует в разных вариантах, начиная с эпистемологического (нам лишь кажется, что законы разных уровней не сводимы друг к другу в силу нашего ограниченного знания) и кончая онтологическим (законы разных уровней нельзя свести друг к другу даже при наличии полной информации). Различают слабый и сильный эмерджентизм. В первом случае законы высшего уровня в принципе можно вывести из законов низшего, хотя они и кажутся «удивительными», во втором случае это невозможно [Chalmers, 2006]. В основе теории слабой эмерджентности лежат три постулата: (1) физический монизм, (2) тезис о системных свойствах и (3) тезис синхронного детерминизма [Bedau, 1997; Stephan, 2002].

(1) Физический монизм постулирует, что системы, обладающие эмерджентными свойствами, состоят исключительно из материальных частей, т. е. отрицается существование нематериальных сверхъестественных сущностей, ответственных за возникновение эмерджентных свойств. (2) Эмерджентные свойства являются системными свойствами. Свойство считается системным, если (и только если) им обладает система, но не обладает ни одна из ее частей. (3) Свойства системы номологически зависят от свойств ее частей и их взаимодействия друг с другом. Не может быть различия между состояниями системы без того, чтобы было различие в свойствах ее частей или в их взаимодействии друг с другом [Stephan, 2002]. Этот последний тезис очень важен, так как

позволяет проверить применимость теории слабой эмерджентности к феноменологии психического. В приложении к содержанию сознания он означает, что, во-первых, разные содержания сознания должны иметь разную физическую основу (т. е. соответствовать разным паттернам мозговой активности) и, во-вторых, одно и то же содержание сознания должно иметь одну и ту же физическую основу как при повторных его появлениях у одного и того же субъекта, так и у разных субъектов. Эти постулаты доступны эмпирической проверке, и мы подробнее рассмотрим их в гл. 3.

Слабый эмерджентизм по сути совместим с редукционизмом, однако, хотя поведение системы на высшем уровне и вытекает из динамики и структуры систем низшего уровня и в принципе может быть вычислено из них, на практике его можно воспроизвести только с помощью симуляции [Bedau, 1997]. Слабый эмерджентизм отрицает существование нисходящей каузальности. То есть, как и в случае с редукционизмом, эмерджентные свойства являются эпифеноменом.

По теории сильной эмерджентности явления высшего уровня могут влиять не только друг на друга, но и на явления нижнего уровня (например, феномены сознания могут изменять активность мозга) [Campbell, 2010]. Наличие сильной эмерджентности не так легко вписывается в парадигмальный материализм. Бидау пишет, что, хотя существование сильной эмерджентности логически возможно, оно создает ощущение дискомфорта, поскольку похоже на магию. «Как возникает нередуцируемая, но супервентная нисходящая каузальная сила, если по определению она не может быть результатом совокупности потенциальных возможностей нижнего уровня? Такая каузальная сила совершенно не похожа ни на что в нашем научном мировоззрении и очень неудобна для любой разумной формы материализма. Ее загадочность только усиливает традиционное беспокойство, вызываемое тем фактом, что нечто появляется как бы из ничего» [Bedau, 1997, p. 377]. Поэтому многие философы отрицают возможность сильной эмерджентности и оставляют лишь слабую, которая, по их мнению, не представляет угрозы для материализма.

Другие, однако, считают сознание и целеполагание примерами сильной онтологической эмерджентности [Chalmers, 1996; Searle, 1997], так как наш опыт целенаправленной деятельности предполагает наличие прямого контроля поведения, не сводимого к сумме каузальных эффектов соответствующих частей мозга

[O'Connor, 2020]. Философы-материалисты с этим, естественно, не согласны и утверждают, что основанные на интроспекции свидетельства – это лишь иллюзия и эпифеномен нейрофизиологических процессов [Dennett, 1991; Ney, 2010].

Как отмечает О'Коннор, такой подход радикально отличается от подхода ученых XX в., которые отвергали витализм при объяснении жизни и доказывали, что она может быть «сведена» к совокупности физико-химических процессов. Эти ученые разрабатывали экспериментальные и аналитические средства, позволяющие им объяснять явления жизни на основе физико-химических процессов, а не заявляли, что жизнь – это лишь иллюзия. В случае сознания, однако, философы-материалисты соглашались, что на субъективном уровне феномены сознания существуют, но отрицают реальность данного уровня. Это создает проблему с точки зрения эпистемологии, так как все наше знание также находится на уровне сознания и, таким образом, должно считаться иллюзорным [O'Connor, 2020].

Далее попытаемся оценить существующие теории и эмпирические данные о мозговых коррелятах ментальных процессов в плане их совместимости с эпифеноменологической (редукционистской или эмерджентистской) концепцией. За редкими исключениями будем опираться на теории и данные, полученные в рамках доминирующего (mainstream) научного направления, развивающегося под сильным влиянием редукционизма. Постараемся иметь это в виду и в важных случаях будем обращаться к сути полученных эмпирических данных, а не к их интерпретации авторами. Сначала рассмотрим теории и данные о мозговых коррелятах сознания и подсознания, затем современные представления о мозговой основе когнитивных процессов, таких как память, внимание, мышление и исполнительный контроль. Далее перейдем к психическим процессам, связанным с ценностными ориентациями и обсудим данные о мозговых коррелятах эмоций, эстетического восприятия и спиритуальности. В конце подведем итог и выберем наиболее правдоподобный вариант представления о природе взаимоотношений между психическими процессами и активностью мозга.

СОЗНАНИЕ

Для человека сознание лежит в основе всех остальных проявлений ментальности. Лишь благодаря наличию сознания мы вообще можем знать о существовании ментальной сферы. Неудивительно поэтому, что дебаты философов о природе ментальности в основном касаются именно сознания. Термин «сознание» используют в отношении как целого организма (существа, обладающие сознанием), так и отдельных умственных состояний и процессов [Rosenthal, 1986; Gennaro, 1995; Carruthers, 2000]. Животное, персона или другая «когнитивная система» могут считаться обладающими сознанием в разных смыслах. В общем смысле это может быть просто способность ощущать мир и реагировать на него [Armstrong, 1981], при уточнении – способность осознать мир должна быть не потенциальной, а актуальной, т. е. относиться только к состоянию нормального бодрствования, а не сна или комы. Более ограниченное определение требует, чтобы сознательное существо было не только способно осознать мир, но и обладало бы самосознанием [Carruthers, 2000]. Субъективность считается неотъемлемым атрибутом сознания. По определению Томаса Нагеля, существо обладает сознанием, если обладает субъективным видением мира (there is «something that it is like» to be that creature) [Nagel, 1974].

Состояния сознания тоже имеют много определений. Наиболее общее из них трактует сознательные умственные состояния как такие, при которых человек осознает, что находится в этом состоянии [Rosenthal, 1986, 1993]. Выраженные таким образом сознательные состояния содержат в себе метаментальность, поскольку требуют наличия ментальных состояний о ментальных состояниях. В этом отношении мысли или желания, в которых мы не отдаем себе отчета, являются неосознанными. Совершенно другой подход к определению состояний сознания подчеркивает их чувственные качественные характеристики. Для их обозначения применяется термин «квалиа». Однако по поводу природы этих состояний и даже самого их существования разные философы придерживаются разных позиций [Churchland, 1983; Dennett, 1991; Chalmers, 1996]. Квалиа иногда называют феноменальными

свойствами, а соответствующий вид сознания – феноменальным сознанием, причем последнее обычно включает не только чувственные ощущения, но и пространственную, временную и концептуальную организацию нашего восприятия мира и своего места в нем.

Неустранимая сложность всего комплекса этих явлений связана с практической невозможностью посмотреть на них со стороны. Конечно, любой человек может рассуждать о природе сознания и разума, привлекая существующие теории и доступные эмпирические данные, так же, как это делается при рассуждении о природе любого явления внешнего мира. Однако невозможно игнорировать тот факт, что предметом рассуждений в данном случае становится то, что эти рассуждения и производит. Можно представить себе какого-либо внешнего наблюдателя, не являющегося человеком, который, не обладая субъективным опытом человека, мог бы тем не менее заметить явные и очевидные признаки разумной деятельности людей, из которых легко догадаться, что они способны понять законы природы и использовать их для конструирования всевозможных замысловатых машин, т. е. обладают разумом. Но, чтобы это заметить и понять (да и даже начать производить подобные наблюдения), наблюдатель должен быть разумным и, стало быть, должен обладать субъективным опытом сознания и мышления. Вот почему субъективность считается неизбежным атрибутом сознания [Nagel, 1974]. Наука же по определению занимается изучением лишь тех явлений, которые могут быть объективизируемы [Popper, 1975]. Нагель поэтому скептически относится к возможности расшифровки природы сознания в рамках материалистического подхода [Nagel, 1974]. Ниже будут проанализированы возможные подходы к научному изучению сознания, но вначале представляется полезным рассмотреть теорию сознания, предложенную Даниэлем Деннеттом, которого по праву считают наиболее крупным авторитетом в области философского истолкования природы сознания и одним из столпов современного материализма.

3.1. «Сознание объяснено»

Свою программную книгу со скромным названием «Сознание объяснено» («Consciousness explained») Деннетт [Dennett, 1991] начинает с утверждения, что он решил наконец проклятую труд-

ную проблему сознания. Первые главы посвящены описанию всевозможных сенсорных иллюзий, затем автор переходит к обсуждению гипотетического научного эксперимента, в котором беспристрастные ученые слушают рассказ испытуемого о состоянии его сознания. Деннетт задается вопросом: почему ученые должны верить этим описаниям? Можно представить, что некоторые испытуемые – это роботы или зомби, у которых нет никакого сознания, но они могут имитировать поведение человека и приводить какие угодно (вымышленные) описания своих ментальных состояний. У ученых нет никаких способов различить описания настоящих людей от описаний зомби.

Далее в качестве второго примера автор приводит этнографов, изучающих дикое племя, представители которого верят в некоего местного бога, автор называет его Финоман (искаженное «феномен»). Этнографы просят описать этого бога и получают достаточно противоречивый и маловразумительный образ, которому они, конечно же, не верят (прежде всего потому, вероятно, что изначально уверены в вымышленности персонажа). Все эти подводные маневры позволяют Деннетту заключить: поскольку нет объективных данных о феноменологии сознания, то все, что мы воспринимаем как собственное сознание, можно считать лишь иллюзией: «...люди, безусловно, верят, что у них есть ментальные образы, ощущение боли, сенсорные образы и прочее и организуют их в фиктивные теории (intentional objects) в своем “гетерофеноменологическом мире”. Вопрос же о том, существуют ли эти образы как реальные объекты, события и состояния мозга – или души, если вам угодно, – это вопрос эмпирического исследования» (подчеркивание мое – тут любопытно отождествление души с мозгом) [Ibid, p. 25]. Если подходящие явления будут обнаружены, это будет решением давней проблемы сознания, если нет, нужно будет объяснить, почему людям кажется, что эти образы существуют.

По Деннетту, единственный способ доказать реальность явления – зарегистрировать его объективными методами. Если у ученых нет таких методов в отношении объектов «гетерофеноменологического мира», значит, эти объекты не существуют. Для каждого индивидуального человека, однако, в целом неважно, могут ли ученые найти для элементов его внутреннего мира аналоги в активности мозга. Вряд ли многие (включая, возможно, и самого Деннетта) готовы согласиться, что их сознание не существует только по причине невозможности его объективизировать уче-

ными. Джон Серл отмечает: утверждая, что субъективное сознание не существует, потому что не может быть зарегистрировано объективными методами, Деннетт делает категориальную ошибку, уравнивая эпистемологическую и онтологическую объективность. Реальность субъективного сознания может подтвердить любой человек, так как обладает этим сознанием, и в данном вопросе он является единственным надежным источником информации [Searle, 1997].

Кроме того, приводимые здесь Деннеттом примеры весьма специфичны. Ученые просят испытуемого описать состояние своего сознания. Эта задача, на мой взгляд, непосильна для большинства людей, так как требует наличия двух редких качеств – навыков интроспекции и способности вербализовать невербальные ощущения. То, что многие не могут описать состояние собственного сознания, или их описания кажутся маловразумительными, не означает отсутствие этих состояний в их субъективном опыте. Подавляющая часть состояний субъективного опыта не воспринимается сознательно просто потому, что человек в обычной жизни не ставит себе такой задачи. Осознание, а тем более описание этих состояний другим людям – отдельная нелегкая работа.

Если же говорить не о неуловимых, трудно вербализуемых элементах сознания, таких как квалиа, а об относительно легко вербализуемых элементах сознания, таких как мысли и рассуждения, то сомнение в их реальности представляется абсурдом, поскольку ставит под вопрос существование всей человеческой цивилизации. Хотя испытуемый в своем словесном отчете может сознательно или неосознанно исказить или скрывать некоторые факты, эти отчеты стали нормальным инструментом исследований в большинстве гуманитарных и многих естественных науках, занимающихся изучением поведения человека, не говоря уже о всех других сферах социальной реальности, несмотря на то, что никому пока не удалось найти соответствия между активностью мозга и содержанием мыслей.

Далее Деннетт переходит к изложению своей «эмпирической теории сознания», называя ее теорией множественных набросков (*multiple drafts theory*). С ее помощью он хочет объяснить, каким событиям в мозге соответствует «поток сознания» (термин, впервые предложенный Джеймсом [Natsoulas, 1983]). Стратегия автора на протяжении всей книги похожа на стратегию осторожного бойца, который не решается атаковать противника «в лоб», а кружит

вокруг него, делая периодически мелкие вылазки в надежде в конечном счете выиграть по очкам. Главный противник в данном случае – гипотетическая фигура, совмещающая в себе зрителя и режиссера в так называемом картезианском театре (термин, придуманный Деннеттом). Фигура эта пришла из философии Декарта и олицетворяет собой нематериальную душу, которая получает от материального мозга всю сенсорную информацию, принимает решения и отдает мозгу приказы действовать.

Деннетт многократно и справедливо отмечает, что избавиться от осознанного или неосознанного представления о наличии в глубине сознания такой фигуры очень трудно. Это видно и из его собственного текста, когда при описании процесса принятия решений он использует выражение «мозг решает». Поскольку механизмы возникновения решения в мозгу остаются неясными (или в лучшем случае прописаны схематично), то замена слова «душа» словом «мозг» ничего по сути не меняет. Аналогично этому при (спекулятивном по собственному признанию автора) описании эволюции человека Деннетт постоянно использует выражение «Мать Природа (с заглавной буквы в оригинале. – Г.К.) одарила нас» (определенными качествами).

Как мы знаем, согласно материалистическому определению, Мать Природа (в отличие от целеполагающего Агента) слепа. Чтобы объяснить, каким образом она могла нас «одарить» тем, что мы имеем, Дарвин предложил свою теорию эволюции, в которой конкуренция и выживание наиболее приспособленных определяют по сути своей хаотический (нецеленаправленный) процесс возникновения порядка из хаоса. Идея оказалась настолько привлекательной, что Деннетт использует ее и для объяснения целенаправленной (по видимости) активности человеческого сознания.

Описывая, например, гипотетический процесс генерации речи, Деннетт справедливо отмечает, что, в отличие от большинства современных компьютеров, устроенных по фон-неймановскому принципу последовательных операций, мозг использует параллельный процессинг. По описанию Деннетта, процесс генерации речи должен проходить через ряд стадий, на каждой из которых имеет место свободная конкуренция множества вариантов. Таким образом, результат в виде произнесенной фразы возникает в процессе естественного отбора на каждом из этапов – спонтанно и как бы из ничего. Такая же схема приложима ко всем остальным процессам. В мозге активно постоянно и одновременно большое

количество функциональных модулей, большинство из которых специализированы для выполнения определенной функции, но могут иногда включаться и в другие задачи, а некоторые не имеют определенной функции (*general purpose device*).

Модули конкурируют друг с другом (за что?). Некоторые теоретики, развивающие сходные идеи, думают, что за доступ к сознанию (например [Block, 2018]), но Деннетт отрицает это, так как отрицает существование сознания как некоей конечной стадии данного процесса. У Деннетта они конкурируют за доступ к каким-то исполнительным механизмам. Вся эта активность создает некий фон, из которого (благодаря конкуренции между модулями и победы наиболее релевантных) в конечном счете появляется необходимое решение (можно думать, в виде какого-либо поведенческого акта). Очевидно, что для выявления наиболее релевантных модулей должен существовать некий компаратор и какой-то критерий, по которому отбираются модули, но об этом Деннетт говорит мало.

Теория множественных набросков имеет сходство с одной из наиболее популярных в настоящее время нейрофизиологических теорий сознания – теорией глобального рабочего пространства [Baars, 1988, 1997, 2002] (она будет рассмотрена ниже). Для работоспособности описываемых Деннеттом механизмов необходимо наличие некоего организующего начала. По Деннетту, этим организующим началом являются частично врожденные наклонности, унаследованные от предков и отражающие отобранные эволюцией биологические механизмы, частично тенденции, усвоенные в процессе воспитания как результат усвоения культурных ценностей (они и сами являются результатом эволюции культуры). Таким образом, устраняется персона, личность, или «Я», как организующее начало сознания и мыслей человека, и результат в виде поведения определяется свободным взаимодействием и конкуренцией факторов культуры и биологических мотиваций.

На самом деле полностью устранить это ненавистное привидение (картезианский театр) не может даже сам Деннетт. Он называет постулируемый им конгломерат конкурирующих модулей «виртуальной машиной». «Каждый, кто обладает такой виртуальной машиной в качестве контрольной системы, обладает сознанием в полном смысле этого слова, и он обладает сознанием, потому что у него есть эта виртуальная машина» (выделено в оригинале [Dennett, 1991, p. 234]). Таким образом, виртуальная машина Деннетта выполняет функции, приписываемые сознанию. Деннетт

продолжает использовать слово «сознание», но для него оно относится не к феноменам субъективного мира человека, а к феноменам в мозге, которые потенциально могут быть объективизированы (хотя, по признанию Деннетта, пока что эти феномены не имеют реального эмпирического наполнения). Итак, теория Деннетта по сути отрицает существование того, что она пытается объяснить [Chalmers, 1996; Searle, 1997]. Отрицается реальность не только сознания, но и «Я», которое, по Деннетту, является центром тяжести конструируемого для удобства (кем конструируемого и для чьего удобства?) иллюзорного нарратива.

3.2. Сознание как эпифеномен активности мозга

Главная тема этой книги – анализ доводов «за» и «против» представления об эпифеноменологической природе психических процессов. Выше уже сделан краткий обзор основных философских концепций о природе ментального, таких как элиминативизм, редукционизм и эмерджентизм. В рамках материалистической философии эти концепции рассматривают все психические процессы как эпифеномен активности мозга. Перед тем, как перейти к обсуждению вопросов о теориях сознания и эмпирических фактах, приведем аргументы эпифеномалистов конкретно в отношении сознания. Это позволит проверить соответствие данных фактов представлениям эпифеномалистов.

Эпифеномен по определению не имеет каузальной силы. То есть все аспекты жизни человека и его поведения не зависят от наличия сознания. Сознание – это что-то вроде шума, который сопровождает работу двигателя, но никак на нее не влияет. Нам только кажется, что мы сознательно принимаем решения и воплощаем их в жизнь. На самом деле решения принимаются в мозге независимо от сознания и наше ощущение свободы воли – это лишь иллюзия [Свааб, 2019; Robinson, 2019]. Возникает вопрос: почему вообще существует сознание? Если человек – продукт эволюции, которая отбирала виды по признаку приспособленности, то легко представить, что эффективно работающий мозг мог бы появиться в процессе такой эволюции без всякого сознания (хотя, как мы видели выше, такая гипотеза не очень-то соответствует палеонтологическим данным, показывающим, что «эффективный» мозг появился раньше, чем поведенческие проявления его эффективности).

Другой, связанный с первым, вопрос: почему содержание сознания, в частности, эмоциональная оценка факторов внешней среды, конгруэнтно значению этих факторов для выживания [James, 1879; Bradley, 2011; Kent, 2018]? Если сознание не имеет каузальной силы, его содержание могло бы вообще не коррелировать ни с поведением человека, ни с обстоятельствами внешней среды. Человек мог бы, например, спастись от тигра и размышлять при этом на отвлеченные темы. Весь наш опыт показывает, однако, что содержание сознания более всего связано с поведением именно в сложных, неоднозначных и потенциально опасных ситуациях. Хорошо известно, что автоматические, заученные действия человек может выполнять, не включая сознания. Но оно активируется в ситуациях неопределенности, новизны и опасности. Если сознание – это эпифеномен работы мозга, почему оно сопровождает одни (потенциально наиболее важные для выживания) его состояния и не сопровождает другие?

Если один человек выполняет команды другого человека, он может произвести два противоположных действия (например, сесть на стул или встать со стула) в зависимости от услышанной команды. Тот или иной поведенческий эффект в этом случае вызывает именно сознательная расшифровка смысла услышанной команды, а не физические свойства звука. Если придерживаться позиции эпифеноменализма, то нужно допустить, что смысл команды должен однозначно кодироваться в активности нейронов. Подчеркну, что кодировка смысла должна быть независимой от процессов первичного восприятия сенсорной информации, так как команды могут подаваться голосом, с помощью сообщения на экране монитора или даже с помощью жеста – их смысл и поведенческий эффект останется тем же. Более того, разные люди и один и тот же человек в разных состояниях способны извлечь один и тот же смысл из этих команд. Это означает, что смысл должен каким-то образом распознаваться в мозге и абстрагироваться от всех сопутствующих характеристик стимула. Эпифеноменологическая редукционистская интерпретация сознания имплицитно предполагает, что это так и есть. Где-то в мозге, на каком-то этапе многоуровневого анализа стимула извлекается его смысл, который «закодирован» в каких-то параметрах активности нейронов и однозначно соответствует тому, как этот стимул представлен в сознании.

Можно рассмотреть приведенный выше пример выполнения звуковой команды, заменив человека компьютером. Современные компьютеры умеют (в некоторых пределах) распознавать человеческую речь и выполнять команды, не обладая сознанием. Казалось бы, это доказывает, что сознание не является непременным условием осмысленного поведения и вполне может быть просто эпифеноменом. Однако поведение компьютера задано алгоритмами, заложенными человеком. При выполнении какого-то действия в ответ на слово команды компьютеру не нужно извлекать смысл этого слова лишь потому, что программист, знающий смысл команды и ответного действия, заложил соответствующую цепочку событий в программу. Если заменить слово команды синонимом, не прописанным в программе, компьютер, в отличие от человека, не сможет ее выполнить. Как отмечал Серл, компьютер «понимает» синтаксис, но не семантику [Searle, 1983]. Кэмпбелл сравнивает «психологические переменные» (к ним относится содержание сознания) с панелью управления [Campbell, 2010]. Так же как манипуляции элементами управления автомобиля имеют неоспоримое влияние на его «поведение», так и компоненты семантики сознания однозначно влияют на поведение человека. Таким образом, если содержание сознания является необходимым условием осмысленного поведения, оно (согласно эпифеноменологической редукционистской интерпретации) должно быть однозначно закодировано в активности нейронов.

В последующем изложении обратимся к нейробиологическим теориям сознания и имеющимся эмпирическим данным.

3.3. Теории сознания

Сформулировано большое количество разнообразных теорий сознания. Все их можно условно разделить на метафизические, которые постулируют онтологический статус сознания по отношению к физическому миру, и специфические, пытающиеся дать более детальное описание природы, свойств и роли сознания [van Gulick, 2014]. Метафизические теории можно разделить на три категории, в зависимости от того, как они решают проблему соотношения сознания и материи. Дуалистические теории признают, что сознание или по крайней мере некоторые его аспекты выпадают из физического мира. Физикалистские теории объясняют феноменологию сознания исключительно действием физических

факторов, а именно – активностью мозга. Наконец, теории идеалистического монизма рассматривают сознание как первичную сущность по отношению к физическому миру.

Дуалистические теории различаются по тому, какие аспекты сознания считаются выпадающими из физического мира. Классический картезианский дуализм называют субстантивным дуализмом, так как он признает существование двух субстанций – материи и сознания [Descartes, 1644]. Хотя эта форма дуализма сейчас не в моде, у нее до сих пор есть последователи [Swinburne, 1986; Foster, 1989, 1996]. Большой популярностью пользуется в настоящее время другая форма дуализма, называемая дуализмом свойств [van Gulick, 2014]. Все теории этой категории постулируют существование определенных свойств сознания, которые не идентичны и не сводимы к физическим свойствам, хотя и могут быть проявлением лежащих в их основе процессов. Различают фундаментальный дуализм свойств, который рассматривает свойства сознания в качестве фундаментального компонента реальности наравне с физическими свойствами. По этой теории свойства сознания могут взаимодействовать с физическими свойствами, но онтологически их существование не зависит и не выводится из физических свойств [Chalmers, 1996]. Эмерджентный дуализм свойств рассматривает сознание как эмерджентную сущность, возникающую из сложной организации физических компонентов, но не сводимую к ним [Hasker, 1999]. Нейтральный монизм утверждает, что как ментальные, так и физические свойства выводятся из более базового уровня реальности, который не является ни ментальным, ни физическим [Russell, 1927; Strawson, 1994a]. Сторонники панпсихизма считают, что все компоненты реальности имеют ментальные или протоментальные свойства, отличные от их физических свойств [Nagel, 1979a]. Одна из версий панпсихизма предложена Тонони в его теории интегрированной информации (ИИТ) [Tononi, 2008; Koch, 2012], суть которой будет раскрыта ниже.

Большинство недуалистических теорий являются разными версиями физикализма. Элиминативизм отрицает реальное существование сознания вообще или некоторых его аспектов [Churchland, 1983; Dennett, 1991]. Теория идентичности отождествляет свойства, состояния и процессы сознания с процессами в мозге. «Если качественное сознательное переживание феноменального красного цвета означает просто наличие состояния мозга с соот-

ветствующими нейрофизиологическими свойствами, тогда такие экспериментальные свойства реальны, но их реальность – это прямая физическая реальность» [van Gulick, 2014, p. 1]. Нередукционный физикализм признает, что теоретических и концептуальных ресурсов описания процессов в мозге недостаточно для описания феноменологии сознания [Boyd, 1980], однако, как отмечают некоторые, он не может предложить адекватной альтернативы [Jackson, 2004].

Специфические теории сознания можно сгруппировать в семь категорий: теории высшего порядка, репрезентативные теории, интерпретационные нарративные теории, когнитивные теории, нейронные теории, квантовые теории и нефизические теории [van Gulick, 2014]. Главная идея теорий высшего порядка состоит в том, что ментальное состояние является сознательным, если оно сопровождается метаментальной реализацией того, что субъект находится в этом состоянии, в противном случае оно будет неосознанным [Rosenthal, 1993]. Теории высшего порядка имеют свое наполнение в области нейрофизиологии, где метаментальная активность связывается с активностью префронтальной коры [Lau, Rosenthal, 2011]. Однако существующие данные показывают сохранность сознания при разрушении областей префронтальной коры [Boly et al., 2017], хотя сторонники этой теории могут считать, что в подобных случаях существенная часть префронтальной коры осталась интактной [Odegaard et al., 2017].

Репрезентативные теории утверждают, что единственными свойствами сознания являются репрезентативные свойства, т. е. репрезентация объектов внешнего и внутреннего мира [Tye, 2000]. Согласно интерпретационным нарративным теориям, содержание сознания не всегда определено и имеет интерпретационный характер. Наиболее известный вариант таких теорий – рассмотренная выше теория множественных набросков Деннетта [Dennett, 1991]. Хотя эта теория имела немалый резонанс, она подверглась и серьезной критике [Block, 1994; Dretske, 1994; Levine, 1994]. Нейрофизиолог Газзанига предложил идею «интерпретирующего модуля», расположенного в левом полушарии, который придает смысл нашим действиям путем построения нарратива [Gazzaniga, 2011].

Когнитивные теории отождествляют состояния сознания с определенной когнитивной архитектурой, или паттерном активности. Наиболее разработанная и популярная теория из этой катего-

рии – теория глобального рабочего пространства (ТГРП, *global workspace theory*), которая описывает сознание как результат соревнования между процессорами за ограниченный ресурс, делающий информацию широко доступной [Baars, 1988]. Ограниченный ресурс понимается в терминах фокуса внимания и рабочей памяти. Широкая доступность информации в глобальном рабочем пространстве делает ее сознательной, доступной для рапорта и гибкого контроля поведения. Дальнейшее развитие модели привязало ее к конкретным структурам мозга, включающим как первичные сенсорные зоны коры, так и лобные и затылочные области, связанные с контролем внимания [Dehaene, Naccache, 2000]. Предполагается, например, что сознательное восприятие сенсорной информации происходит лишь тогда, когда активируется вся глобальная сеть – активности лишь в сенсорных зонах недостаточно, какой бы большой она ни была. Эта теория, по сути, постулирует условия доступности информации для сознания, но ничего не говорит о том, где и как кодируется содержание информации [Block, 2007]. Кроме того, сравнение активности мозга при осознанном и неосознанном восприятии стимула показывает, что активации, локализованной в лобной дорсолатеральной коре, может быть достаточно для сознательного восприятия [Lau, Passingham, 2006].

В отличие от ТГРП, теория рекуррентного процессинга постулирует, что для сознательного восприятия сенсорного стимула достаточно рекуррентной активации в соответствующей сенсорной коре, хотя это восприятие может быть и недоступно для интроспекции [Lamme, 2006, 2010]. Теория рекуррентного процессинга, так же как и ТГРП, ничего не говорит о том, где и как кодируется содержание информации. Согласно теории репрезентации промежуточного уровня (*attended intermediate representation*) сознательное восприятие является репрезентацией промежуточных свойств объекта, таких как цвет, форма, тон. Свойства высшего уровня выводятся путем суждений и не являются предметом сознательного восприятия [Prinz, 2012]. Сознательное восприятие промежуточных свойств объекта возможно, лишь если человек обращает на них внимание. Для каждого компонента сознательного восприятия Принц предлагает нервный субстрат. Он отождествляет репрезентацию промежуточного уровня с гамма активностью (40–80 Гц) в сенсорной коре, а внимание – с синхронизированными осцилляциями, включающими в себя вектор гамма активности [Ibid].

Одной из наиболее популярных в последнее время теорий сознания является ИИТ, предложенная Тонони [Tononi, 2008]. Интеграция информации, приходящей из разных источников, считается важным свойством сознания, которое фигурирует во многих теориях, особенно в теории глобального рабочего пространства. Тонони, однако, идет дальше этого и отождествляет сознание с интегрированной информацией. По его теории, интеграция информации – это необходимый и достаточный атрибут сознания, независимо от субстрата (не обязательно биологического), в котором эта интеграция осуществляется. То есть сознание – это информационно-теоретическое свойство любой системы. Тонони предлагает количественную меру (Фи) для измерения интегрированной информации. С помощью Фи можно оценить уровень интегрированной информации как в системе целиком, так и в разных ее компонентах. Если Фи системы больше, чем у любого из ее компонентов, значит, она обладает сознанием. Даже простые системы (например, фотодиод) обладают некоторым сознанием, если они не включены в качестве компонента в более сложные системы. Таким образом, ИИТ подразумевает некую форму панпсихизма, что Тонони открыто признает. Согласно ИИТ, уровень Фи зависит от наличия в системе петель обратной связи. Поэтому даже, казалось бы, сложные системы, такие как компьютер или мозг, не обладают сознанием, потому что у них нет петель обратной связи.

Нейронные теории пытаются определить нервный субстрат сознания. Если эти теории развиваются в рамках физикализма, то демонстрации наличия корреляций между нервной активностью и содержанием сознания недостаточно, так как подобные корреляции можно объяснить и в рамках дуалистических представлений. Физикалистские нейронные теории должны объяснить природу и механизм этих корреляций и показать, что постулируемый нервный субстрат идентичен, или, по крайней мере, понятным образом соответствует феноменологии сознания [Metzinger, 2000].

Существующие нейронные теории сознания настолько многочисленны, что не могут быть описаны сколько-нибудь подробно. Они включают: модели, которые обращаются к глобальным интегрированным полям [Kinsbourne, 1988; Pockett, 2000; McFadden, 2020]; интеграции информации посредством синхронных колебаний [Crick, Koch; 1990; Singer, 1999]; NMDA – опосредованные

переходные нейронные ассамблеи [Flohr, 1995]; таламически модулированные паттерны корковой активации [Llinas, 2001]; реентерабельные корковые петли [Edelman, 1989]; механизмы компаратора, которые участвуют в петлях прогнозирования действия и оценки результата между лобной и срединной областями мозга [Gray, 1995]; интерпретационные процессы в левом полушарии [Gazzaniga, 1988]; эмоциональные соматосенсорные гомеостатические процессы базирующиеся в лобной и лимбической коре [Damasio, 1999] или в периакведуктальном сером веществе [Panksepp, 1998].

Одна из недавних дискуссий касается вопроса о том, достаточно ли локальной активности в сенсорной коре для возникновения феноменологического сознания, или необходима глобальная активность в разных зонах коры. Сторонники теории глобального рабочего пространства считают, что сознательное восприятие возникает лишь при широкомасштабной активации лобных и затылочных областей наряду с первичной сенсорной корой [Dehaene, Naccache, 2000]. Другие же доказывают, что достаточно повторяющейся активности в первичной сенсорной коре [Lamme, 2006; Block, 2007]. В каждом случае цель состоит в том, чтобы объяснить, как организация активности на соответствующем нейронном уровне может лежать в основе того или иного типа или особенности сознания. Каждая из этих теорий может вполне реалистично описывать процессы в мозге, имеющие отношение к разным сторонам сознательного восприятия реальности, однако ни одна из них не приближается к тому, чтобы объяснить природу и механизм связи между активностью мозга и содержанием сознания.

Квантовые теории пытаются связать феномен сознания с фундаментальными свойствами материи на квантовом уровне. В этой сфере можно выделить два принципиально различных направления. Одно из них рассматривает мозг как квантовую систему и пытается объяснить феномен сознания процессами, происходящими в мозге на квантовом уровне. Другое направление не рассматривает мозг вообще, а применяет абстрактные математические принципы квантовой теории к исследованиям в когнитивной науке.

В рамках первого направления наиболее традиционной, в смысле соответствия классической копенгагенской интерпретации квантовой механики, является теория, предложенная Стэппом [Stapp, 2009]. Она развивает идеи Гейзенберга, фон Неймана и Вигнера о роли человеческого сознания в результатах наблюдений

над физическими квантовыми системами (так называемый эффект наблюдателя). Эти идеи, в неявной форме высказанные Гейзенбергом, были развиты фон Нейманом, который рассматривал процесс физического эксперимента как результат взаимодействия измеряемого объекта (I), аппарата, с помощью которого производится измерение (II), и человека, производящего измерение (III) [von Neumann, 1932]. Он заключает, что для результата измерений не имеет значения, находится ли граница между наблюдаемой и наблюдающей системой между I и (II + III) или между (I + II) и III.

Лондон и Бауэр пошли дальше и предположили, что именно человеческое сознание завершает процесс квантового измерения [London, Bauer, 1939]. Они приписали сознанию решающую роль в понимании результатов квантового измерения с точки зрения обновления знаний наблюдателя. В 1960-х гг. Вигнер [Wigner, 1967] выдвинул предложение, согласно которому сознание влияет на физическое состояние измеряемой системы, а не только на знание наблюдателя. С 1980-х гг. Стэпп развивал свою теорию на базе теорий фон Неймана и Вигнера. Пользуясь постулированной фон Нейманом свободой в размещении интерфейса между наблюдаемой и наблюдающей системами, Стэпп помещает его в мозг наблюдателя [Stapp, 2009]. Он рассматривает предложенное Гейзенбергом разграничение потенциального и актуального, где актуальное связано с измеряемым событием в смысле копенгагенской интерпретации, а потенциальное относится к ситуации до измерения, выражая, таким образом, идею реальности, независимой от измерения [Heisenberg, 1958]. Основной шаг Стэппа состоит в том, чтобы приписать каждому актуальному событию чувственный аспект, который он называет ощущением этого события и который придает ему статус внутренней актуальности.

Таким образом, ментальный и физический полюсы так называемых реальных событий рассматриваются как психологические и физические аспекты реальности. Потенциальные предшественники реальности психофизически нейтральны и относятся к форме существования, при которой разум и материя не разделены. Это выражается, например, в предложенной Стэппом «гибридной онтологии», обладающей как ментальными, так и материальными качествами [Stapp, 2009]. Каждое сознательное ощущение, согласно Стэппу, имеет в качестве физического аналога редукцию квантового состояния, актуализирующую паттерн активности, который иногда называют нейронным коррелятом этого сознательного

ощущения. Предполагается, что возникновение психических состояний соответствует редукции состояний квантовой суперпозиции нейронных ансамблей (в квантовой механике это иногда называют коллапсом волновой функции).

Вероятно, наиболее конкретное предположение о том, как квантовая механика в ее современном виде может играть роль в мозговых процессах, принадлежит Беку и Экклсу [Beck, Eccles, 1992]. Оно относится к конкретным механизмам передачи информации в синаптической щели. Однако какое отношение эти квантовые процессы могут иметь к умственной деятельности и как они могут взаимодействовать с ментальными состояниями, остается неясным до настоящего времени.

Роджер Пенроуз предложил модель, согласно которой сознание возникает вследствие «объективного коллапса» волновой функции, что объясняет «неалгоритмический» характер мышления человека [Penrose, 1989]. Физическая часть сценария, предложенного Пенроузом, относится к будущим разработкам квантовой теории, необходимым для понимания физического процесса, лежащего в основе редукции квантового состояния. Предполагается, что для окончательного понимания квантовых измерений требуется полноценная теория квантовой гравитации. Обоснование необходимости привлечения редукции квантового состояния заключается в том, что элементарные сознательные процессы не могут быть описаны алгоритмически, и, следовательно, не могут быть вычислены. Он ссылается на природу творчества и математического инсайта, на теоремы Геделя, которые доказывают ограниченность любой аксиоматической системы, и на идею платонической реальности за пределами разума и материи. По представлениям Пенроуза, существует три мира – физический, ментальный и платонический мир математических идей. Стюарт Хамерофф, развивая идеи Пенроуза, выдвинул гипотезу, согласно которой «объективный коллапс» происходит в микротрубочках нейронов [Hameroff, Watt, 1982].

В 1960-х гг. Риккарди и Умедзава предложили использовать формализм квантовой теории поля для описания состояний мозга, уделяя особое внимание памяти [Ricciardi, Umezawa, 1967]. Основная идея состоит в том, чтобы моделировать состояния памяти в терминах состояний многокомпонентных систем, как неэквивалентных реализаций вакуумных состояний квантовых полей.

В последнее время в этом направлении был достигнут значительный прогресс за счет включения эффектов диссипации, хаоса, фракталов и квантового шума [Vitiello, 2012].

Сегодня накапливаются свидетельства того, что квантовые концепции, такие как дополнительность, запутанность, дисперсионные состояния, некоммутативные операции и небулева логика, можно применять для описания ментальных процессов. Соответствующие подходы обращаются к чисто ментальным (психологическим) явлениям с использованием формальных функций из квантовой физики, но без привлечения физической основы квантовой механики или квантовой теории поля. Для обозначения этой области исследований был придуман термин «квантовое познание» (*quantum cognition*), в отличие от рассмотренных выше теорий, которые относятся к категории «квантовый мозг» [Atmanspacher, 2020]. Эта сфера исследований еще очень молода и пока неизвестно, какой вклад она в конечном счете внесет в понимание природы сознания. С точки зрения обсуждаемой нами темы эпифеноменологической природы сознания эти данные интересны тем, что демонстрируют концептуально разную природу ментальных и макрофизических (на уровне мозга) процессов. Тем не менее читатель, не интересующийся этой достаточно специальной темой, может легко ее пропустить. Следующий за этим раздел посвящен описанию вычислительной теории разума (ВТР), возникшей на границе между когнитивными и компьютерными науками. Тема эта тоже достаточно специфична и, опять же, ее можно пропустить без ущерба для понимания последующего изложения эмпирических данных о мозговых коррелятах сознания и других ментальных процессов.

3.3.1. «Квантовое познание»

Моделирование ментальных процессов на основе формализма квантовой теории позволяет предсказать некоторые из них лучше, чем классическая теория вероятности, включая байесовскую модель. Интуитивно нетрудно понять, почему некоммутативные операции или небулева логика должны быть актуальны, даже неизбежны для ментальных систем, не имеющих ничего общего с квантовой физикой. Некоммутативность операций означает лишь, что последовательность, в которой применяются операции, имеет значение для конечного результата. А небулева логика относится к предложениям, которые могут иметь нечеткие критерии истин-

ности, не сводимые к простому «да» или «нет», оттенки правдоподобия или неправдоподобия. Отличительной особенностью этих подходов является то, что они привели к четко определенным и конкретным теоретическим моделям с новыми эмпирическими предсказаниями. Ванг с соавторами выделяют пять областей исследования когнитивных процессов, в которых предсказания на основе квантовой теории работают лучше, чем предсказания на основе классической теории вероятности [Wang et al., 2013].

1. Проблема формализации психологических концепций конфликта, двусмысленности и неуверенности. Традиционные когнитивные модели предполагают, что в каждый момент времени человек находится в определенном состоянии по отношению к некоторому суждению; однако истинное состояние человека неизвестно, и, следовательно, модель может лишь присвоить некоторую вероятность состояниям в каждый момент времени. Этот тип моделей является стохастическим только потому, что исследователь не знает, по какой именно траектории следует человек, но предполагается, что его состояние следует четко определенной траектории в своем пространстве состояний. Напротив, квантовый подход позволяет человеку в каждый момент времени находиться в неопределенном (технически дисперсионном) состоянии, называемом состоянием суперпозиции, которое отражает состояние конфликта, двусмысленности или неуверенности, до тех пор, пока не будет принято решение, после чего система коллапсирует в состояние определенности.

2. Проблема формализации чувствительности когнитивной системы к измерениям. Традиционные когнитивные модели предполагают, что то, что мы измеряем в определенный момент времени, отражает состояние когнитивной системы непосредственно перед измерением. Один из уроков, извлеченных из квантовой теории, состоит в том, что измерение системы не только регистрирует ее свойство, но и формирует его. Непосредственно перед измерением квантовая система находится в состоянии суперпозиции. Ответ, который мы получаем, является результатом взаимодействия состояния суперпозиции и задаваемого вопроса [Bohr, 1958]. Это взаимодействие создает четко определенное состояние из неопределенного состояния суперпозиции. Квантовый принцип построения реальности из взаимодействия между неопределенным состоянием человека и задаваемым ему вопросом согласуется с конструктивистской позицией в психологических исследованиях

и идеей о том, что совершенный выбор может изменить предположения [Sharot et al., 2010]. Он лучше соответствует психологической интуиции в отношении сложных суждений и решений, чем предположение, что извлеченный ответ – это просто регистрация ранее существовавшего состояния.

3. Проблема формализации эффектов последовательности когнитивных измерений. Изменение когнитивного состояния, возникающее в результате ответа на вопрос, может вести к тому, что человек иначе отвечает на последующие вопросы [Moore, 2002]. Другими словами, первое измерение изменяет контекст следующего измерения и порядок измерений становится важным. Это означает, что мы не можем определить совместную вероятность ответов на совокупность вопросов. Мы можем присвоить вероятность только последовательности ответов. В квантовой физике измерения, зависящие от порядка, называются некоммутативными операциями, и многие математические свойства квантовой теории возникли из разработки вероятностной модели для некоммутативных измерений, включая принцип неопределенности Гейзенберга [Heisenberg, 1958]. Эффекты последовательности измерений, такие как эффекты порядка вопросов, являются серьезной проблемой в исследовании суждений и принятия решений, нуждающейся в теоретическом обосновании подобном тому, которое достигнуто в квантовой теории.

4. Проблема понимания причины отклонений от классических вероятностных законов в сфере когнитивных исследований. Человеческие суждения не всегда подчиняются классическим законам логики и вероятности. Классические вероятностные модели, используемые в современных моделях когнитивных процессов и принятия решений, выводятся из аксиом Колмогорова, которые назначают вероятности событиям, определенным как множества. Семейство множеств в теории Колмогорова подчиняется аксиомам булевой логики. Важное место среди них занимает распределительная аксиома, из которой выводится закон общей вероятности, лежащий в основе байесовских сетей. Однако закон общей вероятности нарушается во многих психологических экспериментах [Tversky, Shafir, 1992]. Квантовая теория вероятностей выводится из аксиом фон Неймана [von Neumann, 1932], которые приписывают вероятности событиям, определенным как подпространства в векторном пространстве Гилберта, не подчиняющиеся дистрибутивной аксиоме булевой логики [Hughes, 1989]. Это

значит, что квантовые модели не всегда подчиняются закону общей вероятности и лучше подходят для описания когнитивных процессов.

5. Проблема понимания неразложимости мышления. В когнитивной науке концепции и процессы считаются «разложимыми», так что целое можно понять с точки зрения составляющих его компонентов. Это отражается в предположении, что существует полное совместное распределение вероятностей по всем измерениям когнитивной системы, из которого можно восстановить попарные совместные распределения для любой пары измерений. Однако психологическая литература предоставляет обширные данные, указывающие на обратное. Например, узнавание лица опирается на его целостное восприятие, так же как восприятие двусмысленных изображений. Концептуальные комбинации тоже не поддаются разложению на осмысленные части. В отличие от классических, квантовые модели позволяют описывать когнитивные системы как неразложимые; при этом парные вероятности не могут быть получены из общего совместного распределения вероятностей. Это говорит о радикально новом по сравнению с классическим моделированием типе корреляций, известном в квантовой теории как корреляции запутанности.

Четыре группы ключевых концепций и принципов квантовой теории лучше всего иллюстрируют, как квантовые когнитивные модели отличаются от традиционных когнитивных моделей при изучении когнитивных процессов.

Во-первых, это дисперсионные состояния и квантовые вероятности. В монографии «Математические основы квантовой механики» фон Нейман представил математическую основу гильбертова пространства для разработки формализма статистического ансамбля в квантовой теории [von Neumann, 1932]. Важно отметить, что правила для квантовых вероятностей отличаются от правил классической вероятности. Не только состояния смешанного ансамбля, но и состояния чистой квантовой суперпозиции являются дисперсионными: они не могут быть представлены точно, как мы привыкли делать в классической механике. Правило Борна [Born, 1926], которое определяет вероятность обнаружения конкретного квантового состояния квадратом его амплитуды, вводит в уравнение интерференционные члены, отсутствующие в классической теории вероятностей. Следствием этого является неаддитивный характер квантовых вероятностей. Кроме того, поскольку

состояния изменяются в результате измерения, квантовые вероятности переходов между состояниями решающим образом зависят от предшествующих измерений.

Во-вторых, запутанность, нелокальные корреляции и неравенства Белла. Специфическая форма квантовой суперпозиции называется квантовой запутанностью. Запутанные состояния демонстрируют акаузальные нелокальные корреляции между результатами измерений на пространственно разнесенных подсистемах системы. Запутанные системы могут возникать в результате любых взаимодействий, и запутанность – это общая черта квантового мира. Поскольку эти корреляции повсеместны для запутанных систем, такие системы называются неразборными. Нелокальные корреляции являются основой того, что называют квантовым холизмом. Ключевой проблемой для понимания квантовых систем является не то, как можно составить их из элементарных компонентов, а, скорее, как можно разложить их на составные части. Таким образом, картина квантового мира принципиально целостна [Wang et al., 2013].

В-третьих, некоммутативные наблюдаемые и дополнительность. В то время как системы, изучаемые в классической физике, ограничены коммутативными наблюдаемыми, квантовые системы могут иметь как коммутативные (например, масса и заряд), так и некоммутативные наблюдаемые (например, положение в пространстве и моментум). Некоммутативность наблюдаемых A и B можно охарактеризовать как $AB \neq BA$. Это выражает эффект порядка, неизвестный для классических наблюдаемых, которые подчиняются закону коммутативности ($AB = BA$). Если наблюдаемые A и B не коммутируют, они несовместимы и не имеют полной ортонормальной системы общих собственных состояний, т. е. они могут разделять некоторые, но не все собственные состояния. A и B максимально несовместимы (или дополняют друг друга), если у них нет ни одного общего собственного состояния. Бор [Bohr, 1928] первоначально ввел понятие дополнительности для объяснения неоднозначности волноподобного и частицеподобного поведения квантовых систем. Интересно, что Бор фактически заимствовал эту концепцию из психологии, где она была предложена Уильямом Джеймсом [James, 1890]. В узком смысле дополнительность относится к несовместимости измерений, которые не коммутируют, т. е. их последовательность имеет значение для конечного результата. В более общем плане дополнительность от-

носится к описаниям, которые исключают друг друга, но вместе необходимы для исчерпывающего описания ситуации. В этом смысле дополнительность относится к несовместимым аспектам системы, которые нельзя объединить в одном булевом описании.

В-четвертых, небулева и частично булева логика и комплементарность. Одна из предложенных фон Нейманом формализаций квантовой теории опирается на неклассическую (небулеву) логику [Birkhoff, von Neumann, 1936]. В отличие от двузначной (да/нет) логики классических бинарных альтернатив, где дополнением утверждения является его отрицание, дополнения квантового утверждения обычно отклоняются от его отрицания. Уильям Джеймс [James, 1890] приводит такой пример: «Истинной противоположностью веры является сомнение, а не неверие». В булевом смысле дополнением и отрицанием веры является неверие, а в квантовом, небулевом смысле – сомнение, а не отрицание веры.

Успешное применение этих принципов квантовой теории для описания когнитивных процессов продемонстрировано при моделировании принятия решений, сенсорного восприятия, семантических сетей, оценки вероятностей событий, памяти и влияния последовательности измерений [Wang et al., 2013]. На первый взгляд, адекватность квантового формализма для описания когнитивных процессов кажется парадоксом, так как явления, изучаемые квантовой механикой и когнитивной психологией, лежат на противоположных полюсах спектра сложности – с одной стороны элементарные процессы на уровне атома, с другой – процессы, предположительно связанные с активностью наиболее сложной из известных систем. Может быть, это свидетельствует о правильности физикалистского представления о том, что в основе ментальности лежат физические процессы? В конечном счете квантовая механика – это лишь физическая теория.

Физикалистские редукционистские теории ментальности обычно описывают процессы в мозге в рамках классических представлений, а приведенные выше данные как раз показывают, что классические представления менее адекватны при описании когнитивных процессов, чем квантовый формализм. Может быть, это аргумент в пользу важности квантовых процессов в работе мозга, как это представлено в рассмотренных выше теориях Стэппа, Экклса, Пенроуза и Витиэлло? Однако, как мы видели, теории эти пока слишком спекулятивны и не могут объяснить, как квантовые

эффекты в активности мозга могут отражаться в ментальной сфере. Кроме того, адепты исследований в области «квантового познания» подчеркивают отличие их работы от исследования квантовых механизмов мозга [Ibid]. Одним из вариантов объяснения похожести закономерностей, описывающих процессы на ментальном и базовом физическом уровнях, могут быть философские теории нейтрального или двойственного монизма, согласно которым материя и ментальность – это два аспекта подлежащей фундаментальной реальности [Atmanspacher, 2014].

3.3.2. Вычислительная теория разума

Завершает обзор теорий сознания очень важное и популярное в настоящее время направление, которое обычно описывают общим термином «вычислительная теория разума» (ВТР, the computational theory of mind). Это направление возникло и развивалось под большим влиянием компьютерных наук и исследований в области искусственного интеллекта (ИИ). Следствием несомненных успехов в этой области явилось представление (подпитываемое большой традицией научно-фантастических романов) о том, что уже в ближайшем будущем будут созданы компьютеры, ничем не уступающие и даже существенно превосходящие возможности человеческого разума, и что разум вообще – это лишь способность производить вычисления, используя ограниченный набор алгоритмов. Эта идея казалась очень соблазнительной, в первую очередь потому, что вычисления, производимые компьютером с использованием алгоритмов, имеют хорошо разработанную математическую базу, которая могла бы позволить описать работу человеческого разума.

Прежде чем приступить к обсуждению основных постулатов ВТР и ее развития в последние десятилетия, нужно кратко остановиться на математической теории алгоритмических вычислений. Алгоритм (само слово происходит от имени арабского математика Мухаммада ибн Муса аль-Хорезми, жившего в IX в. н.э.) – это конечная последовательность однозначно интерпретируемых инструкций для решения класса математических проблем или выполнения компьютерных операций [Definition of ALGORITHM. Merriam-Webster Online Dictionary, <https://web.archive.org/web/20200214074446/https://www.merriam-webster.com/dictionary/algorithm>]. То есть вычисления с использованием алго-

ритмов не требуют какой-либо сообразительности или способности интуитивно выбрать какой-то вариант из множества возможных. Все, что нужно, – это знание соответствующих инструкций и их скрупулезное выполнение.

В 1928 г. немецкие математики Дэвид Гилберт и Вильгельм Акерманн сформулировали так называемую проблему принятия решений (нем. Entscheidungsproblem) [Hilbert, Ackermann, 1950] – можно ли, используя алгоритмы, решить любую математическую задачу? То есть можно ли для любого адекватно сформулированного математического утверждения решить, верно оно или нет, исходя из набора аксиом и используя правила математической логики? Сам Гилберт был уверен в положительном ответе на этот вопрос. Австрийский логик Курт Гедель, однако, опубликовал в 1931 г. две теоремы (Gödel's incompleteness theorems), которые доказывали ограниченность любой аксиоматической системы в ее способности моделировать даже базовые арифметические операции [Gödel, 1931]. В теоремах Геделя рассматриваются формальные системы, которые должны соответствовать требованиям полноты и последовательности.

Формальная система – это система конечного количества аксиом и правил логики (алгоритмов), позволяющих выводить новые теоремы. Такая система считается полной, если любое утверждение, сформулированное с использованием аксиом в соответствии с правилами, или отрицание этого утверждения могут быть выведены в рамках системы. Система считается последовательной (непротиворечивой), если нельзя одновременно вывести и утверждение, и его отрицание. Первая теорема Геделя утверждает, что любая последовательная система, в которой можно производить элементарные арифметические операции, не может быть полной, т. е. в ней есть утверждения, которые нельзя ни доказать, ни опровергнуть. Вторая теорема утверждает, что непротиворечивость любой системы не может быть доказана в рамках самой системы. Этот негативный ответ на сформулированную Гилбертом проблему принятия решений был в дальнейшем независимо подтвержден Алонзо Чёрчем [Church, 1936] и Аланом Тьюрингом [Turing, 1936]. Пенроуз считает, что человеческое мышление имеет в своей основе неалгоритмическую природу, и это позволяет находить решения в условиях неопределенности и для прежде неизвестных задач [Penrose, 1989].

Дальнейшее совершенствование теория алгоритмических решений получила в работах Алана Тьюринга, которые имели колоссальное значение для всего последующего развития компьютерных наук, ИИ и ВТР. Принципиально важную роль сыграла предложенная Тьюрингом концепция универсальной вычислительной машины. Машина Тьюринга (МТ), описанная в уже упомянутой выше статье [Turing, 1936], – это абстрактная модель идеализированного вычислительного устройства, имеющего в своем распоряжении неограниченные ресурсы времени и памяти. Машина манипулирует символами, природу которых Тьюринг не конкретизирует.

Предполагается, что есть ограниченный набор простых символов, которые машина может писать или стирать в ячейках памяти. Бесконечное количество ячеек памяти (это могут быть клеточки на бумажной ленте или силиконовые чипы) организовано в линейную последовательность. Центральный процессор в каждый момент времени имеет доступ лишь к одной ячейке, наподобие сканера, который проходит по бумажной ленте, последовательно сканируя каждую ячейку. Сам процессор может быть в одном из конечного набора состояний. Процессор может совершать одно из четырех действий: записывать символ в ячейку, стирать символ из ячейки, передвигаться к следующей ячейке («двигаться по ленте направо»), возвращаться к предыдущей ячейке («двигаться по ленте налево»). Какое из этих действий будет совершать процессор в каждый момент времени, зависит исключительно от текущего состояния машины и от символа, записанного в актуальной ячейке.

Определяющие работу машины правила записаны в таблицу, где указано, какие действия выполняет процессор в зависимости от текущего состояния (из конечного набора состояний) и символа (из конечного набора символов). В таблице также указано, как меняется состояние машины исходя из этих же факторов. Таким образом, в таблице дается ограниченный набор рутинных механических инструкций, управляющих вычислениями. Тьюринг представляет это неформальное описание в виде строгой математической модели и утверждает, что такая машина, несмотря на ее простоту, способна проделать любую выполняемую людьми механическую операцию над конфигурациями символов.

Далее Тьюринг доказывает, что может существовать универсальная МТ (УМТ), которая может воспроизводить поведение любой МТ, если в качестве входной информации она получает таблицу, описывающую соответствующее поведение. Таким образом,

УМТ является программируемым компьютером общего назначения. В некотором приближении все современные компьютеры являются УМТ, так как могут выполнять операции любой МТ при наличии соответствующей программы, с той оговоркой, что у них, в отличие от МТ, ограниченный ресурс памяти.

Прогресс компьютерных технологий подталкивал многих, включая Тьюринга, к мысли о том, что в недалеком будущем можно будет создать компьютер, способный мыслить. Однако первоначальный успех в этой области (например, знаменитая программа Logic Theorist, решившая 38 теорем из 52, описанных в Principia Mathematica) сменился периодом застоя, охладившего энтузиазм конструкторов ИИ. Стало понятно, что даже ментальные процессы относительно низкого уровня, такие как сенсорное восприятие, требуют операций, недоступных современным компьютерам. Одна из трудных проблем, с которой столкнулись исследователи в области ИИ, – решение задач в условиях неопределенности. В реальной человеческой жизни принятие практически любого решения происходит в условиях неопределенности.

Стандартная математическая модель для такого рода задач – байесовская теория принятия решений. Неопределенность в рамках этой модели кодируется в терминах вероятности. Могут быть прописаны алгоритмы, как изменять оценку вероятности при появлении новых данных и какие действия выбирать исходя из этой вероятности. В 1980–1990-х гг. технологический и концептуальный прогресс позволил создать эффективные компьютерные программы для использования байесовской модели в реалистичных сценариях, и это направление остается одним из наиболее перспективных в исследовании ИИ [Murphy, 2012]. Использование байесовской модели, а также нейронных сетей и машинного обучения, которые мы рассмотрим ниже, позволили достичь существенного прогресса в некоторых сферах. Так, созданная в IBM программа Deep Blue смогла победить в 1997 г. чемпиона мира по шахматам Гари Каспарова. Другие программы способны управлять автомобилем и распознавать человеческую речь. В свете этих успехов опять актуальным становится вопрос о возможности создания искусственного интеллекта, ничем не отличающегося от человеческого или даже превосходящего его.

Тьюринг в свое время представил свой вариант решения этой проблемы. Он рекомендовал заменить вопрос «может ли компьютер мыслить», который он считал безнадежно туманным, вопросом

«может ли компьютер успешно пройти специфический, предложенный им тест». Этот тест, получивший название теста Тьюринга, состоит в том, что эксперт задает вопросы двум невидимым собеседникам, один из которых – человек, а другой – компьютер. Результат теста для компьютера будет успешным, если эксперт не сможет различить, кто есть кто [Turing, 1950]. Философы, однако, критиковали предложение Тьюринга [Block, 1981; Searle, 1980]. В частности, Джон Серл выдвинул в качестве контраргумента модель «китайской комнаты» – человек, не знающий китайского языка, получает таблички с текстами на китайском и должен направлять респонденту ответы на этом же языке, следуя подробным инструкциям (написанным на его родном языке). Если инструкции достаточно хороши, у респондента может создаться впечатление, что он общается с человеком, владеющим китайским языком, хотя в реальности его собеседник не понимает ни слова по-китайски [Searle, 1980]. Этот аргумент имеет целью показать, что компьютерная имитация, какой бы искусной она ни была, всегда останется лишь имитацией.

Классическая ВТР (КВТР) берет свое начало в работе Уоррена Маккаллока и Вальтера Питтса, которые предположили, что что-то, напоминающее машину Тьюринга, может быть хорошей моделью для человеческого разума [McCulloch, Pitts, 1943]. В 1960-х гг. эта модель, ставшая центральной в когнитивной науке, исследовала разум методами психологии, ИИ, лингвистики, философии, экономики (особенно теории игр), антропологии и нейробиологии. Согласно КВТР, человеческий разум – это вычислительная машина, принципиально похожая на МТ, и главные ментальные процессы, такие как рассуждение и принятие решений, – это вычисления, похожие на вычисления в МТ. В отличие от господствовавших до 1960-х гг. теорий, таких как бихевиоризм, пытавшийся соотнести ментальные состояния с поведенческими паттернами, или теория идентичности, провозглашающая идентичность ментального и нейрофизиологического состояний, КВТР использует функциональный подход. Ментальные состояния трактуются в нем как состояния, соответствующие определенной функциональной организации, независимо от субстрата, на котором эта организация реализована (биологическая ткань, или, например, чипы компьютера) [Putnam, 1967].

Таким образом, Путнам предлагает версию функционализма, названную затем машинным функционализмом. По этой версии отдельные ментальные состояния – это состояния центрального про-

цессора. Таблица состояний определяет функциональную организацию системы и ту роль, которую играют отдельные состояния в этой организации. В отличие от МТ, переходы от состояния к состоянию не детерминированы, а имеют вероятностный характер, т. е. разум представляет собой вероятностный автомат. Критикуя теорию Путнама, Блок и Фодор отмечали, что человеческий разум имеет потенциально бесконечное количество состояний, в отличие от конечного количества состояний вероятностного автомата, определяемого таблицей состояний [Block, Fodor, 1972]. В предложенной Фодором версии КВТР, названной им «репрезентативная теория разума», особое внимание уделяется символам, которыми манипулирует разум в процессе тьюринг-подобных вычислений [Fodor, 1975]. Набор этих символов он называет «языком мысли». Смысл сложных композиций является функцией смысла отдельных частей и способа их комбинирования, задаваемого набором алгоритмов.

Классическая ВТР, в частности теория Фодора, нейтральна в отношении субстрата, производящего вычисления. Она потенциально приложима и к дуалистическим интерпретациям, когда вычисления производит картезианская душа, и к физикалистским интерпретациям, которые субстратом для вычислений видят мозг. На практике все адепты КВТР являются физикалистами и считают, что символы «языка мысли» закодированы в активности нейронов, а алгоритмические операции над символами представляют собой какие-то нервные процессы, природа которых пока неизвестна. По утверждению Фодора и многих других приверженцев КВТР, все ментальные процессы сводятся к тьюринг-подобным вычислениям с использованием набора символов и правил манипулирования этими символами [Fodor, 2005; Gallistel, King, 2009].

В 1980-х гг. появилась коннекционистская теория (КонТ), альтернативная классической. Создатели КонТ, в отличие от приверженцев КВТР, черпают вдохновение не в компьютерных науках, а в нейрофизиологии. Основой для вычислений они считают не машину Тьюринга, а нейронную сеть. Существенно отличаясь от МТ, нейронная сеть представляет собой коллекцию связанных друг с другом узлов, которые можно разделить на три категории: входные, выходные и скрытые (расположенные между входными и выходными). Каждый узел в каждый момент времени можно охарактеризовать уровнем активации, выраженным реальным числом и взвешенными связями с другими узлами, веса которых

(т. е. силу связи) можно также выразить реальным числом. Активация входных узлов задается экзогенно и служит входным сигналом для вычислений. Общая активация любого скрытого или выходного узла является функцией суммы взвешенных (на силу связи) активаций питающих его узлов и его собственной активации. Форма этой функции может быть разной для разных сетей.

Вычисление в сети представляет собой волну активации, распространяющуюся от входных к выходным узлам в соответствии с весами связей в сети. В прямой сети волна распространяется лишь в одном направлении. В рекуррентных сетях есть петли обратной связи. Рекуррентные сети труднее для математического представления, чем прямые сети, однако они очень важны для моделирования многих психологических феноменов [Elman, 1990]. Веса в нейронной сети обычно могут меняться в соответствии с алгоритмом «обучения». Описано множество разных алгоритмов, но основная идея состоит в том, чтобы постепенно подстраивать веса так, чтобы результат вычислений в сети приближался к целевому результату, который можно ожидать для соответствующей входной информации. Алгоритм обратного распространения (backpropagation algorithm) является широко используемым примером такого алгоритма [Rumelhart et al., 1986a].

Так же как и КВТР, КонТ берет свое начало в классической работе Уоррена Маккаллока и Вальтера Питтса, которые исследовали сети логических операторов «И» и «ИЛИ» [McCulloch, Pitts, 1943]. Эти сети можно рассматривать в качестве прототипа нейронных сетей, в которых уровень активации узла имеет лишь два значения (0 или 1), а функция активации определяется верностью соответствующего выражения. Маккаллок и Питтс считали логические операторы идеализированной моделью нейронов. Современные цифровые компьютеры фактически представляют собой сети, построенные из логических операторов. В когнитивных науках, однако, сети, построенные из бинарных элементов, не удовлетворяют исследователей в качестве прототипа нейронных сетей. В частности, обычно подчеркивается аналоговый характер уровня активации нейронов.

Начиная со второй половины 1980-х гг. нейронные сети стали привлекать все большее внимание для объяснения разнообразных когнитивных феноменов, включая распознавание объектов, восприятие речи, понимание предложений и так далее [Rumelhart et al., 1986b]. В 2010-х гг. в компьютерных науках особую популярность

приобрели модели, известные как «глубокие нейронные сети», или «сети глубокого обучения», которые содержат большое количество скрытых узлов (иногда сотни слоев) и тренируются на больших базах данных с использованием того или иного алгоритма обучения (обычно алгоритм обратного распространения) [LeCun et al., 2015]. Использование этих сетей позволило получить замечательные результаты во многих областях ИИ, таких как распознавание объектов и стратегические игры. Сейчас они получили широкое распространение в коммерческих приложениях, а также для моделирования когнитивных процессов [Marblestone et al., 2016].

Хотя КВТР и КонТ часто рассматривают как альтернативные теории, они по сути не являются взаимоисключающими – нейронные сети могут моделироваться и классической МТ, как это делается в современных цифровых компьютерах, и, с другой стороны, классические вычисления по типу машины Тьюринга могут осуществляться и с использованием нейронных сетей [Graves et al., 2014]. Сторонники КонТ обычно подчеркивают, что нейронные сети гораздо лучше моделируют работу мозга, чем вычисления по типу МТ. Узлы сети моделируют нейроны, а соединения между узлами – синапсы. Однако реальное устройство мозга гораздо сложнее, чем конструкция любой из современных нейронных сетей. Нейроны гораздо более гетерогенны, чем узлы сети, в частности, они имеют разную нейрохимическую природу. Кроме того, они сочетают аналоговый мембранный потенциал с дискретными потенциалами действия.

Самое важное отличие состоит в том, что большинство алгоритмов машинного обучения (например, алгоритм обратного распространения) требуют наличия эталона, т. е. результата, к которому нужно стремиться в процессе обучения. Этот эталон программа получает от оператора, знающего правильный ответ. Ничего подобного нет в работе мозга и сознания. В некоторых нейронных сетях вместо алгоритма обратного распространения используется алгоритм обучения с подкреплением [Pozzi et al., 2018] и другие алгоритмы, для которых не нужен эталон [Krotov, Hopfield, 2019], но они пока менее успешны. Кроме того, коннекционистские модели пока не способны предложить правдоподобных механизмов долговременного хранения информации в мозге, кроме безнадежно неэффективных петель ревербирующей активности.

Еще один вопрос, на который пока не может ответить ни КВТР, ни КонТ: как мозг, построенный из относительно гораздо медленнее (по сравнению с компьютером) работающих элементов (достаточно вспомнить время синаптической задержки и время распространения импульса по аксону), может осуществлять такие сложные вычисления так быстро [Gallistel, King, 2009]. Наконец, пока нет понимания того, как вычислительные теории разума могут объяснить каузальную силу содержания (семантики) сознания. Очевидно, что традиционные вычислительные системы манипулируют символами на основе синтаксических правил, описываемых алгоритмами (например, таблица состояний в машине Тьюринга). Они, таким образом, «слепы» в отношении семантики, т. е. смысла символов, наподобие того, как обитатель «китайской комнаты» «слеп» в отношении смысла китайских фраз, которыми он манипулирует.

Адепты КВТР настаивают на том, что семантические интерпретации «языка мыслей» не влияют прямо на результаты вычислений, т. е. не имеют каузальной силы. Результаты вычислений зависят исключительно от формальных правил манипулирования символами, т. е. разум является «синтаксической машиной» [Fodor, 1980]. Сторонники КонТ более осторожны в отрицании роли семантики, но в большинстве своем также согласны с выводом о вычислении, имеющем преимущественно синтаксический характер. Многие философы, критикуя этот тезис сторонников КВТР и КонТ, считают поведение человека и результаты его рассуждений принципиально зависимыми от семантики его мыслей [Block, 1990; Kazez, 1995; Figdor, 2009]. Некоторые критики вычислительных теорий разума доказывают, что возможности человеческого разума превосходят возможности алгоритмических машин типа МТ [Nagel, Newman, 1958; Lucas, 1961; Penrose, 1989]. Интуиция, креативность, инсайт не могут быть смоделированы на основе алгоритмов, так же как и многие другие особенности человеческого сознания [Dreyfus, 1992]. Даже такие апологеты КВТР, как Фодор (особенно в его последних трудах), часто выражают скептицизм в отношении того, что КВТР может объяснить все важные свойства человеческого разума [Fodor, 2000].

В целом вычислительные теории разума представляют интересную попытку формального описания ментальных операций. Они являются своего рода интерфейсом между когнитивными и компьютерными науками и пока, похоже, более полезны для по-

следних, чем для первых. Появившиеся в процессе развития этих теорий и их приложений наработки, такие как различные варианты нейронных сетей и машинного обучения, получили широкое применение в разных сферах и позволили достичь существенного прогресса в области ИИ. Эти достижения сейчас находят применение в том числе и при изучении и моделировании активности мозга и ментальных свойств. Успехи, достигнутые при использовании нейронных сетей глубокого обучения в решении таких задач, как классификация объектов по семантическим категориям и распознавание зрительных образов, показывают, что похожие механизмы может использовать и мозг для решения аналогичных задач. Сомнительно, однако, что КВТР, КонТ и их варианты смогут описать и объяснить все особенности человеческого сознания и разума.

При рассмотрении вычислительных теорий разума нельзя обойти вниманием стандартное среди когнитивных психологов описание активности мозга термином «обработка информации» (information processing). Понятие информации стало очень популярным начиная с работы Клайда Шеннона, заложившего в 1948 г. основы математической теории коммуникации [Shannon, 1948]. Шеннон понимал информацию как меру уменьшения неопределенности, которая выражается в виде изменения распределения вероятностей возможных состояний [Cover, Thomas, 2006]. Есть и другие формулировки информации. Фред Дретске трактовал информацию, содержащуюся в какой-либо переменной, через ее корреляцию с другой переменной [Dretske, 1981]. Например, количество колец на срезе дерева коррелирует с его возрастом, значит, оно содержит информацию о возрасте дерева. Еще одно определение, «семантическая информация», относится к содержанию репрезентации (например, репрезентация объекта внешнего мира в сознании человека) [Fodor, 1998; Sprevak 2010].

Понятие информации приобрело колоссальное значение в разнообразных науках, включая физику (например, квантовая теория информации), многие разделы инженерных наук, компьютерные науки, ИИ и когнитивные науки (например, рассмотренная выше теория интегрированной информации). Ряд ученых и философов, начиная, пожалуй, с Джона Уиллера [Wheeler, 1989], предлагают считать информацию одной из основ мироздания, наряду с материей [Lloyd, 2007; Seife, 2007; Davis, 2010; Vedral, 2010; Gleick, 2011]. При этом значение, в котором используется термин «информация», часто не оговаривается.

Информация, в каком бы смысле она не определялась, имеет значение лишь для существа, способного ее извлечь и понять. Как известно, для обладающего сознанием и разумом человека информация в разных ее видах имеет первостепенное значение. Можно думать, что собака, обнюхивающая забор со следами, оставленными другими собаками, извлекает информацию, имеющую для нее значение. Менее очевидно, но потенциально возможно использование понятия информации для объяснения поведения растений или бактерий. В отношении неорганической природы, если речь не идет о ее описании человеком, понятие информации теряет смысл. Поэтому философ Джон Серл считает объяснение феномена сознания с помощью информации примером циркулярного мышления, поскольку само понятие информации предполагает наличие обладающего сознанием существа, способного ее понять [Searle, 2013].

Из поверхностного обзора теорий сознания и разума можно видеть, что данная проблема пока остается недоступной для редуccionистского решения. Именно это мотивирует некоторых теоретиков искать ее решение путем пересмотра базовых представлений о природе реальности [Chalmers, 1996; Rosenberg, 2004; Stapp, 2009]. Существующие в рамках редуccionистского подхода теории в значительной степени умозрительны и ни одна из них пока не имеет надежного эмпирического обоснования. Более того, ни одна из рассмотренных теорий даже и не пытается объяснить природу и механизм корреляций между нервной активностью и содержанием сознания и постулировать нервный субстрат, который идентичен, или, по крайней мере, понятным образом соответствует феноменологии сознания [Metzinger, 2000]. Переходя далее к обзору эмпирических данных о соотношении нервной активности и феноменологии сознания, важно на начале определить, о каких феноменах сознания идет речь и как они могут изучаться.

3.3.3. Феноменология сознания

Очевидно, что экспериментальные подходы к вопросу о том, как содержание сознания возникает в результате активности мозга, и теоретические интерпретации эмпирических данных в решающей степени зависят от того, что мы подразумеваем под содержанием сознания. Опустив пространные философские трактовки различных свойств сознания [см., например [van Gulick, 2014]], сосредоточимся на том, что чаще всего подразумевается под со-

держанием сознания в соответствующих работах и называется «феноменальным сознанием» (the phenomenal consciousness; см., например [Dennett, 1991; Chalmers, 1995; Frankish, 2016]). Хотя философы не испытывают беспокойства, рассуждая о феноменальном сознании, не совсем понятно, как это явление может быть представлено в научных терминах. Философские описания сознания обычно относятся к таким его характеристикам, как качественный характер субъективного опыта, его феноменальные свойства и его субъективность (см., например [van Gulick, 2014]). Все эти характеристики являются предметом серьезных споров среди философов.

Качественный характер субъективного опыта относится к некоему типу «ощущений» (например, красноты красного – the redness of red), которые могут сопровождать не только чувственное восприятие, но и любой вид опыта, включая мысли и желания [Siewert, 1998]. Эти ощущения, которые называются квалиа, часто описываются в терминах «каково это» (what it is like), и их существование считается центральным для философского понимания сознания [Nagel, 1974, Chalmers, 1996]. Термин «феноменальные свойства» иногда используется взаимозаменяемо с термином «квалиа», но, кроме этого, он часто относится к интенциональным и репрезентативным свойствам и содержанию. Однако в последнем качестве сознательные феноменальные состояния трудно отделить от бессознательных состояний, также имеющих интенциональные и репрезентативные аспекты [Siewert, 1998]. Обсуждается вопрос о том, ограничиваются ли феноменальные свойства сенсорными ощущениями или сопровождают любое ментальное содержание, такое как вера или мышление [Prinz, 2012]. Субъективность также часто приравнивается к качественным или феноменальным аспектам сознания и относится к тому факту, что содержание сознательного опыта доступно только субъективно и не может быть полностью постигнуто с точки зрения третьего лица [Nagel, 1974].

Различные состояния сознания объединены понятием сознательного опыта. «Все эти состояния (состояния сознания) объединяет то, что в них есть что-то, что характерно для каждого состояния (there is something it is like to be in them). Все они являются состояниями сознательного опыта» [Chalmers, 1995, p. 200]. Трудная проблема сознания (the hard problem of consciousness) в определении Чалмерса состоит в том, чтобы объяснить, как мозг про-

изводит эти состояния. «Когда мы думаем и воспринимаем, есть поток обработки информации, но есть и субъективный аспект. Этот субъективный аспект и есть сознательный опыт» [Ibid]. В основе всех приведенных описаний лежит знаменитая фраза Нагеля [Nagel, 1974]: «есть что-то, что характеризует...» (пребывание в этом состоянии, или бытие этого существа). Данный аспект называется «субъективным характером (или качеством) опыта», и это (what-it's-likeness) получило название «феноменальное сознание».

Определение Нагеля чрезвычайно двусмысленно и является предметом многочисленных споров среди философов. Например, Хакер считает, что «переживания, как правило, индивидуализируются не в зависимости от свойств самого переживания, а в зависимости от свойств предмета, который является объектом переживания» [Hacker, 2012, p. 16]. Красный цвет – это не качество восприятия, а качество воспринимаемого объекта (например, красной розы). Релевантность таких философских конструкций, как квалиа, для научного описания процесса восприятия также может быть подвергнута сомнению на основании эмпирических данных, которые показывают, что мы обычно воспринимаем объекты целостно и, если это не является специальной задачей, не анализируем их путем извлечения различных свойств, таких как цвет или форма. Понятия «краснота» или «округлость» – примеры безграничной способности человека к обобщению и абстракции. Понятие «краснота» не появляется как «ощущение», когда мы воспринимаем какой-либо красный объект, скорее, оно может быть извлечено при сравнении различных красных объектов и выяснении того, что у них общего, несмотря на явные различия в форме и размере.

Кроме всего прочего, накоплено много свидетельств того, что наше восприятие всегда является прогнозирующим (так называемое предсказывающее кодирование, predictive coding), т. е. оно основано на априорном ожидании того, каким может быть воспринимаемый объект с учетом ситуационного контекста [Friston et al., 2006]. Воспринимаемые объекты классифицируются на очень ранних стадиях восприятия. Хотя мы не знаем, как именно мозг делает это, нам известно, что извлечение смысла постепенно уточняется на разных уровнях вентрального визуального пути, позволяя классифицировать воспринимаемый объект, например, как лицо или место [Kanwisher et al., 1997]. Эта категоризация пред-

положительно основана на одновременном использовании множества воспринимаемых комбинаций характеристик и их сравнении с информацией, хранящейся в памяти. Заключительный этап – распознавание, т. е. отождествление объекта с чем-то, что мы уже знаем, с автоматическим присвоением ему всех уже известных свойств. Предположительно, это происходит даже с объектами, которых мы никогда раньше не видели, – в данном случае мозг также пытается извлечь смысл, сравнивая их с тем, что он уже знает.

Важно подчеркнуть, что все процессы категоризации и распознавания происходят очень быстро и полностью бессознательно, и только результат появляется в сознании. Таким образом, невозможно сознательно «почувствовать» все многообразие квалиа, связанных с восприятием свойств объекта. Такое чувство может возникнуть только в том случае, если мы сознательно сконцентрируем внимание на определенном свойстве объекта, и даже тогда оно будет неразрывно связано с этим объектом. Просто попытайтесь почувствовать «красноту» или «округлость» без связи с каким-либо объектом. Мы легко манипулируем подобными абстрактными понятиями, но это не «чувства». Вообще говоря, сознание не имеет дела с «сырыми чувствами», оно основано на смыслах, извлеченных из сырых данных на бессознательных стадиях, и это относится не только к восприятию, но и ко всем ментальным явлениям, включая мысли. Опять же, мы можем сознательно концентрироваться на «чувствах», но только если намеренно стремимся к этому.

Утверждение о качественной природе любого субъективного опыта спорно и по другой причине. Ценностные ориентации в отношении объектов и событий действительно являются общей чертой людей и, предположительно, всех живых существ. Мы склонны не просто отмечать начало дождя, но и сопровождать это наблюдение оценочным комментарием, например, «хорошо для урожая» или «плохо для прогулки». У других живых существ, которые не могут поделиться с нами своими чувствами, оценочный характер их восприятия часто проявляется в их поведении приближения или избегания [Schneirla, 1939]. Эти ценностные ориентации очень адаптивны, но не считаются обязательным атрибутом любого восприятия или другого ментального процесса. «Каждый опыт является возможным субъектом предикатов отношения, например, быть приятным или неприятным, интересным или скуч-

ным, привлекательным или отталкивающим. Но неверно, что каждый опыт является действительным субъектом такого предиката отношения. Что касается большинства переживаний, то на вопрос “каково это было?”, или “как ты это оцениваешь?” правильным ответом будет: “ничего особенного” и “это было совершенно безразлично» [Hacker, 2012, p. 17].

Субъективность и недоступность сознательного опыта также относительны, когда речь идет о людях. Человеческий язык позволяет нам делиться содержанием субъективного сознательного опыта с другими и сравнивать разные точки зрения. Важно помнить, что только благодаря этой способности человека мы вообще можем знать о существовании сознания. Нам остается только фантазировать, обладают ли существа, не являющиеся людьми, сознанием, и упражнять свое воображение, пытаясь почувствовать, «каково это быть летучей мышью». Возможности языка, безусловно, ограничены, и трудно точно описать смутные чувства (одни справляются с этим лучше, чем другие). Однако мы умеем расшифровывать эмоциональные состояния людей по их жестам и мимике, и, в целом, для обычных людей в повседневной жизни делиться друг с другом своим субъективным опытом не является непреодолимым препятствием.

В научных исследованиях был разработан ряд экспериментальных поведенческих парадигм, которые, помимо самоотчетов, позволяют нам получать информацию о содержании субъективного сознательного опыта. Наконец, любая мысль, включая научную теорию, сначала появляется в некотором уме как субъективный сознательный опыт, а затем может стать культурным наследием. Существование человеческой культуры было бы невозможным, если бы люди не могли делиться друг с другом содержанием своего субъективного сознательного опыта. Следовательно, заявление о том, что содержание субъективного сознательного опыта в принципе недоступно с точки зрения третьего лица, может рассматриваться как преувеличение, которое искусственно создает непреодолимое препятствие для научного исследования сознания. «Истинная природа» любого объекта научного исследования (например, элементарных частиц в квантовой физике) всегда недоступна во всей своей глубине, но это не означает, что мы должны отказаться от изучения этого объекта. Прагматические подходы, такие как копенгагенская интерпретация квантовой механики, которые игнорируют то, что в принципе недоступно, и использу-

ют доступную информацию (знаменитый принцип «заткнись и вычислай»), см., например [Kaiser, 2014]), могут дать отличные результаты.

Таким образом, учитывая неоднозначный и противоречивый характер философских определений феноменальных свойств, не представляется разумным использовать их в научном исследовании сознания. Выяснить, «как мозг производит» субъективное «ощущение», недоступное с позиции стороннего наблюдателя, кажется практически невозможным. В темной комнате действительно сложно найти черную кошку. На мой взгляд, было бы разумнее использовать то, что Нед Блок называет «доступным сознанием» – информацию, которая доступна сознательному существу для использования и руководства действием. Эта информация может включать в себя ментальные представления внешних объектов и их свойств или внутренне генерируемые мысли и чувства, если они доступны для использования «в рассуждении и рациональном управлении речью и действием» [Block, 1995, p. 227].

3.4. Эмпирические данные о нервном субстрате сознания

Самый очевидный источник данных о содержании сознания – интроспекция. Надежность этого источника может быть подвергнута сомнению (см., например [Schwitzgebel, 2011]), однако во многих экспериментальных контекстах он считается вполне надежным [Spener, 2015]. Если испытуемый может рапортовать о каком-то содержании своего сознания или использовать его для сознательного рационального поведения, значит, это содержание доступно для интроспекции [Block, 1995]. Таким образом, широкое определение доступности содержания сознания рассматривает целенаправленное преднамеренное поведение как критерий доступности. Если некоторое содержание сознания может использоваться в таком поведении, значит, оно доступно. Это определение позволяет использовать показатели поведения, в том числе и у животных, для оценки содержания сознания [Wu, 2018].

Содержанием сознания может быть восприятие объекта, например, лица, или свойства, например, направления движения объекта, а также воображаемые объекты, воспоминания, мысли и вообще все то, что наполняет наше сознание. Если в основе чувственной репрезентации объекта X_c лежит его нервная репрезентация X_n , то X_n должно быть необходимым и достаточным условием

появления $Xч$. Необходимость можно проверить, устраняя нервное содержание. Если $Xч$ исчезает после устранения $Xн$, значит, наличие $Xн$ является необходимым условием для возникновения $Xч$. Достаточность можно проверить путем генерации $Xн$, что должно привести к возникновению соответствующего $Xч$. Демонстрация необходимости и достаточности $Xн$ для возникновения $Xч$ ничего не говорит, однако, о причинной роли того и другого в отношении поведения. Согласно эпифеноменологической редукционистской интерпретации, $Xн$ порождает $Xч$ и независимо от этого вызывает соответствующую поведенческую реакцию $Xп$. Устранение $Xч$ никак не повлияло бы на $Xп$, однако экспериментально это проверить невозможно, так как предполагается, что $Xн$ неизбежно сопровождается появлением $Xч$. Единственным доступным способом доказательства эпифеноменологической природы $Xч$ является демонстрация содержательной идентичности $Xн$ и $Xч$, поскольку субъективный опыт говорит нам, что именно содержание сознания каузально связано с нашим поведением.

Представим, например, простой эксперимент, в котором испытуемому предъявляется на экране монитора движущаяся точка. Задача испытуемого – нажать на клавиатуре стрелку, направление которой совпадает с направлением движения точки. Например, нужно различить два варианта – движение слева направо ($Xн1$) и движение снизу вверх ($Xн2$). Если скорость выполнения задания не ограничивается, можно ожидать, что для подавляющего большинства испытуемых задача будет решаться правильно в 100 % случаев. Если после каждого нажатия спрашивать испытуемого, какое движение он видел, он подтвердит, что видел точку, перемещавшуюся слева направо ($Xч1$), или, соответственно, снизу вверх ($Xч2$).

Таким образом, будет подтверждено соответствие между содержанием интроспективно доступного зрительного восприятия ($Xч$) и соответствующим ему поведением ($Xп$). Чтобы доказать, что $Xч$ в данном случае эпифеномен, нужно выявить паттерны нервной активности ($Xн1$ и $Xн2$), которые позволили бы предсказать $Xн1$ и $Xн2$ не хуже, чем это делают $Xч1$ и $Xч2$. Более того, надежность предсказания $Xн1$ и $Xн2$ на основе $Xн1$ и $Xн2$ в выборке испытуемых должна быть не хуже, чем надежность предсказания $Xн1$ и $Xн2$ на основе $Xч1$ и $Xч2$. Это значит, что $Xн1$ и $Xн2$ должны быть идентичны у разных испытуемых, аналогично тому, как у них идентичны $Xч1$ и $Xч2$. Можно возразить, что $Xч$ и $Xп$ – это по

сути одно и то же, измеренное разными способами (соответственно с помощью интроспекции и с помощью целенаправленного поведения) и, значит, они заведомо должны коррелировать друг с другом лучше, чем с нервной активностью, измерение которой неизбежно сопряжено с ошибкой в силу неточности доступных методов.

Представим, например, что как X_c , так и X_n являются результатом активности одного и того же нервного субстрата (X_n), который каузально необходим и достаточен для обеспечения X_n . Тогда X_c и X_n должны идеально коррелировать друг с другом и гораздо хуже коррелировать с X_n , так как ошибки интроспекции и измерения соответствующего поведения должны быть заведомо меньше, чем ошибки измерения X_n . Такое возражение, однако, можно поставить под сомнение в свете недавно выявленной роли дорзального и вентрального зрительных трактов в обеспечении поведения и сознательного восприятия соответственно (см. ниже раздел 4.4.1.). Эти данные показывают, что ее использование, по крайней мере в отношении зрительной информации, для целенаправленного поведения и для сознательного восприятия связаны с активностью анатомически различных структур мозга, так что корреляция между X_c и X_n не является в данном случае следствием того, что оба они – результат активности одного и того же нервного субстрата.

3.4.1. Феноменология и неврология зрительного восприятия

Имеет смысл начать с феноменологии зрительного восприятия, во-первых, потому, что нервные механизмы зрительного восприятия лучше изучены и, во-вторых, потому что она проще, чем феноменология, например, понимания речи, или феноменология мышления. Пять ученых получили Нобелевскую премию за исследования в области физиологии зрительного восприятия [Morais, 2017]: Грэнит, Уолд и Хартлайн – по медицине в 1967 г. за изучение нервных и химических процессов в сетчатке глаза, дающих начало зрительному восприятию [Raju, 1999]; Хьюбел и Уизел – по физиологии в 1981 г. за изучение обработки информации в зрительной коре. Хьюбел и Уизел показали, как информация, поступающая от сетчатки, обрабатывается в зрительной коре, чтобы генерировать сигналы, соответствующие восприятию краевых линий, движения, стереоскопической глубины и цвета, которые являются строительными блоками визуальной сцены [Wurtz, 2009].

Начиная с работ Ангерлейдера и Мишкина [Ungerleider, Mishkin, 1982], принято различать в корковом представительстве зрительной системы приматов две подсистемы – дорзальный и вентральный пути. Первый проецируется в теменную долю, а второй – в височную [Kravitz et al., 2011]. Ангерлейдер и Мишкин изначально постулировали, что функции этих путей связаны с ответом на вопрос «где» и «что» соответственно, т. е. дорзальный путь участвует в пространственном восприятии, а вентральный – в категориальном. Позднее Милнер и Гудэйл доказывали, что дорзальный путь поставляет зрительную информацию для руководства оперативным поведением, а вентральный – для восприятия, т. е. для понимания смысла образов, памяти и планирования [Milner, Goodale, 1995]. Соответственно, по предположению авторов, дорзальный путь связан с неосознанным восприятием, а вентральный – с сознательным [Goodale, Milner, 2004]. Это предположение во многом основывалось на наблюдениях за пациентами с повреждениями соответствующих областей мозга. Как показали результаты наблюдений, повреждение дорзального пути не влияет на осознанное восприятие зрительной информации, но вызывает оптическую атаксию, т. е. неспособность использовать зрительную информацию для управления действиями [Andersen et al., 2014]. Повреждения вентрального пути, наоборот, нарушают сознательное восприятие зрительной информации, приводя к зрительной агнозии, но оставляют сохранной способность использовать зрительную информацию для руководства действиями [Fahrah, 2004].

Еще один феномен неосознанного восприятия зрительной информации – так называемое слепое зрение (*blindsight*), наблюдающееся у пациентов с повреждениями в области первичной зрительной коры (V1). Такие повреждения обычно сопровождаются неспособностью видеть часть пространства, контралатеральную поврежденной стороне мозга [Weiskrantz, 1986]. При этом часто сохраняется способность использовать не воспринимаемую зрительную информацию для руководства своим поведением [Gelder et al., 2008].

Функции дорзального и вентрального путей до сих пор остаются предметом дискуссий (см., например [Schenk, McIntosh, 2010]), однако большинство исследователей признает, что сознательное восприятие зрительной информации связано преимущественно с вентральным зрительным трактом, а ее использование

для поведения – с дорзальным. Аналогичное разделение показано и для слуха. Вентральный слуховой путь начинается в первичной слуховой коре и участвует в выделении фонем из остальных звуков и преобразования их в слоги [Nickok, Poeppel, 2007; DeWitt, Rauschecker, 2012]. Этот путь затем сливается с вентральным зрительным трактом в средней височной извилине и височном полюсе, где слуховые объекты превращаются в аудиовизуальные [Gow, 2012]. Дорзальный слуховой путь входит в заднюю верхнюю височную извилину и затем к началу дорзального зрительного тракта около сильвиевой борозды. Функция дорзального слухового пути состоит в том, чтобы преобразовывать слуховые сенсорные репрезентации в артикуляционные моторные, что необходимо при обучении речи [Nickok, 2012].

Приведенные выше данные ставят под сомнение правомерность эпифеноменологической интерпретации сознательного восприятия. Если функция мозга (отобранная, по представлениям эпифеноменологов, в процессе эволюции) состоит лишь в обеспечении поведения, а сознание возникает как эпифеномен этой функции, то трудно объяснить возникновение в процессе данной эволюции двух анатомически различных трактов, участвующих в обеспечении поведения и сознательного восприятия соответственно.

Далее рассмотрим данные в пользу нейронной основы специфического содержания сознания на примере восприятия зрительной информации. Они должны ответить на вопрос: какие нейронные состояния или свойства необходимы и/или достаточны для того, чтобы сознательное восприятие имело содержание X , а не Y ? Один из подходов к ответу на поставленный вопрос – эксперименты с бинокулярным соперничеством (binocular rivalry), когда каждый глаз получает различную информацию. Испытуемый при этом не воспринимает два образа одновременно, а попеременно видит то один, то другой [Blake et al., 2014]. Теоретически это можно объяснить «соперничеством» между монокулярными нейронами на ранних стадиях обработки стимула, или между бинокулярными нейронами, которые получают сигнал от обоих глаз на более поздних стадиях обработки. Данные ранних исследований на обезьянах свидетельствовали в пользу второго варианта – показатели стимул-зависимого поведения обезьян лучше коррелировали с активностью нейронов нижней височной коры (inferotemporal cortex, IT), чем с активностью нейронов первичных

зрительных областей (V1 или V2) [Logothetis et al., 1996]. Функциональные МРТ-исследования показали, однако, что у людей активность первичной зрительной коры (V1) также коррелирует с восприятием [Polonsky et al., 2000].

Более поздние работы продемонстрировали важность процессов на разных уровнях зрительного анализатора [Tong et al., 2006]. Если конкурирующие образы содержат части, которые могут быть слиты в одно целое, испытуемый воспринимает этот интегрированный образ [Ngo et al., 2000], т. е. результат бинокулярного соперничества возникает в процессе интеграции, чувствительной к глобальным свойствам стимула, а не просто отражает «победу» одного нервного сигнала над другим [Baker, Graf, 2009].

В целом результаты экспериментов с бинокулярным соперничеством позволяют выявить, на каком уровне зрительного анализатора происходит выбор варианта, представленного в сознании, но, так же как и в теоретически постулируемых вариантах соперничества между разными модулями (теория множественных набросков и ТГРП), механизм этого выбора, или интеграции, неясен. Более того, эти данные ничего не говорят о том, как содержание выбранного варианта закодировано в активности мозга. Тут важно отметить, что в большинстве экспериментов с бинокулярным соперничеством используются стимулы, относящиеся к разным семантическим категориям (например, лицо и здание, или мужское и женское лицо). Поэтому обнаруживаемые различия в активности мозга на уровне конечных стадий зрительного анализатора, соответствующие разному содержанию сознания, в этом случае лишь подтверждают, что восприятие зрительных объектов сопровождается их категоризацией в относительно широкие семантические категории.

Другой пример – исследование восприятия направления движения. Известно, что важную роль в этом восприятии играет часть средней височной извилины (MT или V5 у людей). Разрушение MT приводит к неспособности видеть движение объектов [Zihl et al., 1983]. Регистрация ответов нейронов MT на восприятие движущихся в разных направлениях объектов в экспериментах на обезьянах показывает, что каждый из них имеет максимальный ответ на определенное направление движения и менее выраженный ответ на другие направления движения. То есть ответ каждого нейрона имеет вероятностный характер и может быть описан с помощью функции распределения плотности вероятности [Lee, Maunsell, 2009].

В последние годы стало популярным применять для описания их поведения байесовскую модель. Согласно этой модели, извлечение информации из активности популяции нейронов дает не направление движения (например, слева направо), а, скорее, функцию распределения вероятностей в пространстве возможных направлений движения. По байесовской модели, нейроны как бы рассчитывают апостериорную вероятность с учетом эффекта стимула и вероятности события на основе предшествующего знания [Colombo, Seriès, 2012; Shea, 2014; Rescorla, 2015]. То есть мозг представляет сенсорную информацию в вероятностных терминах, рассчитывая функцию распределения вероятностей.

Однако содержание чувственного образа не имеет вероятностного характера – мы определяем направление движения объекта однозначно, а не в вероятностных терминах. Теннисист во время матча даже долю секунды не сомневается, летит ли мяч в правый или в левый сектор. Это подчеркивает несоответствие между нейронным и чувственным содержанием сенсорного восприятия [Morrison, 2017; Block, 2018] и напоминает ситуацию в квантовой механике, где движение квантового объекта описывается функцией распределения вероятностей, и в момент измерения эта вероятностная функция «коллапсирует» в определенность. Где и как происходит «коллапс» функции распределения вероятностей сенсорной информации и из нее извлекается определенное чувственное содержание?

Как показывают данные фМРТ-исследований, информацию об ориентации зрительных образов можно извлечь с помощью математического анализа даже из активности первичной зрительной коры [Kamitani, Tong, 2005]. Непонятно, однако, где и как эта информация извлекается в мозге. Предложены гипотезы для объяснения того, как вероятностная репрезентация на уровне нейронов превращается в определенность на уровне сознания. По гипотезе Блока, например, имеет место соревнование между нейронными популяциями, представляющими разные возможные черты (например, движение налево или направо) и победитель в этом соревновании «получает всё» [Block, 2018]. Бросается в глаза сходство этой гипотезы с рассмотренной выше теорией множественных набросков Деннета. В обоих случаях, как и в теории эволюции Дарвина, решающая роль отводится свободной конкуренции и естественному отбору. Однако, если в теории эволюции критерии отбора примитивно просты (проигравшие физически элимини-

руются, не успев дать потомства), то в данном случае остается неясным вопрос, как и где происходит выявление «победителя». В последних строках своей статьи Блок не забывает подчеркнуть, что он не допускает наличия в мозге конечной стадии процессинга, на которой принимается решение (картезианский театр в терминологии Деннета), однако альтернативного объяснения не приводит [Ibid].

Выявление «победителя» требует как минимум существования некоего «компаратора», на котором должна сходиться активность всех соревнующихся модулей. Если придерживаться эпифеноменологической интерпретации сознания, то этим компаратором может быть только некая популяция нейронов. Она должна отвечать на приходящие от первичных модулей сигналы разной интенсивности по принципу «всё или ничего», например, она будет пропускать сигналы одного модуля и блокировать все остальные. Активность этой популяции и всех вышестоящих в иерархии сенсорного анализа популяций нейронов должна однозначно соответствовать содержанию чувственного образа. Только в таком случае можно было бы думать, что субъективное чувственное восприятие как бы и не нужно – один и тот же физический процесс (активность нейронов «компаратора») дает начало как субъективному чувственному восприятию, так и всем последующим в иерархии активности мозга физическим процессам, ведущим в конечном счете к поведению. Каузальной силой в этой схеме должен обладать именно «компаратор».

Вопрос, почему вообще существует субъективное чувственное восприятие, все равно остался бы неясным, но ответ на него можно было бы отнести к сфере метафизики. Существование компаратора постулируется в некоторых теоретических моделях сознания (см., например, теорию Грея [Gray, 1995]), однако эмпирически ничего похожего на компаратор в мозге пока обнаружить не удалось. Различные характеристики стимула и даже субъективную уверенность в наличии той или иной характеристики можно предсказать (с той или иной степенью надежности) из активности нейронов, записанной либо инвазивными (на животных, или на людях при нейрохирургических операциях), либо неинвазивными (фМРТ) методами [Burge, 2010; van Bergen et al., 2015; Peters et al., 2017]. Эти предсказания имеют вероятностный характер, и остается непонятным, как вероятностная информация, содержащаяся в мозге, превращается в определенную информацию в сознании.

На людях механизмы представления семантической информации в мозге изучаются в основном неинвазивными методами, преимущественно с использованием фМРТ. Общее направление этих исследований можно описать в терминах кодирования и декодирования информации. При изучении кодирования характеристики стимула используются для предсказания активности мозга, а при изучении декодирования активность мозга – для предсказания характеристик стимула [Naselaris et al., 2011]. Заметим в скобках, что, строго говоря, данные исследования не выявляют нервный субстрат содержания сознания, потому что нервная активность сравнивается не с содержанием сознания, а с характеристиками стимула, которые предположительно должны быть представлены в сознании.

В качестве математической модели иногда используются обычная линейная регрессия или итеративная байесовская модель, а в последние годы общепринятым стало применение нейронных сетей и алгоритмов машинного обучения. Для интерпретации результатов этих исследований важно учитывать, какие характеристики стимула используются в модели. При анализе как кодирования, так и декодирования традиционно рассматривают три набора переменных, каждый из которых можно представить в виде многомерного пространства.

Пространство стимула отражает физические характеристики стимула. Например, в случае монохромных зрительных образов пространство стимула представляет собой n -мерное пространство, где n – это количество пикселей в изображении, а значение по соответствующей шкале – яркость пикселя. Каждый образ является уникальной комбинацией значений по шкалам и задается точкой в этом n -мерном пространстве. Пространство мозга моделируется в виде m -мерного пространства, где m – количество вокселей в соответствующей области мозга (например, в первичной зрительной коре), а значение по каждой шкале выражает активность вокселя. Общее понимание в физиологии сенсорного восприятия состоит в том, что мозг реагирует на определенные признаки стимула, как это показано, например, для зрительного восприятия в работах Хьюбела и Уизела [Wurtz, 2009]. Поэтому важно поместить между пространствами стимула и мозга промежуточное k -мерное пространство признаков, где k соответствует количеству извлекаемых признаков. Каждое изображение является уникальной комбинацией признаков и представлено точкой в этом пространстве.

Обычно считается, что связь между пространством стимула и пространством признаков имеет нелинейный характер, а связь между пространством признаков и пространством мозга линейна [Wu et al., 2006]. Признаки могут извлекаться разными способами. Например, в работе Кая с соавторами в качестве признаков использовали вейвлет-преобразование переменных пространства стимула [Kay et al., 2008]. Однако в большинстве исследований пространство признаков представляет собой некую предположительно осмысленную для испытуемого категоризацию стимулов (например, лица версус здания, или различные сегменты пространства в задачах на ориентацию).

В исследовании Назелариса с соавторами сравнивалась предсказательная сила извлеченных разными способами признаков в отношении активности разных зон зрительного анализатора. Оказалось, что структурные признаки изображения, такие как положение в пространстве, ориентация и пространственная частота, извлеченная с помощью вейвлет-преобразования входного сигнала, лучше предсказывали активность ранних областей зрительной коры (V1, V2 и V3), а семантические признаки (категоризация стимулов по содержанию) – более высоко расположенных в иерархии областей зрительного анализатора (передняя затылочная кора) [Naselaris et al., 2009]. Для категоризации стимулов по содержанию они сравнивались с набором из 1750 изображений, которые были предварительно классифицированы на 23 семантические категории независимыми наблюдателями. Был использован алгоритм оптимизации максимизации ожидания для определения вероятности активации каждого вокселя в конкретной области мозга в ответ на предъявление изображения, относящегося к той или иной категории.

В результате для каждого вокселя были получены распределения вероятностей его ответа на стимул каждой категории. Показано, что распределения вероятностей ответа на структурные и семантические признаки не перекрываются, т. е. одни воксели преимущественно отвечают на структурные, а другие на семантические признаки. Кроме того, показано, что точность предсказаний для вокселя сравнима с описанной в экспериментах на животных точностью предсказаний для отдельных нейронов. Отобранные воксели (т. е. те, которые отвечали на семантические признаки) были объединены в общую модель. Когда использовались лишь две семантические категории (одушевленные и неоду-

шевленные объекты), максимальная точность предсказания этой модели достигала 90 %; когда использовались 23 категории, максимальная точность была 40 % [Naselaris et al., 2009].

Из этого относительно подробного описания исследования, которое можно считать типичным, можно сделать несколько выводов. Во-первых, структурные и семантические признаки образа представлены в разных отделах зрительного анализатора. Механизм извлечения семантической информации из структурной остается неизвестным. Во-вторых, ответ отдельных вокселей, так же как и отдельных нейронов, имеет вероятностный характер. То есть каждый воксель с большей вероятностью отвечает на стимулы определенной категории, но может отвечать и на стимулы других категорий. Как из этих вероятностей извлекается определенность – остается неизвестным. В-третьих, в данной работе для отнесения изображения к той или иной категории использовалась база данных из большого количества изображений. Можно думать, что и в мозге отнесение стимула к той или иной категории происходит путем сравнения с образами, хранящимися в памяти, однако механизм этот пока неизвестен. Наконец, в-четвертых, при использовании 23 семантических категорий точность предсказания составляла всего лишь 40 %. То есть на основе активности мозга в 40 % случаев можно правильно определить, к какой из 23 категорий относится стимул, а в 60 % случаев определение будет неверным.

Нужно подчеркнуть, что 23 категории выбраны произвольно и в реальности количество категорий при восприятии зрительной информации несопоставимо больше. При увеличении же количества категорий резко снижается точность предсказания. Можно ожидать (хотя такие данные в работе и не собирались), что при использовании вербального отчета точность предсказания была бы близка к 100 %. То есть по этим данным семантическая информация, предположительно представленная в сознании, лишь очень грубо соответствует активности мозга, и эти данные, таким образом, трудно интерпретировать в свете редукционистской концепции.

3.4.2. Байесовская модель

Как было выше сказано, в когнитивных науках все бóльшую популярность приобретают байесовские вероятностные модели принятия решений в условиях неопределенности. Можно кратко

рассмотреть этот подход на примере сенсорного восприятия, которое лучше других ментальных процессов соответствует байесовской модели. Сенсорное восприятие – классический образец принятия решений в условиях неопределенности. Еще Гельмгольц подчеркивал, что проблема неопределенности является эндемической для сенсорного восприятия [Helmholtz, 1867]. Например, у зрительного анализатора нет непосредственного доступа к отдаленному окружению, у него есть доступ только к сенсорным стимулам, возникающим в сетчатке. Как же зрительный анализатор принимает решение о свойствах отдаленного объекта на основе ограниченной информации, поставляемой сетчаткой? Гельмгольц предполагал, что стимулы, поступающие из сетчатки, запускают неосознаваемый процесс принятия в качестве решения наиболее правдоподобной гипотезы о свойствах отдаленного объекта. Подход Байеса к моделированию сенсорного восприятия основан на этом предположении Гельмгольца. Байесовская модель восприятия включает пространство гипотез, в котором каждая гипотеза h относится к какому-либо аспекту воспринимаемого объекта (форма, размер, цвет и так далее). Исходная вероятность $p(h)$ – это начальное правдоподобие, приписываемое h . Исходная вероятность сенсорного сигнала e в свете гипотезы h оценивается как $p(e|h)$. После получения сигнала e система пересчитывает вероятность h с учетом e . По теореме Байеса апостериорная вероятность рассчитывается как $p(h|e) = \eta \times p(h) \times p(e|h)$, где η – размерностная константа, обеспечивающая возможность того, чтобы все вероятности суммировались в единицу. На основе апостериорной вероятности система выбирает наиболее правдоподобную гипотезу. Этот выбор управляется функцией максимизации полезности, которая оценивает последствия ошибки [Rescorla, 2019].

Сейчас «байесовскую когнитивную науку» выделяют даже в виде отдельной дисциплины, которая рассматривает байесовские модели восприятия, моторного контроля, причинных рассуждений, социального познания, интуиции, навигации и анализа естественного языка. Предполагается, что в основе всех этих процессов лежит вероятностный анализ информации байесовского типа, который происходит на подсознательном уровне. Критики байесовской парадигмы, однако, указывают на большое количество ментальных процессов, протекающих не по байесовскому типу. По убеждению Канемана и Тверского, рассуждения на сознательном уровне обычно не используют исчисления вероятностей, а

при принятии решений часто не учитывается ожидаемая максимизация пользы [Kahneman, Tversky, 1979; Tversky, Kahneman, 1983]. Некоторые считают, что и на подсознательном уровне ментальная активность часто не подчиняется законам байесовской парадигмы [Morales et al., 2015]. В разделе 3.3, посвященном теориям сознания, мы рассмотрели данные, согласно которым активность мозга и ментальная активность подчиняются разным законам – классическая теория вероятности, включая байесовскую модель, и булева логика лучше описывают активность мозга, а квантовая теория вероятности и небулева логика лучше описывают ментальную активность.

Кроме того, байесовская модель не объясняет, откуда берутся гипотезы. Можно думать, что гипотезы генерируются с учетом ситуационного контекста (предсказывающее кодирование) на основе информации, хранящейся в памяти, однако как рассчитывается их исходная вероятность и как все это закодировано в активности мозга, остается неизвестным [Orlandi, 2016]. Сторонники байесовского подхода соглашаются с тем, что конкретные механизмы пока неясны и вряд ли можно объяснить все ментальные феномены на основе теории Байеса, но некоторые из последних, особенно сенсорное восприятие и моторный контроль, лучше всего описываются именно в рамках этой теории [Rescorla, 2019].

Так или иначе, мало кто сейчас сомневается, что активность мозга (так же как неосознанная ментальная активность, к которой она эксплицитно или имплицитно приравнивается) имеет вероятностный характер, независимо от того, можно ли ее описать в рамках теории Байеса. Как из глубин этого вероятностного океана всплывают в сознании вполне определенные образы и решения, до сих пор остается загадкой. Примечательны в этом отношении приобретенные недавно импульс со стороны исследования мозга дебаты о реальности или иллюзорности свободы воли. Данный вопрос будет рассмотрен более подробно далее, но в рамках текущего обсуждения можно бросить взгляд на некоторые аспекты упомянутых дебатов. Они имеют глубокие корни в философии, но недавний всплеск интереса к теме связан с экспериментами нейрофизиолога Бенджамина Либета, показавшего, что электрический потенциал готовности появляется в двигательной коре на 300–500 мс раньше времени, которое испытуемые рапортуют как время принятия сознательного решения [Libet, 1985].

Эксперименты имели большой резонанс и широко интерпретировались как доказательство отсутствия свободы воли – все на самом деле решает мозг и нам только кажется, что мы принимаем решения сознательно [Свааб, 2019]. Такая интерпретация использовалась как веский аргумент в пользу эпифеноменологической природы сознания. Эксперименты Либета и их интерпретация критиковались специалистами, но суть в данном случае не в этом. Если активность мозга и неосознанная ментальность имеют вероятностный характер, а определенность возникает лишь на уровне сознания, то как это можно совместить с представлением о том, что именно активность мозга обладает каузальной силой в отношении поведения и принятия решений, а содержание сознания – это лишь эпифеномен? Согласно всему нашему опыту, именно определенность содержания нашего сознания каузально связана с принимаемыми нами решениями.

Возвращаясь к зрительному восприятию, нужно подчеркнуть, что, по существующим представлениям, кодирование зрительных образов в мозге организовано в виде многоуровневой системы, которая послужила прототипом для создателей нейронных сетей глубокого обучения. В первичной зрительной коре происходит извлечение базовых атрибутов нижнего уровня, таких как форма, пространственные отношения (включая положение в пространстве и размер), движение, текстура, яркость и цвет [Burge, 2010]. Бердж, обсуждая атрибуты высшего уровня, предполагает, что для биологически важных свойств, таких как пища, опасность, укрытие, могут существовать специальные атрибуты высшего уровня.

В целом исследования кодирования и декодирования, в частности, при восприятии зрительных стимулов, показывают, что атрибуты как низшего, так и высшего уровня могут быть с некоторой степенью надежности извлечены из активности мозга с помощью математического анализа [Hasson et al., 2010; Wen et al., 2018]. Кодирование зрительной информации в мозге изучено лучше всего, но есть исследования кодирования и слуховой информации, например, фрагментов музыки [Hoeffle et al., 2018]. Механизмы извлечения атрибутов высшего уровня в мозге неизвестны. Также неизвестно, как на основе этих атрибутов происходит категоризация объектов. Исходя из успешности использования нейронных сетей глубокого обучения для решения задач категоризации и распознавания образов, можно думать, что похожие механизмы использует и мозг.

3.4.3. Нейролингвистика

Вторая после зрительного восприятия лучше всего исследованная область – кодирование в мозге вербальной информации, изучаемое нейролингвистикой. Цель нейролингвистики – расшифровать нейрональную основу знания и использования языка. Эти исследования берут свое начало в работах французского невролога XIX в. Поля Брока, обнаружившего связь между повреждением определенной области левой нижней лобной извилины и нарушением продукции речи. Работы Брока и его немецкого коллеги и современника Карла Вернике легли в основу классической модели нейролингвистики, согласно которой продукция речи связана с областью Брока, а понимание с областью Вернике, расположенной в задней части левой верхней височной извилины. Хотя эта классическая модель многими критиковалась в конце XIX – начале XX в., она получила новый импульс в 1960-х гг. в основном благодаря работам Нормана Гешвинда [Geschwind, 1965; Geschwind, Levitsky, 1968].

Неоклассическая модель постулировала, что продукция и понимание речи связаны с областью Брока и Вернике соответственно, а также зависят от связывающего эти области дугообразного пучка (*arcuate fasciculum*). С появлением нейровизуализационных методов эта модель подверглась существенной модификации. Хотя важность областей Брока и Вернике в этих исследованиях получила подтверждение, было показано участие и большого количества областей за их пределами, в том числе и в правом полушарии [Bates, Dick, 2000]. Было также доказано, что продукция и понимание речи не могут быть однозначно связаны с дихотомией областей Брока и Вернике: область Брока участвует и в понимании, а область Вернике – в продукции [Stowe et al., 2005]. Кроме того, дихотомия «продукция – понимание» не исчерпывает всех сфер лингвистики.

Много внимания в последних работах уделяется вопросу: способен ли мозг различать разные уровни лингвистической структуры, такие как фонология, синтаксис и семантика [Vambini, 2012]. По результатам исследований, повреждение области Брока избирательно нарушает синтаксический компонент и в продукции, и в понимании речи [Caramazza, Zurif, 1976]. Была выдвинута альтернативная классической гипотеза, согласно которой область Брока участвует как в продукции, так и в понимании синтаксиса, а область Вернике важна как для продукции, так и для понимания семантики [Swinney, 1999].

В более поздних нейровизуализационных исследованиях представление о «центрах речи» было заменено представлением о сетевой репрезентации лингвистических процессов, а дихотомия «продукция – понимание» заменена изучением разных компонентов лексической структуры, таких как грамматические классы или семантические категории [Price, 2010]. Важные области таких исследований, например, – изучение дихотомий «существительное – глагол» [Vigliocco et al., 2011] или «одушевленные – неодушевленные» объекты [Martin, 2007]. Однако в этих исследованиях остается нерешенной «проблема несоответствия гранулярности», т. е. несоответствия элементов лингвистической структуры (слог, словосочетание, параграф) и лингвистических операций (конкатенация, генерация фразеологической структуры и семантическая композиция) элементам нейробиологической структуры (нейрон, или нейронная колонка коры) [Poeppel, Embick, 2005; Grimaldi, 2012]. Несмотря на это, как показывают полученные данные, элементам лингвистической структуры можно найти корреляты в активности мозга. Например, восприятие только иерархически организованных грамматических правил специфически активирует область Брока и соседние с ней области, что расценивается как свидетельство коркового представительства «универсальной грамматики» [Tettamanti, Perani, 2012].

В связи с обсуждаемыми вопросами наибольший интерес представляет кодирование в мозге семантической, т. е. смысловой информации. Определение смысла слова на первый взгляд кажется очевидным. Это именно объект или явление, обозначаемое данным словом. Оно может быть написанным или озвученным. Вместо конкретного слова можно подобрать синонимы. Можно, наконец, использовать слова других языков – смысл при этом не меняется. Лингвисты, однако, до сих пор не пришли к единому мнению в отношении того, как определить смысл слова. Споры идут, например, по поводу того, в какой степени связанное с данным объектом или явлением знание может быть частью смысла слова. С точки зрения минималистов, значение слова – это встроенное языковое понятие, которое нельзя рассматривать в отрыве от языка. То есть никакое относящееся к данному объекту или явлению знание не может быть частью значения слова. Максималисты же, наоборот, утверждают, что значение слова уходит корнями в человеческое знание, опыт, ментальные репрезентации и другие нелингвистические концепции.

В вычислительной лингвистике принято считать, что значение слова распределено через все контексты, в которых это слово может быть использовано, и количественно его можно определить путем простого подсчета всех возможных контекстов. Этот подход называют распределительной семантической моделью (PCM, *distributional semantic model*). Модель позволяет представить значение слова в виде реального числа, путем подсчета количества контекстов, в которых данное слово используется в достаточно большой базе письменных текстов. Однако она не может определить, в чем именно состоит смысл слова, и вся полезность модели ограничивается возможностью ее использования в простых инженерных расчетах. Тем не менее PCM наиболее часто применяется в когнитивных и нейролингвистических исследованиях.

Существует два основных варианта PCM. В первом из них составляется контекстуальная матрица слов. Каждое слово в ней задано рядом значений, равных количеству случаев, когда это слово встречается в контексте с каким-либо другим словом. В последние годы гораздо чаще используется другой вариант PCM, в котором с помощью нейронных сетей рассчитывается вероятность найти слово в определенном контексте. Используется прямая (*feed-forward*) нейронная сеть. В ее первом слое есть вектор начальных весовых коэффициентов для каждого слова в словаре. Проходя через тело текста методом скользящего окна, сеть учится предсказывать вероятность нахождения каждого слова в определенном контексте, заменяя начальные весовые коэффициенты на рассчитанные вероятности. Полученная модель называется моделью встраивания слова (*MBC, word embedding*). Эту модель можно получать и другими методами, основная идея остается той же и была сформулирована уже в 1950-х гг. – «слово характеризуется его компанией» [Firth, 1957]. Слова, относящиеся к одному и тому же семантическому домену, чаще встречаются в одинаковых контекстах. Степень сходства можно оценить корреляцией весовых коэффициентов в матрице. Например, для слов «месяц» и «неделя» корреляция равна 0.74, а для слов «месяц» и «высокий» –0.22 [Huth et al., 2016].

Первая работа, в которой PCM применялась в нейролингвистическом исследовании, была опубликована в 2008 г. [Mitchell et al., 2008]. Была использована фМРТ, записаная в эксперименте с предъявлением 60 конкретных существительных, принадлежащих

12 семантическим категориям, которые предъявлялись девяти испытуемым в случайном порядке 6 раз каждое. Рассчитывался фМРТ-ответ на каждое слово как среднее всех ответов на это слово.

Вероятность активации вокселя в фМРТ-образе мозга представлялась в виде взвешенной суммы семантических значений слова, умноженной на коэффициент, рассчитанный описанной в статье моделью. Слово было представлено в виде вектора, в котором каждое значение соответствовало его встречаемости с одним из 25 сенсомоторных глаголов (например, «видеть», «слышать» и т. д.) в большом теле текста (триллион слов). Средняя точность предсказания семантической категории на основе фМРТ была 0.77, а предсказания слов внутри категории 0.62. Для случайно сгенерированных (бессмысленных) категорий точность предсказания была 0.60. Наиболее информативные области мозга были расположены в левой нижней височной и лобной извилинах, а также в зрительных областях и моторной коре и, в случае глаголов, с большей вероятностью обнаруживались поблизости от функционально когерентной области коры (например, моторная кора для глагола «толкать»). Наиболее яркий результат этой работы состоял в том, что локализация семантических центров и сила ответа на стимулы были более или менее одинаковы у разных испытуемых.

Следующая веха в исследованиях в этой области заложена в работе [Huth et al., 2016], ее авторы поставили себе задачу описать «семантическую карту» мозга, каждая единица которой отвечает сильнее на слова с определенными семантическими характеристиками. В отличие от предшествующих работ, в качестве стимулов использовались не отдельные слова, а прослушивание фрагмента текста; МВС каждого слова из прослушанной истории строили путем вычисления его сочетания с каждым из 985 обычных английских слов (например, «выше», «беспокойство», «мать») в большом теле английского текста и использовали для выявления ответов каждого вокселя в фМРТ-образе мозга. Кластеризация семантических признаков всех слов позволила создать 12 семантических категорий: «тактильная» (включала слова типа «пальцы»), «зрительная» (например, слово «желтый»), «числовая», «локационная» (например, «стадион»), «абстрактная» («естественный»), «время», «профессиональная» («встреча»), «насильственная» («летальный»), «коммунальная» («школы»), «ментальная» («спящий»), «эмоциональная» и «социальная» («ребенок»).

Анализ фМРТ-ответов методом главных компонент выявил четыре компонента, объясняющих большую часть вариации. Первый, наиболее сильный в плане объясненного разнообразия компонент одним своим концом представлял категории, относящиеся к людям и социальным взаимодействиям (социальная, эмоциональная, насильственная и коммунальная), а другим перцептивные, количественные и пространственные дескрипторы. Остальные три компонента труднее поддавались интерпретации. Можно заметить, что полученные семантические категории охватывают далеко не все возможные смысловые категории и, более того, некоторые из них кажутся достаточно случайными. Это может быть связано с использованием ограниченной лингвистической базы, на основе которой строилась модель.

В исследовании [Pereira et al., 2018] была поставлена задача создать универсальный декодер, потенциально способный извлечь из активности мозга, записанной в процессе чтения текстов, значения слов, фраз и предложений на любую тему, включая абстрактные идеи. Семантические векторы были посчитаны для всех слов базового словаря (~30 000 слов). Кластеризация этих векторов позволила выявить 200 семантических категорий, 20 из которых были отброшены в силу трудности интерпретации. Затем из каждой категории было выбрано представительное слово и декодер тренировали путем предъявления этих слов. Далее декодер тестировали на новом лингвистическом материале. Для отдельных слов сравнивали реальный семантический вектор слова с предсказанным на основе данных активности мозга. Средняя точность предсказаний была 0.7 (при уровне случайности – 0.5). После этого авторы показывают, что декодер можно использовать и для предсказания смысла фраз и предложений, описывающих понятия, представленные в словах, использованных для тренировки декодера.

Средняя корреляция между реальным семантическим вектором фразы и декодированным из фМРТ-данных вектором была 0.35. Для конструкции декодера использовали 5000 наиболее информативных вокселей у каждого испытуемого. Как и в предыдущих работах, локализация этих вокселей была похожа у разных испытуемых. Оказалось, что они достаточно широко разбросаны по разным областям коры и известным сетям покоя: 21 % – языковая сеть; 15 % – дефолтная сеть; 23 % – сети внимания; 19 % – зрительная сеть; 22 % – другие области мозга.

Рассмотрев этапные и авторитетные исследования в области кодирования семантической информации, попробуем разобраться в том, как эти данные можно интерпретировать в плане соответствия представленного в сознании смысла слов выявляемым с помощью фМРТ паттернам активности мозга. Прежде всего нужно отметить, что использование МВС в качестве идентификатора смысла слова имеет свои ограничения. Модель встраивания слова зависит от размеров и содержания базы текстов, использованных для ее построения. На одних и тех же данных разные МВС могут давать разные результаты [Abnar et al., 2017]. Кроме того, два слова (или фразы) с высокой корреляцией их семантических векторов могут иногда обозначать противоположные смыслы. Например, модели слов «налево» и «направо» будут очень похожи, так как эти слова в текстах обычно сочетаются с одними и теми же глаголами (например, «иди налево» и «иди направо»), однако их смысл по сути противоположен. Поэтому полученные с помощью МВС «семантические карты» мозга непригодны для расшифровки истинного смысла услышанных фраз.

Семантические категории, будь то 12 категорий, как в работе [Huth et al., 2016], или 200 категорий, как в работе [Pereira et al., 2018], – это лишь грубая тематическая классификация смыслов, наподобие того, как в библиотеке книги расставляют по темам – химия на одной полке, а детективные истории на другой. Это, кстати, повторяет организацию, например, зрительного анализатора, где восприятие лиц связано с одной частью веретенообразной извилины (face area), восприятие мест – с другой (place area), а восприятие форм – с третьей (shape area) [Kanwisher et al., 1997]. Для того чтобы такая тематическая классификация была возможна, необходимо одно условие – тематические категории должны каким-то образом распознаваться на предшествующем этапе анализа. Однако механизмы этого распознавания пока неизвестны. Предположительно, оно должно включать извлечение информации из семантической памяти и сравнение признаков воспринимаемых объектов с признаками, характерными для разных семантических категорий. В любом случае существующие данные о том, как семантическая информация отражена в активности мозга (будь то содержание зрительных стимулов или слов) не укладываются в представление о каузальном превосходстве мозга над ментальностью.

Строго говоря, рассмотренные выше работы не исследуют связь между активностью мозга и содержанием сознания. Смысл воспринимаемых образов или слов оценивался не по отчетам испытуемого, а косвенно, по характеристикам стимула, интуитивно связанным со смыслом. Можно думать, что, если бы в этих работах отчеты испытуемых регистрировались, они бы коррелировали с этими характеристиками на уровне, близком к единице. Активность мозга, как было показано, коррелирует с семантическим вектором на уровне 0.35 [Pereira et al., 2018]. То есть активность мозга позволяет с некоторым приближением извлечь грубую семантическую характеристику стимула, но не более того. Средняя точность предсказания семантической категории, к которой относится слово, на основе активности мозга была достоверно выше уровня случайных совпадений, но не очень от него отличалась (0.7 и 0.5 соответственно, см.: [Pereira et al., 2018]). Это значит, что так же, как в случае восприятия зрительных образов, активность мозга при восприятии лингвистической информации имеет вероятностный характер и механизмы извлечения определенного смысла каждого слова из этого распределения вероятностей остаются непонятными.

3.5. Как же семантика «кодируется» в мозге?

В рассмотренных выше двух примерах восприятия семантической информации – распознавание зрительных образов и речи – есть немало общего, что позволяет выявить общие принципы организации отражения семантики в активности мозга. В обоих случаях имеет место классификация воспринимаемых стимулов по тематическим категориям, которые представлены в мозге в анатомически разных областях (например, лица, места и формы для зрительных образов или семантические категории для слов). По каким признакам и с помощью каких механизмов происходит классификация – пока непонятно. Категории эти грубые, что не позволяет с достаточной точностью предсказать по активности мозга смысл воспринимаемого стимула. Кроме того, активация структур мозга имеет вероятностный характер, и для извлечения из нее семантической информации необходим сложный математический анализ. Сложность извлечения семантической информации из сенсорных стимулов хорошо иллюстрируется трудностями моделирования этих процессов в компьютерных науках. Хотя в

последние годы в области распознавания зрительных образов и понимания речи компьютером достигнут большой прогресс, эффективность компьютера пока уступает эффективности человека. Прогресс достигнут благодаря применению методов байесовского анализа и машинного обучения. Предполагается, что похожие механизмы использует и мозг, однако как конкретно эти механизмы реализованы в мозге – неизвестно.

В частности, большой пласт работ посвящен механизмам так называемого предсказывающего кодирования (predictive coding), однако мы не будем здесь подробно рассматривать эту большую тему. В общих чертах, согласно теории предсказывающего кодирования, восприятие сенсорной информации является активным процессом. Мозг постоянно генерирует и обновляет ментальную модель окружающего мира. На основании этой модели генерируется предсказание (ожидание) сенсорного входного сигнала, которое сравнивается с реальным сигналом, и рассогласование между ожиданием и реальностью (ошибка предсказания) используется для коррекции ментальной модели. Идея предсказывающего кодирования при восприятии внешнего мира была высказана Гельмгольцем еще в 1860 г., разрабатывалась американским психологом Джеромом Брюнером в 1940-х гг. и стала предметом экспериментальных исследований после опубликования в 1981 г. статьи МакКлелланда и Румелхарта [McClelland, Rumelhart, 1981]. Современные модели предсказывающего кодирования постулируют механизмы, основанные на байесовском подходе и воплощенные в мозге в виде иерархических многослойных сетей с восходящими и нисходящими связями между слоями [Clark, 2013].

На интуитивном уровне идею предсказывающего кодирования легко понять из повседневного опыта. Если человек входит, например, в хорошо знакомое помещение, он ожидает увидеть хорошо ему знакомые предметы на привычных местах и, если картина соответствует ожиданию, он ее практически не замечает (нулевая ошибка рассогласования). Если же, вопреки ожиданию, он обнаружит посреди комнаты новый и неизвестный ему предмет (большая ошибка рассогласования), все внимание будет направлено на то, чтобы понять, что это за предмет, и объяснить его появление в этом месте. Одна из самых больших проблем теории заключалась в смутном понимании того, как именно минимизация ошибок предсказания может быть устроена в мозге. В некоторых фМРТ-исследованиях увеличение BOLD-сигнала интерпретиру-

ется как сигнал ошибки, в других же – как входной сигнал [Kogo, Trengove, 2015]. Ключевой нерешенный вопрос – что именно представляет собой сигнал ошибки и как он рассчитывается на каждом уровне обработки сигнала в мозге [Bastos et al., 2012]. Другая проблема состоит в том, что вычисления, которые, согласно модели, должны производиться на каждом уровне иерархической сети, могут быть компьютерно неразрешимыми [Kwisthout, van Rooij, 2020]. В целом модель предсказывающего кодирования теоретически, безусловно, очень привлекательна, однако понимание того, как она может быть воплощена в активности мозга, пока очень смутное.

И самое, пожалуй, главное – никто пока не показал наличие в мозге активности, один к одному отражающей представленную в сознании семантику стимула. Как уже было отмечено выше, связь содержания сознания с поведением не вызывает сомнения. Чтобы доказать, что сознание – это эпифеномен активности мозга, необходимо продемонстрировать наличие в мозге активности, соответствующей текущему содержанию сознательного восприятия. Какие бы методы анализа не использовал мозг (байесовский анализ, предсказывающее кодирование или методы, похожие на машинное обучение), на выходе должен быть результат в виде нейронной активности, однозначно соответствующей сознательному восприятию.

Но, может, как утверждает Деннетт, мы и не должны искать в мозге этот финальный результат, аналог картезианского театра? Может быть, по крайней мере, в случае поведения, результат в виде победителя в соревновании конкурирующих нейронных модулей замыкается непосредственно на двигательном эффекторе и проявляется в виде однозначной активации двигательной коры в ответ на стимулы? То есть компаратором, о котором говорилось выше, являются непосредственно двигательные нейроны? Однако исследование организации моторного контроля показывает, что он также имеет вероятностный характер [Wolpert, 2007] и хорошо аппроксимируется байесовской моделью принятия решений [Wolpert, Landy, 2012].

Таким образом, активность мозга, сопровождающая как сенсорное восприятие, так и моторный контроль, имеет вероятностный характер. Во многих случаях эта активность хорошо описывается байесовской моделью, в которой вероятность получения некоего результата (будь то результат сенсорного восприятия оп-

ределенного стимула или результат движения) оценивается на основе предшествующего знания и сравнивается с реальным результатом, а ошибка рассогласования используется для корректировки знания и для принятия решения с учетом цены ошибки в конкретной ситуации. Для нашего рассмотрения даже не так важно, будет ли в конечном счете принята или отвергнута байесовская модель для описания этих процессов в мозге. Важно то, что эти процессы в любом случае имеют вероятностный характер и принятие решения происходит в условиях неопределенности. Необходимо подчеркнуть, что речь тут идет не об ошибках, связанных с погрешностью методов измерения активности мозга, а о природе самой нервной активности.

Как же из этой вероятностной картины возникает определенность, присущая нашему восприятию реального мира и наших действий в нем? Представьте, какой сложной была бы жизнь, если бы при пересечении перекрестка нам каждый раз нужно было бы рассчитывать вероятность того, что мы видим именно зеленый цвет светофора, или что машина движется слева направо, а не наоборот, или что мы сами движемся в том направлении, в котором решили двигаться? Все эти расчеты, оценки вероятностей и механизмы принятия решений не достигают нашего сознания. Мы обычно не сомневаемся, что видим то, что видим, и делаем то, что хотим сделать, и основанное на этом поведение оказывается эффективным, иначе мы не смогли бы выжить в этом мире.

Важно еще раз особо отметить следующее: мозг содержит информацию о семантических свойствах стимула и использование сложных методов анализа позволяет извлечь ее с какой-то степенью вероятности, и даже информацию, коррелирующую с субъективной оценкой надежности этой информации [Li et al., 2021a]. Нет, однако, никаких намеков на то, может ли сам мозг осуществлять подобного рода анализ и продуцировать таким образом содержание нашего сознания. Показать наличие такого механизма необходимо для доказательства редукционистской интерпретации сознания.

НЕОСОЗНАННОЕ

В предшествующей главе были рассмотрены данные о соотношении между содержанием сознания и активностью мозга. Невозможно отрицать, однако, что значительная часть нашей ментальной жизни происходит за пределами сознательного восприятия. Накоплен большой эмпирический материал, показывающий, что эта неосознаваемая часть ментальности может влиять на поведение, иногда помимо нашей воли. Хотя размышления о роли неосознаваемой ментальности можно найти у многих ученых и философов прошлого, наиболее систематическая разработка этой темы впервые появилась в трудах психоаналитиков Зигмунда Фрейда и Карла Юнга. Однако, несмотря на широкую популяризацию и колоссальное влияние идей психоаналитиков на все последующее развитие психологии, мы не будем здесь их обсуждать, поскольку они относятся все-таки к донаучному этапу и не имеют весомого эмпирического обоснования. В этой главе будем использовать термины «неосознанное», «имплицитное» и «подсознательное» как синонимы, хотя исторически разные авторы в разное время придавали им разный смысл. Фрейд в первых своих трудах использовал термин «подсознание» (*subconscious*), но потом отказался от него, заменив его термином «неосознанное» (*unconscious*). Современные авторы часто предпочитают употреблять термин «имплицитное знание».

В отличие от психоаналитической традиции, движимой спекулятивными теоретическими концепциями, такими как комплекс Эдипа или комплекс Электры, когнитивная традиция исследования бессознательных процессов основана на относительно небольшом количестве теорий и преимущественно ориентируется на сбор эмпирических данных. Когнитивные исследования показали, что автоматически и явно неосознанно люди регистрируют и получают больше информации, чем то количество, которое они могут получить сознательно [Augusto, 2010]. В тех случаях, когда испытуемые демонстрируют поведение, указывающее на обладание некоторым знанием, но при этом не могут его выразить словами, можно предположить, что у них есть бессознательное или имплицитное знание, т. е. знание, которое доступно для использования в поведении, но не доступно для сознания.

Таким образом, под бессознательным знанием понимается знание, которое выявляется только при выполнении задания, когда субъекты не знают, что они получают к нему доступ, тогда как о сознательном знании мы говорим, когда субъекты осознают, что обладают им и получают к нему доступ [Schacter, 1992]. Примером бессознательного знания может быть уже рассматривавшийся нами феномен слепого зрения (blindsight), наблюдающийся у пациентов с повреждениями в области первичной зрительной коры. Такое знание можно интерпретировать как знание низшего уровня, а сознательное знание – как знание высшего уровня. По этой интерпретации сознательное знание – это метазнание, при котором сознание получает доступ к информации, хранящейся в «бессознательном» хранилище [Dienes, Perner, 2003]. Но часто делаются и более сильные утверждения. Многие считают, что бессознательное знание – это не просто знание, которое не достигает сознания, а качественно другой вид знания, получаемый с помощью других когнитивных (и мозговых) механизмов [Greenwald, 1992]. Примером могут служить уже рассмотренные нами данные о вентральном (сознательном) и дорзальном (бессознательном) зрительных трактах.

В эксперименте бессознательное знание изучается с помощью соответствующих методов. Например, испытуемым предъявляются стимулы, которые они не могут воспринимать сознательно потому, что они слишком слабы, кратки, сложны или замаскированы другими стимулами, но можно, тем не менее, показать, что имели место бессознательное восприятие, хранение и извлечение информации об этих стимулах. Другой подход – исследование состояний, при которых сознательное восприятие отсутствует (сон, кома, анестезия и пр.), или клинических случаев, связанных с неспособностью сознательно воспринимать некоторые стимулы (например, слепое зрение, геминеглект или прозопагнозия), или просто манипуляция вниманием в процессе эксперимента, когда с помощью инструкции осознанное внимание испытуемых направляется не на те стимулы или аспекты стимулов, которые реально являются предметом исследования. Важно подчеркнуть, что, по крайней мере в некоторых случаях, сознание может получить доступ к имплицитному знанию, но в эксперименте наличие этого знания можно тестировать и без участия сознания.

Во многих случаях бессознательное знание – это процедурное знание, т. е. знание, которое проявляется исключительно в каких-то процедурах и действиях и не может быть вербализовано. Обыч-

ные примеры такого рода знаний – езда на велосипеде, говорение на родном языке, распознавание лиц. Отмечен целый ряд различий между имплицитным и эксплицитным знанием, что подтверждает их принципиально разную природу. Есть данные, что знания, полученные без участия сознания, лучше сохраняются и более устойчивы к амнезии [Graf et al., 1984], имеют холистическую природу, т. е. их невозможно разбить на компоненты, в отличие от декларативного знания, структуру которого можно анализировать [Fodor, Pylyshyn, 1988].

Процедурное знание обычно негибкое и рутинное. Так, основанное на процедурном знании поведение нарушается, когда в условия задачи вносят изменения [Bailey et al., 2005]. Процедурное знание не коррелирует с эксплицитным знанием – улучшение поведения, основанного на процедурном знании, не сопровождается аналогичным улучшением вербализации [Berry, Broadbent, 1988] и, в отличие от эксплицитного знания, оно не коррелирует с IQ [Reber, 1992]. Хотя поведение на основе процедурного знания может быть достаточно сложным и требующим навыка, оно предсказуемо и может быть задано с помощью набора инструкций и алгоритмов, которым можно обучить с помощью классического или оперантного подкрепления. Предполагается, что процедурное знание – это эволюционно более раннее приобретение, а осознанное знание зарождается лишь с появлением развитого неокортекса [Augusto, 2010]. Соответственно, у рыб, амфибий и пресмыкающихся есть лишь процедурное знание. Человек на ранних (доязыковых) стадиях индивидуального развития приобретает колоссальный объем процедурных знаний, которыми пользуется потом всю оставшуюся жизнь [Fifer et al., 2010].

Однако сфера неосознанного отнюдь не ограничивается лишь приобретением автоматических поведенческих навыков. Уже в XIX в. некоторые психологи, интерпретируя результаты экспериментов по восприятию стимулов, утверждали, что мы можем различать стимулы внешнего мира совершенно неосознанно [Peirce, Jastrow, 1884]. Более поздние данные о зрительном восприятии у пациентов с повреждениями разных участков зрительного тракта, рассмотренные нами ранее, это бесспорно подтверждают. Тот факт, что в экспериментах с замаскированным праймингом испытуемые могут неосознанно обрабатывать смысл стимулов, показывает, что бессознательное восприятие может происходить и на более высоких уровнях обработки, а данные исследований паци-

ентов в состоянии анестезии и комы подтверждают гипотезу о возможности людей строить и активировать обширные базы знаний совершенно бессознательно [Augusto, 2010].

Согласно многочисленным примерам неосознанного обучения и имплицитной памяти, люди способны не только усваивать информацию (например, обучение во сне), но и извлекать сложные смысловые паттерны без участия сознания. Например, люди, как правило, не способны описать большинство грамматических правил, которые они используют, говоря на своем родном языке. Особенно поразителен этот факт в случае детей 5–6 лет, еще не знакомых с какими-либо правилами грамматики. Мы можем быстро выносить суждения о людях, опираясь на стереотипы и не зная о правилах и конструкциях, к которым прибегаем в таких случаях [Steele, Morawski, 2002]. Как видно из экспериментов с применением искусственной грамматики, достаточно просто предъявлять строки предложений, созданных по правилам искусственной грамматики, чтобы испытуемые бессознательно усвоили новые грамматические правила [Dienes, Perner, 2003]. Интересно, что грамматики в этих экспериментах слишком сложны для осознанного изучения даже в течение длительного периода времени. Кроме того, они не обязательно основаны на алфавите. Например, используя формы, цвета и т. д., можно продемонстрировать передачу бессознательного знания между модальностями [Manza, Reber, 1997].

Применение сложных систем правил и бессознательное приобретение знаний об этих системах исследовалось с конца 1970-х гг. Бродбентом и его коллегами [Broadbent et al., 1986]. Их исследования направлены на то, чтобы показать, что правильное выполнение задачи (достижение и поддержание заданных целевых значений путем изменения одной входной переменной) не зависит от способности субъектов вербализовать знания о системах, которыми их просят управлять, или о том, как им удастся успешно их контролировать. Хотя выполнение задачи улучшается с практикой, это не отражается на улучшении способности испытуемых отвечать на вопросы о работе системы. В отличие от этого, устные инструкции, данные испытуемым, улучшают их способность отвечать на вопросы, но не имеют никакого значения для их реальной работы по управлению системами. Это убедительно подтверждает гипотезу о том, что мы можем успешно действовать физически в ситуациях, в которых единственное знание, используемое нами, является бессознательным.

Принципиально важная сфера, в которой роль сознательного и неосознанного знания нуждается в прояснении, – мышление и принятие решений. Всегда ли мы осознаем свои решения? Какие бессознательные факторы определяют принятие решения? Принятие решений включает такие аспекты, как «рациональность» и «логичность», которые чаще всего отождествляются с сознанием. Однако данные исследований убедительно показывают, что «рациональное» и «логическое» мышление или обработка информации, содержащейся в базе знаний, может осуществляться в значительной степени или полностью бессознательно.

Часто выбор решения в сложных ситуациях происходит на основе интуиции, внутреннего чувства (*gut feeling*), не поддающегося немедленному описанию. Иногда рациональное обоснование принятых решений происходит постфактум, когда нужно объяснить себе или другим, почему было принято именно это решение. Для объяснения этого явления Дамасио с соавторами предложили гипотезу соматического маркера (ГСМ), которая постулирует решающую роль соматических состояний (эмоциональных изменений) в принятии решений. По этой гипотезе, когнитивные состояния связаны с соматическими изменениями, которые возникают в биорегуляторных процессах, и эти ассоциации, однажды сохраненные в памяти, вспоминаются в контексте, аналогичном тому, в котором они впервые произошли [Bechara, Damásio, 2005]. Согласно ГСМ, вентромедиальная префронтальная кора (ВМПФК) является областью мозга, в которой когнитивно обрабатывается и сохраняется ассоциация между фактическими знаниями и биорегуляторными процессами. Формирование этой ассоциации нарушено у пациентов с повреждением ВМПФК (например, знаменитый случай Финеса Гейджа) [Damásio, 1996].

Другая теория (*unconscious thought theory*) постулирует существование неосознанного мыслительного процесса и базируется на эмпирическом наблюдении, согласно которому люди часто принимают более правильные решения, когда они сознательно не контролируют процесс принятия решений [Dijksterhuis, Nordgren, 2006]. Эксперименты, в которых тестируется эта теория, включают периоды так называемой инкубации, в течение которой внимание испытуемого отвлекается от выполнения основной задачи и после которого решение этой задачи «внезапно» само собой появляется в сознании. Согласно этой теории, ограниченные ресурсы сознания препятствуют быстрому нахождению решения, в то вре-

мя как ресурсы подсознания практически неограниченны. Кроме того, подсознание работает дивергентно и по принципу «снизу вверх», а сознание – конвергентно и по принципу «сверху вниз» [Ibid]. Эффективность сознательной мысли линейно снижается при увеличении сложности задачи, а эффективность подсознания от нее не зависит. В этой теории, однако, не прописаны конкретные механизмы неосознанного мышления, и она критикуется как в плане теоретических постулатов, так и в плане используемых методов [Asker, 2008].

Нужно подчеркнуть, что, несмотря на немалое количество накопленных эмпирических фактов, существование неосознанного знания до сих пор остается предметом дебатов. По убеждению многих критиков, то, что приписывают неосознанным процессам, на самом деле является результатом работы сознания, а экспериментальные подтверждения существования неосознанного знания объясняют либо методическими ошибками, либо неверной интерпретацией (см., например [Shanks, St. John, 1994]). Часто дебаты в этой области крутятся вокруг субъективного порога осознания, т. е. точки, в которой испытуемые еще не знают, что они владеют определенной информацией, а также вокруг приписываемой людям способности иметь мысли более высокого порядка – осознавать свои собственные психические состояния [Dienes, 2007]. Если, однако, определять знания как установление успешных отношений с окружающей средой [James, 1904], то эмпирические данные в поведенческой и когнитивной психологии убедительно указывают на существование качественно отличного вида знания, приобретаемого, хранимого и воспроизводимого совершенно бессознательно [Augusto, 2010].

Для обсуждаемой нами темы важно, что сознательное и неосознанное знание качественно различны и связаны (в большинстве или во всех случаях) с активностью разных отделов мозга. Это убедительно опровергает представление о том, что сознание – лишь эпифеномен активности мозга, как шум является эпифеноменом работы двигателя. Представим себе двигатель, в котором было бы две отдельные части – одна чтобы бесшумно производить работу, а вторая – чтобы, не участвуя в работе, производить шум. Существующие теории постулируют, что сознание появилось в эволюции позже, чем бессознательный контроль поведения, и непропорциональное развитие связанных с сознанием областей неокортекса именно и определяет человеческий мозг [Ibid].

Как мы видели в разделе, посвященном эволюции человека (см. гл. 1), человеческий мозг – очень затратный орган и его появление связано с большим количеством сложностей, в сумме снижающих потенциал воспроизводства потомства. Чтобы тем не менее появиться и сохраниться в эволюции, связанные с сознанием области мозга должны были обеспечивать функции, перевешивающие по своей полезности все неизбежные потери. В свете наиболее популярных сейчас теорий о роли социальных факторов в эволюции человеческого мозга можно думать, например, что бессознательная часть мозга должна обеспечивать поведение, а сознание и связанные с ним области мозга нужны лишь для коммуникации и вербализации неосознанно принятых решений. Этому противоречит, однако, факт, согласно которому сознательные и неосознанные процессы не коррелируют друг с другом [Broadbent et al., 1986]. То есть сознание – это не просто вербализация неосознанного поведения, а качественно отличный способ управления поведением, и, чтобы отбираться в процессе эволюции, оба способа в сочетании друг с другом должны давать преимущество для выживания.

Альтернативное объяснение появления связанных с сознанием областей мозга в эволюции человека допускает, что оно не обеспечивало немедленное преимущество для выживания. По палеонтологическим данным, поведенческие проявления сознания, которые потенциально могли бы отбираться естественным отбором, возникли как минимум на 100 тыс. лет позже, чем мозг, анатомически неотличимый от мозга современного человека (см. гл. 1). Это объяснение, однако, предполагает действие каких-то факторов, не вписывающихся в классическую теорию Дарвина.

При обсуждении роли сознания и неосознанного в нашей ментальной жизни, принятии решений и контроле поведения часто в явной или неявной форме неосознанное отождествляется с активностью мозга. Подразумевается, что если природа связи между содержанием сознания и активностью мозга пока неизвестна, то уж ту часть ментальности, которая не отражается в сознании, мы можем с легкостью отождествить с активностью мозга. В этом рассуждении верно лишь одно – активность мозга не транслируется один к одному в содержание сознания. У нас, однако, нет никаких оснований думать, что активность мозга транслируется один к одному в содержание неосознанного. Природа связи между активностью мозга и содержанием неосознанного еще бо-

лее туманна, чем природа связи между активностью мозга и содержанием сознания, так как содержание неосознанного, в отличие от содержания сознания, можно оценивать лишь по косвенным поведенческим признакам. И в том и в другом случае, несмотря на немалое количество теорий и спекуляций, конкретные механизмы мозга остаются неизвестными.

В психологии, особенно после появления психоанализа, был большой запрос на методы определения содержания бессознательного. Однако в большинстве своем предложенные методы (интерпретация снов или произвольных оговорок, проективные тесты, наподобие теста Люшера и Роршаха) оказались ненадежными и не используются в научной психологии. Исключение составляет имплицитный ассоциативный тест (ИАТ), предложенный Гринвальдом с соавторами в 1998 г. [Greenwald et al., 1998]. Задачей испытуемых в этом тесте является категоризация слов, относящихся к двум параллельным концепциям (например, «Я версус другие» и «тревожный версус спокойный»), используя две клавиши клавиатуры (например, левая стрелка для «Я» и «тревожный», а правая – для «другие» и «спокойный»). Предполагается, что испытуемый будет быстрее нажимать на клавиши, если присвоенное клавишам сочетание концепций соответствует его имплицитному представлению. Например, испытуемые, которые неосознанно считают себя тревожными, будут работать быстрее, если «Я» и «тревожный» присвоены одной и той же клавише. Имплицитный ассоциативный тест быстро приобрел популярность, и за прошедшие годы выполнено большое количество исследований с его применением. Показано, что имплицитные и эксплицитные (с помощью опросников) меры одних и тех же психологических конструктов слабо коррелируют друг с другом и первые лучше предсказывают неосознанное поведение, а вторые – сознательно контролируемое [Egloff, Schmukle, 2002; Oswald et al., 2013]. Это еще раз подтверждает относительную независимость указанных двух видов регуляции поведения.

В бытовом сознании, в художественной, да и в научной литературе часто можно встретить противопоставление сознательного и неосознанного, ярко представленное в известной повести Стивенсона в образах доктора Джекилла и мистера Хайда. Литература, в том числе научная, изобилует примерами, иллюстрирующими неспособность сознания проникнуть в подсознательные мотивы наших сознательных действий. В большом количестве

экспериментов показано, что имплицитный прайминг какой-либо информацией вызывает «перекос» в последующем сознательно контролируемом поведении и ни наличие, ни причина этого перекоса человеком не осознается. Все это ставит под сомнение эффективность сознательного контроля поведения. Может быть, нам только кажется, что мы способны сознательно планировать и осуществлять наши замыслы, что у нас есть свобода воли? Может быть, все это лишь иллюзия и на самом деле все решения тайно принимает мистер Хайд? Пожалуй, нигде темы сознания, неосознанного и их связи с активностью мозга не стоят так остро, как при обсуждении проблемы свободы воли, которую мы обсудим далее (см. гл. 6), но перед этим мы рассмотрим современные представления об организации базовых когнитивных процессов, таких как память, внимание, мышление, исполнительный контроль, и их связи с активностью мозга.

КОГНИТИВНЫЕ ПРОЦЕССЫ И МОЗГ

Эффективность организации когнитивных процессов в значительной степени определяет человеческий интеллект. Когнитивная психология и когнитивная нейробиология накопили большой объем эмпирических данных и теоретических концепций, связывающих базовые когнитивные процессы с активностью мозга. Рассмотрим эти данные конспективно, более подробно останавливаясь на отдельных фрагментах, представляющих, на мой взгляд, наибольший интерес в контексте расшифровки природы связи между ментальностью и мозгом. Начнем с памяти, которую можно считать компонентом практически любого когнитивного процесса, так как наиболее популярные в настоящее время теории как сенсорного восприятия, так и контроля поведения, такие как предсказывающее кодирование и байесовская модель, в явном виде включают использование предшествующей информации. Далее рассмотрим современные представления о хранении и извлечении информации в мозге.

5.1. Память

Исследование механизмов памяти имеет длинную историю и, естественно, не может быть детально рассмотрено в этом коротком разделе. Модель организации памяти, которая с небольшими изменениями сохранилась до настоящего времени, была предложена Ричардом Аткинсоном и Ричардом Шиффрином в 1968 г. [Atkinson, Shiffrin, 1968]. Согласно этой модели, память имеет три компонента: сенсорный регистратор, кратковременная (или рабочая) память и долговременная память. Отдельные компоненты и вся модель критиковались и уточнялись в последующих работах, но, во всяком случае, деление на кратковременную и долговременную память, которые используют разные механизмы и имеют разное представительство в мозге, никем сейчас не оспаривается.

Наиболее популярная модель организации кратковременной памяти была предложена Аланом Бэдди и Грэмом Хитчем [Baddeley, Hitch, 1974]; она содержит четыре компонента. Центральный исполнительный модуль выполняет функции контроля потоков информации между подчиненными системами, которые

включают фонологический (вербальная информация), зрительно-пространственный и эпизодический модули. Последний представляет собой временное хранилище информации, объединяющей сенсорную репрезентацию эпизода и релевантную информацию, извлеченную из долговременной памяти [Baddeley, 2011]. Фонологический модуль связан с активностью вербальных зон в левой височной коре, зрительно-пространственный модуль представлен активностью затылочной и теменной коры, эпизодический модуль связывают с билатеральной активностью лобных и височных долей и гиппокампа [Rudner et al., 2007]. Центральный исполнительный модуль до сих пор окутан тайной, хотя, предположительно, он должен быть связан с активностью лобных долей мозга [Baddeley, 2011].

Долговременную память принято делить на декларативную (эксплицитную) и процедурную (имплицитную). Процедурная память – это память о том, как пользоваться объектами и управлять телом в разных ситуациях. Считается, что у нас нет сознательного доступа к процедурной памяти и она связана преимущественно с активностью базальных ганглиев [Foerde, Poldrack, 2009]. Как отдельный вид памяти иногда выделяется эмоциональная память – память о событиях, вызвавших сильные эмоции. Она доступна для сознания, но может включать как декларативные, так и процедурные компоненты и сопровождаться неосознаваемыми вегетативными реакциями. Активность амигдалы связана с эмоциональными процессами и ее взаимодействие с гиппокампом и префронтальной корой участвует в кодировании эмоциональной информации [Buchanan, 2007]. К декларативной памяти относят все виды памяти, доступные для сознания. По современным представлениям, она кодируется гиппокампом и расположенными по соседству областями медиальной височной коры (entorhinal и perirhinal cortex), но в консолидации и хранении информации участвует большое количество и других областей мозга.

Принято делить эксплицитную память на три компонента – эпизодическая, автобиографическая и семантическая память. Эпизодическая память – это память на специфические события в определенный момент времени (ее называют еще контекстуальной памятью, так как воспоминание о событии неразрывно связано с контекстом ситуации) [Tulving, 2002]. Автобиографическая память – это разновидность эпизодической памяти, которая содержит лишь события, непосредственно относящиеся к собственной

персоне [Conway, Pleydell-Pearce, 2000]. Наконец, семантическая память – это сумма знаний о мире, включая, например, значения слов и всевозможных понятий. Если связь эпизодической памяти с гиппокампом не вызывает сомнений, то анатомическая основа семантической памяти до сих пор является предметом дискуссий.

По существующим данным, нейронные структуры, лежащие в основе эпизодической и семантической памяти, по крайней мере частично независимы друг от друга [Yee et al., 2013], хотя, согласно исследованиям функциональных связей, сети, участвующие в эпизодической и семантической памяти, в значительной степени перекрываются [Palacio, Cardenas, 2019]. В отличие от эпизодической памяти, которая ухудшается с возрастом, семантическая память дольше остается сохранной, что свидетельствует о разных механизмах двух видов памяти [Spaniol et al., 2006]. Этому выводу соответствуют и данные, показывающие, что повреждения в области медиальной височной коры (МВК) связаны с дефицитом эпизодической, но не семантической памяти, а повреждения в новой коре, в частности, в области височного полюса, связаны с семантической деменцией при относительной сохранности эпизодической памяти [Patterson et al., 2007].

Однако различие между эпизодической и семантической памятью не такое отчетливое, и их правильнее представлять не как независимые, а, скорее, как взаимодействующие системы. Воспоминание эпизода может сопровождаться извлечением семантической информации и наоборот [Greenberg, Verfaellie, 2010]. На нейрональном уровне считается, что нейроны МВК кодируют произвольные ассоциации, а нейроны новой коры – более упорядоченные, иерархически организованные ассоциации [Kumaran et al., 2016; Quian Quiroga, 2016]. По стандартной модели МВК кодирует эпизодическую память только в процессе обучения, но не после того, как она консолидировалась в неокортексе [Squire et al., 2015]. Согласно теории множественных следов, МВК сохраняет критическую роль в хранении эпизодической памяти и после консолидации [Moscovitch et al., 2016].

Рассмотрим сначала мозговые корреляты семантической памяти. В экспериментах по распознаванию смысла стимулов, предъявляемых разными способами (например, один и тот же смысл может быть представлен словом или картинкой), активация выявляется как в специфических для данного способа, так и в общих, независимых от способа областях мозга. Данные исследова-

ний, изучающих доступ к одной и той же семантической памяти при использовании разных стимулов и разных типов задач, свидетельствуют о том, что семантические концепции представлены в мозге в виде распределенных паттернов активности, которые могут быть по-разному задействованы как стимулами разных форматов, так и особенностями задания [Yee et al., 2013]. Извлечение знания о каком-то атрибуте объекта (например, цвет) сопровождается активацией тех же областей коры, которые активируются при сенсорном восприятии этого атрибута [Hsu et al., 2011]. То же отмечено в отношении моторных атрибутов [Chao, Martin, 2000].

В целом данные нейровизуализации, нейропсихологии и исследования эффектов транскраниальной стимуляции мозга показывают, что семантические знания об объектах построены вокруг их сенсорных и моторных атрибутов и что эти атрибуты хранятся в соответствующих сенсорных и моторных областях мозга [Yee et al., 2013]. Где хранятся более абстрактные знания – не до конца понятно. По мнению некоторых исследователей, семантические знания разного уровня абстракции хранятся в разных регионах и более высокие уровни абстракции больше связаны с лингвистическими зонами коры [Hart, Kraut, 2007]. Если семантические знания организованы в виде мультимодальной, высоко интерактивной, распределенной системы, то как мозг «выбирает» паттерн активации, соответствующий данному объекту, каналу восприятия и требованиям поставленной задачи? Предполагается, что функции координатора в такого рода задачах выполняют некоторые области префронтальной коры, в частности, левая нижняя лобная извилина [Mechelli et al., 2004]. Конкретный механизм этой координации остается предметом спекуляций.

Модель распределенного сенсомоторного представительства в мозге знания об определенном объекте усматривает существование различий, связанных с опытом приобретения указанного знания. Это может относиться как к индивидуальным различиям между людьми, так и к различиям между состояниями одного и того же человека. Действительно, присутствие таких различий экспериментально показано [Kiefer et al., 2007]. То есть, несмотря на согласие двух людей в отношении свойств какого-либо объекта, его представительство в активности мозга каждого из них будет различным. Интересно, что это распространяется на семантические концепции на всех уровнях абстракции. Например, тенденция ассоциировать добро с правым, а зло с левым проявляется у

праворуких в активации соответствующей области сенсомоторной коры [Casasanto, 2009]. Как оказалось, паттерн может измениться на противоположный, если доминирование правой руки нарушено из-за инсульта или вследствие экспериментальной манипуляции [Casasanto, Chrysikou, 2011]. То есть, если, например, правша и левша рассуждают о природе добра и зла, паттерны активации мозга у них могут быть различны, хотя смысл понятий добра и зла одинаков. Возникает вопрос: в какой степени семантическая память, представленная в активности мозга, соответствует знанию, которым мы пользуемся в собственных мыслях и при общении друг с другом?

Другой подход к моделированию структуры семантической памяти и знания вообще развивается в рамках коннекционизма [Jones et al., 2014]. Коннекционистские модели предполагают наличие знаний в виде сети взаимосвязанных единиц, конструкция которой описывается в терминах рассмотренной выше коннекционистской теории разума и нейронных сетей. Наиболее популярны сейчас распределенные коннекционистские модели структуры знания используют идеи и методы, описанные выше при рассмотрении нейролингвистики. В этих моделях смысл определенного понятия извлекается на основе вероятностного анализа контекстуальных связей. Очень часто в таких моделях используется байесовский подход, при котором латентный смысл представлен в виде вероятностного распределения в семействе близких тем [Griffiths et al., 2007]. Распределенные коннекционистские модели семантической памяти иногда критикуют за то, что, в отличие от рассмотренных выше моделей, основанных на учете сенсорного компонента и контекста ситуации, они придают слишком большой вес лингвистическому компоненту знания. Компьютерное моделирование показывает, что модели, которые одновременно используют как сенсорный, так и лингвистический компоненты, лучше аппроксимируют смысл понятий [Baroni et al., 2010].

Рассмотрим теперь данные о кодировании эпизодической информации в активности мозга. Как мы уже отмечали ранее, процесс восприятия, например, зрительных образов описывается сейчас в терминах предсказывающего кодирования (predictive coding). Согласно этой концепции, сенсорное восприятие – активный процесс, при котором поступающая из органов чувств информация сравнивается с содержащейся в памяти информацией. На ее основе происходит категоризация и узнавание воспринимаемых объек-

тов внешнего мира. Это означает, что процесс сенсорного восприятия нужно рассматривать в неразрывной связи с процессами кодирования и извлечения информации из памяти. Для зрительного восприятия анатомическую основу такого взаимодействия обеспечивают мощные связи между структурами вентрального зрительного тракта (ВЗТ) и МВК [Lavenex, Amaral, 2000]. И восприятие, и память связаны с построением смысла путем отбора небольшого количества информации и использования умозаключений и абстракций. Интересно сравнить, как смысл воспринимаемого отражается в активности нейронов ВЗТ и МВК.

Как мы уже видели раньше, зрительное восприятие связано с категоризацией воспринимаемых образов, которая сопровождается абстрагированием от большого количества деталей и индивидуальных различий и извлечением прототипов соответствующих категорий. То же происходит и при запоминании эпизодической информации – мы запоминаем субъективный смысл эпизодов, абстрагируясь от деталей [Koriat et al., 2000]. Сходство иллюзий восприятия и памяти еще больше подчеркивает связь между ними. При восприятии мы постоянно делаем предположения о том, что видим, и это иногда проявляется в зрительных иллюзиях, основанных на ложных интерпретациях [Eagleman, 2001]. В памяти аналогичные процессы вызывают возникновение ложных воспоминаний [Baddeley et al., 2009].

В зрительных корковых областях отражение атрибутов зрительных стимулов в активности нейронов топографически организовано и распределено, что идеально подходит для извлечения субъективного смысла и семантической категоризации. Начало ответа на верхних уровнях ВЗТ (нижняя височная извилина, НВИ) происходит примерно через 100 мс после предъявления стимула, что намного раньше, чем ответы, получаемые в МВК (примерно 300 мс) [Quiari Quiroga, 2016]. Если нейроны первичной зрительной коры (область V1) отвечают на локальные атрибуты стимула (ориентация, цвет, форма), то нейроны НВИ отвечают на принадлежность стимула к той или иной категории (лица, объекты, части тела) [Gross, 2008].

В свое время высказывалось предположение, что последовательная обработка зрительной информации в иерархически организованных слоях зрительного анализатора может заканчиваться на клетках, активность которых репрезентует индивидуальные стимулы специфическим и неизменным образом (так называемая

клетка бабушки (grandmother cell), которая отвечает только на образ бабушки) [Barlow, 1994]. Однако степень селективности и инвариантности нейронов НВИ относительно ограничена и очень далека от представления о бабушкиной клетке. Фактически эти нейроны демонстрируют распределенное представительство и реагируют на большое количество разных стимулов, в основном в рамках предпочитаемой категории стимулов [Tsao et al., 2006]. Более того, информация об идентичности конкретного стимула кодируется модуляцией силы ответа на различные стимулы в популяции нейронов. Другими словами, в то время как информация о категории стимула является явной на уровне отдельной клетки (например, по срабатыванию нейрона мы можем определить, был ли стимул лицом или нет), информация об идентичности конкретного стимула в категории имеет неявный вид (т. е. по срабатыванию нейрона мы не можем сказать, чье это лицо) и распределяется в популяции нейронов [Quian Quiroga, Kreiman, 2010].

В целом эти результаты подтверждают идею о том, что нейроны вдоль вентрального зрительного пути (и префронтальной коры) участвуют в извлечении субъективного значения предъявляемого стимула. Более того, их активация связана именно с сознательным восприятием стимула. Например, если испытуемым на короткий промежуток времени предъявляли зашумленное изображение и просили определить, является ли оно изображением лица или любого другого объекта (уровень зашумления подбирался так, чтобы испытуемые могли правильно определить категорию объекта примерно в 50 % случаев), то большая амплитуда компонента N170 вызванного потенциала регистрировалась в тех случаях, когда изображение распознавалось как лицо [Navajas et al., 2013]. Таким образом, перцептивная осведомленность о том, к какой категории принадлежит стимул, появляется уже через 170 мс после его предъявления и связана с активностью НВИ, а узнавание конкретного лица происходит через 300 мс и связано с активностью МВК, так как требует извлечения разнородной информации, относящейся к этой персоне [Quian Quiroga, 2020].

Кажется странным, что активность нейронов НВИ позволяет определить лишь то, к какой категории относится объект. Если извлечение смысла и распознавание конкретных объектов (например, лицо конкретного человека) происходит на основе информации, приходящей из зрительного анализатора, то активность нейронов НВИ должна в той или иной форме эту информацию со-

держат. Действительно, в одной из последних работ с регистрацией активности нейронов НВИ у обезьян, которым показывали изображения человеческих лиц, обнаружено, что это так и есть [Chang, Tsao, 2017].

Показано, что можно с достаточной точностью определить идентичность лица по активности (средняя частота импульсации за 50–300 мс от начала предъявления стимула) лишь 200 нейронов в НВИ, если каждое лицо представлено в виде точки в 50-мерном пространстве. Для формирования этого пространства каждое лицо представлялось в виде показателей формы (координаты ключевых точек, таких как линия волос, углы носа, глаз, губ и т. д.) и яркости (значение яркости каждого пикселя в изображении), из которых методом главных компонент извлекались 50 размерностей. Уникальность каждого лица определялась его положением в этом 50-мерном пространстве. Оказалось, что каждый из зарегистрированных нейронов НВИ преимущественно отвечал на какую-то одну размерность из 50 и по сумме ответов 200 нейронов можно было вычислить уникальное сочетание размерностей для каждого лица. Максимальная точность предсказания идентичности лица по активности нейронов НВИ была 75 % при общем количестве лиц, равном 40, и уменьшалась при увеличении количества лиц. Каждый нейрон отвечал на «свою» размерность не по принципу «все или ничего» – сила ответа линейно снижалась для размерностей, отклоняющихся от целевой на все больший угол, и падала до уровня случайной активности в плоскости ортогональной предпочитаемой размерности.

Эти данные показывают, что информация, достаточная для идентификации лица, содержится в активности нейронов НВИ, хотя и ничего не говорят о том, как эта информация извлекается в мозге. Как, например, мозг извлекает информацию о координатах ключевых точек лица? По предположению авторов [Ibid], это достигается с помощью архитектуры, аналогичной иерархической глубокой сети с прямой связью (hierarchical feedforward deep-network), но это пока лишь гипотеза. Важно подчеркнуть, что активность нейронов НВИ кодирует не идентичность лица (например, его похожесть на какой-то прототип), а выраженность определенных характеристик. «Цель» нейронов НВИ состоит в том, чтобы настроить систему координат для измерения лиц, а не в том, чтобы идентифицировать лица. Ниже мы увидим, что это не так для нейронов МВК, связанной с процессами эпизодической памяти.

Критически связанная с декларативной памятью МВК включает гиппокамп и прилежащие области височной доли, организованные иерархически. В нижней части иерархии находятся периринальная (perirhinal) и парагиппокамповая (parahippocampal) кора, которые получают информацию из корковых сенсорных областей и посылают ее в энторинальную (entorhinal) кору, откуда она поступает в гиппокамп [Lavenex, Amaral, 2000]. Амигдала, тоже часть медиальной височной коры, имеет прямые связи с сенсорными областями и другими структурами МВК. Последняя связана с возникновением некоторых форм эпилепсии и является поэтому типичной областью, предназначенной для имплантации внутричерепных электродов при нейрохирургических операциях, что дает исследователям уникальную возможность изучать процессы памяти с помощью внутричерепных записей у людей. Накопленные в этих исследованиях данные указывают на то, что кодирование эпизодической памяти в мозге человека может происходить не так, как у животных.

Изучение активности отдельных нейронов в процессе обучения показало, что у грызунов эпизодическая информация (например, информация о двух разных лабиринтах) кодируется в неперекрывающихся нейронных ансамблях в зубчатой извилине гиппокампа, что обеспечивает, как считается, отсутствие интерференции между разными эпизодами в памяти [Knierim, Neunuebel, 2016]. Это явление назвали разделением паттернов (ПП, pattern separation). Оно наблюдается не только для различных сцен, но и для любого изменения условий эксперимента [Moser et al., 2017]. Сходные результаты получены и у обезьян [Gulli et al., 2020].

У людей косвенные данные о наличии ПП получены с помощью фМРТ [Leal, Yassa, 2018]. Однако регистрация нейронной активности у людей в процессе нейрохирургических операций выявила другую картину. Было показано существование в МВК так называемых концептуальных клеток (concept cells), которые отвечают на определенные концепции (например, знакомый человек), а не на отдельные атрибуты [Quian Quiroga et al., 2005]. Ответы этих клеток отличаются высокой специфичностью (например, только на этого, но не на какого-либо другого человека) и высокой инвариантностью (на все очень разные атрибуты этого человека, включая внешний вид, звук голоса, написанное или произнесенное имя) [Quian Quiroga, 2012]. Концептуальные клетки

найлены только у людей, но не у животных. У людей выявлено также наличие нейронов, кодирующих ассоциации между концепциями. Такие нейроны могли отвечать, например, на изображения актера и самолета, если актер снимался в фильме под названием «Аэроплан» [Rey et al., 2020].

Изучение нейронной репрезентации эпизодической памяти у людей обнаруживает отсутствие разделения паттернов. По этим данным, индивидуальные концепции представлены конкретными инвариантными энграммами, а связь между концепциями закодирована через частично перекрывающиеся нейронные ансамбли [Quian Quiroga, 2020]. По мнению автора, данные о наличии РП у людей, полученные с помощью фМРТ, объясняются методическими особенностями использованного эксперимента. Кирога указывает, что кодирование с помощью РП может в конечном счете привести к нехватке нейронов в гиппокампе для репрезентации нарастающего объема эпизодической памяти. Это может не создавать проблемы у животных, но у человека спровоцировало бы «комбинаторный взрыв». Следовательно, эволюция мозга человека шла не просто по пути увеличения его объема, но и по пути изменения механизма кодирования информации [Ibid].

Воспроизведение эпизодической памяти, таким образом, является не просто «проигрыванием» записи, а, скорее, реконструкцией на основе сохраненной схемы коактивации энграмм, поддерживаемой более детальной активацией в неокортексе. Такая организация памяти облегчает генерализацию, создание абстрактных концепций, креативность и воображение, а также использование информации в новых контекстах. Ниже приведены основные характеристики «концептуальных» нейронов МВК, описанные в обзоре Кироги [Quian Quiroga, 2016].

(1) Концептуальные нейроны показывают высокую избирательность ответов, которая увеличивается с 2 до 5 % (от общего количества предъявляемых стимулов) по мере продвижения от парагиппокампальной коры к гиппокампу [Quian Quiroga, 2012]. Такая высокая избирательность (реакция на конкретных людей, а не на более широкие категории, такие как лица, мужчины и т. д.) соответствует предполагаемой роли этих нейронов в формировании и воспроизведении эпизодических воспоминаний, поскольку мы склонны запоминать эпизоды с участием конкретных людей и мест. Важно подчеркнуть, однако, что, несмотря на такую высокую избирательность, эти нейроны не следует считать «бабушки-

ными клетками», потому что они все-таки отвечают не исключительно на одну концепцию и могут быть обучены отвечать на новые ассоциации [Ison et al., 2015].

(2) Концептуальные нейроны демонстрируют очень высокий уровень визуальной и мультимодальной инвариантности, поскольку они выборочно отвечают на различные изображения конкретных людей, мест или объектов и даже на их написанные и произнесенные имена. Степень инвариантности также увеличивается в иерархии МВК – 52 % нейронов показывают визуальную инвариантность в парагиппокампальной коре, по сравнению с 85 % в гиппокампе, и ни один нейрон в парагиппокампальной коре не проявляет реакции на текст или звуковое представление, в то время как примерно половина нейронов в гиппокампе это демонстрирует [Quian Quiroga, 2012]. Следовательно, концептуальные нейроны не отвечают на специфические визуальные черты, так как их ответы на разные изображения одного и того же человека аналогичны. Это согласуется с тем фактом, что мы склонны помнить отношения между концепциями и забываем конкретные детали [Bartlett, 1932].

(3) Концептуальные нейроны представляют эксплицитную репрезентацию «смысла» стимула. По активации этих нейронов можно предсказать, какую специфическую концепцию (из множества других) видит или даже воображает испытуемый [Cerf et al., 2010], но невозможно предсказать, какое конкретное изображение этой концепции он видит. Надежность предсказания линейно увеличивается с увеличением количества регистрируемых нейронов, показывая, что каждый нейрон кодирует дополнительную информацию о данной концепции [Quian Quiroga et al., 2007]. В противоположность этому, из активности отдельных нейронов НВИ невозможно предсказать, какой стимул воспринимается (потенциально можно сказать, лицо это или нет, но невозможно сказать, чье это лицо), и надежность предсказания увеличивается нелинейно с увеличением количества регистрируемых нейронов [Kreiman et al., 2006].

(4) Активность концептуальных нейронов связана с сознательным восприятием стимула по принципу «все или ничего», что показано в экспериментах с кратковременным предъявлением и маскированием стимула. То есть нейроны активировались только в тех случаях, если стимулы сознательно узнавались, и молчали в остальных случаях, хотя картинка и все физические параметры

стимула оставались одинаковыми [Quian Quiroga et al., 2008]. Более того, при предъявлении преобразованных изображений, в которых совмещались черты двух известных актрис, активировались нейроны, отвечающие на концепцию либо одной, либо другой актрисы, в соответствии с тем, какую из них испытуемый «узнавал» [Quian Quiroga et al., 2014]. Тем самым подчеркивается, что эти нейроны реагируют на субъективное восприятие (то, что, по мнению испытуемого, он видел), независимо от характеристик стимула, запускающего его.

(5) Концептуальные нейроны могут отвечать на более чем одну концепцию, но в таком случае концепции, на которые они отвечают, обычно как-то связаны друг с другом (например, два актера, которые играли в одном и том же фильме) [Quian Quiroga, Kreiman, 2010]. Предполагается, что нейроны, отвечающие на более чем одну концепцию, кодируют ассоциации между этими концепциями. Показано, что ассоциации не являются жесткими и могут изменяться в процессе эксперимента [Ison et al., 2015], в отличие от жесткого категориального семантического кодирования в новой коре.

(6) Концептуальные нейроны имеют тенденцию отвечать на персонально релевантные концепции, т. е. такие, которые достаточно важны для формирования новой эпизодической памяти. Так, наибольшая вероятность ответов показана на изображения экспериментаторов, с которыми испытуемые хорошо познакомились за время исследований, и членов семьи, затем следовали фотографии знаменитостей и лишь потом – остальных людей [Viskontas et al., 2009], что говорит о быстром и преимущественном кодировании новой и значимой информации.

(7) Ответы концептуальных нейронов МВК регистрируются уже при первом предъявлении стимула и снижаются при повторных предъявлениях, но остаются выше фонового уровня [Pedreira et al., 2010], что свидетельствует в пользу гипотезы, согласно которой МВК продолжает сохранять координирующую роль и после консолидации следа памяти в новой коре.

(8) Репрезентация концепций в концептуальных нейронах МВК не имеет топографического характера – близко расположенные нейроны не обязательно отвечают на похожие или связанные концепции [Quian Quiroga et al., 2007]. Это может способствовать быстрому кодированию новых ассоциаций между разрозненными концепциями.

(9) Латентный период ответа концептуальных нейронов МВК относительно велик – около 300 мс, что гораздо больше, чем для нейронов, участвующих в зрительном восприятии [Quiñ Quiroga et al., 2009]. Такие поздние ответы предполагают, что нейроны МВК активируются после относительно длительной обработки в других корковых областях, вероятно, для извлечения смысла стимула, который передается в МВК для запоминания. То, что эти нейроны не участвуют в зрительной перцепции, также подтверждается тем фактом, что они срабатывают, когда субъекты представляют концепцию после определенного сигнала, думают о ней для решения задачи, или просто вспоминают в отсутствие какой-либо внешней стимуляции [Quiñ Quiroga, 2020].

(10) Возникновению ответа концептуальных клеток предшествует отклонение потенциала локального поля (ПЛП, local field potential) после распознавания стимула [Rey et al., 2014]. В экспериментах с предъявлением стимулов на пороге сознательного восприятия показано, что глобальный (во всех частях МВК) ПЛП-ответ в частотном диапазоне тэта-осцилляций наблюдается только в том случае, если имело место сознательное распознавание стимула [Quiñ Quiroga et al., 2008]. Этот ответ не был селективным (т. е. возникал в ответ на любое изображение, если оно было сознательно опознано) и, вероятно, запускался проекциями из верхних уровней зрительного анализатора в НВИ. Предполагается, что ПЛП в тэта-диапазоне открывает временное окно для синхронизации информации из разных сенсорных модальностей и конвергенции ее на клетках МВК. В том же временном диапазоне наблюдался всплеск гамма-активности, который, вероятно, отражает кодирование определенной концепции в клетках МВК [Quiñ Quiroga, 2020].

Подводя итог, скажем, что концептуальные нейроны МВК связаны с инвариантным и явным представлением смысла сознательно воспринимаемых стимулов и критически участвуют в формировании и воспроизведении эпизодических воспоминаний, кодируя ассоциации между концептами [Quiñ Quiroga, 2012]. Каждая концепция кодируется в МВК ансамблем концептуальных нейронов, которые срабатывают вместе всякий раз, когда субъект осознает конкретную концепцию [Waydo et al., 2006]. В отличие от сенсорных модальностей, где трудно выделить ансамбль нейронов, так как каждый нейрон отвечает на множество стимулов (хотя и с разной силой), концептуальные нейроны МВК отвечают на не-

большое количество концепций, и в этом случае концепции как-то связаны друг с другом. Связанные концепции кодируются частично перекрывающимися ансамблями нейронов, что дает простой и эффективный механизм быстрого кодирования ассоциаций [Quiñ Quiroga, 2020]. Этот механизм основан на представлении относительно небольшого числа очень специфических концепций – тех, которые достаточно актуальны, чтобы их запомнить, и ассоциаций, созданных между ними. Это составляет только скелет эпизодических воспоминаний, который обогащен указателями на области коры, кодирующие важные детали следа памяти, а также связанной семантической информации, основанной на предыдущих абстракциях и категоризациях.

Таким образом, поток сознания включает в себя активацию последовательности ансамблей концептуальных клеток гиппокампа и связанных с ними корковых репрезентаций. Как описано ранее, нейроны МВК демонстрируют мультимодальную инвариантность, отвечая на совершенно разные изображения одного и того же человека, а также на его написанное или произнесенное имя. Можно думать, что при первых предъявлениях разные изображения человека могут вызвать срабатывание одного ансамбля нейронов МВК, а его имя – другого. Поскольку эти два события – видение лица и слушание (или воспоминание) имени – будут иметь тенденцию происходить вместе, получится достаточно большое перекрытие между нейронными ансамблями и в дальнейшем они сольются в единый ансамбль, который затем может быть активирован изображением или именем. Предполагается, что при встрече с незнакомым объектом активируется большое количество нейронов МВК, которое при повторных встречах снижается до устойчивого нейронного ансамбля. При отсутствии повторных встреч ансамбль исчезает.

Как мы видели ранее, в иерархии областей мозга, участвующих в обработке зрительных стимулов, по мере продвижения к высшим уровням увеличивается инвариантность и избирательность по отношению к сложным характеристикам стимула. Визуальные области высокого уровня проецируются в МВК, где этот процесс продолжается дальше и достигает своей вершины в гиппокампе, нейроны которого избирательно реагируют на разные изображения и даже имена конкретных людей или объектов. Хотя общий процесс повышения инвариантности и избирательности

одинаков для зрительного анализатора и МВК, существуют серьезные различия в типах кодирования этими областями для выполнения различных функций. Зрительные области имеют топографически организованное представительство, тогда как гиппокамп содержит случайные связи между нейронами и действует как автоассоциативная сеть, которая способна быстро создавать связи между любыми заданными представлениями. Области высокого уровня вдоль вентрального зрительного тракта имеют относительно распределенное кодирование стимулов, которое идеально подходит для надежного распознавания и хранения семантической информации, тогда как представление в МВК намного реже и инвариантно, что идеально для эпизодической памяти.

Таким образом, в то время как зрительная кора высокого уровня демонстрирует начальную дифференциальную активацию, разделяющую различные стимулы примерно через 100 мс после их предъявления, процесс извлечения смысла, приписываемого стимулам, происходит в течение последующих 200 мс, после чего активация передается нейронам МВК для эксплицитного кодирования концепций и запоминания и воспроизведения ассоциаций между ними (а также связи с корковым представительство). Активация ансамблей концептуальных нейронов сопровождается двумя процессами, происходящими параллельно: с одной стороны, последовательная активация ансамблей является отражением потока сознания между связанными концепциями и, с другой стороны, указатели от этих ансамблей к связанным корковым репрезентациям обогащают эти воспоминания деталями (например, чертами лица) и связанной семантической информацией.

Как же модель частично перекрывающихся нейронных ансамблей в МВК может объяснить эпизодическую память? Кироба описывает два гипотетических варианта. Во-первых, ассоциации в МВК могут указывать на более детальное представительство в коре, и воспоминание эпизодов определяется взаимодействием между схематической активацией в МВК и обогащенной деталями активацией в различных сенсорных зонах коры. Вторая возможность состоит в том, что эпизодическая память, так же как зрительное восприятие, – это конструирование на основе лишь нескольких ассоциаций. Мы помним лишь, что встретили определенного человека в определенном месте, или обсуждали определенную тему, а остальное мы выводим на основе предположений. Например, ут-

верждается, что мы на самом деле не помним время так, как нам кажется, и что идея времени, которая является ключевой в эпизодической памяти, может быть основана на предположениях [Friedman, 2004]. По мнению Кирогги, ответом может быть комбинация обеих возможностей, а именно: то, что мы называем эпизодической памятью, основывается на нескольких ассоциациях между концепциями в МВК и несколькими указателями на детали, хранящиеся в коре головного мозга, а остальное – ложные воспоминания, которые мы додумываем [Quian Quiroga, 2016].

Модель эпизодической памяти Кирогги в значительной степени спекулятивна, как признает и сам автор, и ключевые ее аспекты, например, взаимодействие между сенсорными зонами коры, эпизодической и семантической памятью и содержанием потока сознания, вряд ли возможно проверить в обозримом будущем. Тем не менее она очень интересна, так как прописывает гипотетический механизм извлечения субъективного смысла из потока сенсорной информации. На примере зрительного анализатора можно видеть, что процесс анализа зрительной информации в мозге по мере продвижения от низших уровней к высшим состоит в последовательно сужающейся семантической категоризации. Прагматический смысл этого очевиден. Например, если в течение 100 мс зрительный анализатор способен определить, что воспринимаемый объект является лицом или зданием, резко снижается количество возможных вариантов развития событий и организации поведения. Остается неясным, как конкретно данная категоризация осуществляется в мозге.

На входе, в первичной зрительной коре, нейроны отвечают на зрительную стимуляцию распределенной активацией, которая вероятно связана с простыми свойствами зрительного объекта (яркость, цвет, движение, форма). По наиболее популярной сейчас байесовской модели, сверху спускается гипотеза о том, какого рода объект мы можем увидеть в текущей ситуации, и на основе этой гипотезы рассчитывается комбинация визуальных свойств, которая сравнивается с входной комбинацией. В результате сравнения гипотеза либо принимается и сигнал передается в верхние отделы для проверки уточняющих гипотез, либо отвергается и заменяется альтернативной гипотезой. Как мы уже отмечали выше, эта модель пока остается лишь теоретической концепцией и нет понимания того, как конкретно все описанные здесь процессы могут происходить в мозге.

Для того чтобы генерировать гипотезы, мозг должен извлекать информацию о сходных ситуациях из семантической и/или эпизодической памяти. Например, если мы стоим на берегу моря и, всматриваясь вдаль, смутно видим какой-то объект на горизонте, вряд ли мы ожидаем, что этот объект окажется человеческим лицом. Скорее всего, это должен быть корабль. Соответственно, количество возможных категорий резко сужается просто исходя из контекста ситуации. Но знание о том, в какой ситуации мы находимся и какие объекты обычно можно встретить в подобных ситуациях, извлекается из эпизодической и семантической памяти. Значит, уже самые первые этапы сенсорного восприятия неразрывно связаны с извлечением информации из памяти. Например, знание о том, в какой ситуации я нахожусь, извлекается из памяти о том, где я находился и что делал в предшествующие моменты времени.

Так или иначе, на выходе вентрального зрительного тракта через 100 мс кодируется лишь принадлежность воспринимаемого объекта к какой-либо семантической категории (например, лица). Далее, по теории Кирогги, в течение последующих 200 мс происходит определение субъективного смысла воспринимаемой информации (например, что это не просто лицо, а лицо моего соседа). Процесс извлечения смысла заканчивается активацией ансамбля нейронов МВК, соответствующего концепции моего соседа. Однако как происходит извлечение смысла? Можно думать, что данный процесс тоже градуальный (например, сначала определяется, что это не просто лицо, а лицо мужчины, затем – что это мужчина с голубыми глазами и светлыми волосами и т. д.). На каждом этапе, опять же, требуется извлечение информации из семантической памяти. По гипотезе Кирогги, эти стадии происходят уже не в зрительном анализаторе, а на разных уровнях МВК плюс в разных областях новой коры.

Кирогга подчеркивает различия в представлении смысла в активности НВИ и МВК. В нейронах верхнего слоя вентрального зрительного тракта принадлежность к той или иной семантической категории закодирована в виде распределенной активности, имеющей вероятностный характер. То есть каждый нейрон отвечает на практически все категории и сила его ответа (выраженная в виде частоты импульсации) распределена между категориями по закону Гаусса. В МВК лишь небольшая часть нейронов отвечает на какую-либо концепцию и ответы эти происходят по принципу

«все или ничего». Что же происходит за 200 мс, что радикально меняет характер ответа нейронов? По терминологии Кироги, происходит извлечение смысла.

Если по активности нейронов НВИ лишь с какой-то долей вероятности можно определить принадлежность объекта к той или иной категории, то по активности нейронов МВК можно с достаточной долей уверенности идентифицировать объект. Где и как происходит извлечение смысла, Кирога не уточняет, ограничиваясь утверждением, что это происходит в результате обработки дополнительной информации из разных сенсорных модальностей. Предполагается, что в этом как-то участвуют тэта- и гамма-осцилляции, которые синхронизируют активность МВК с активностью других областей [Quian Quiroga, 2016].

Модель Кироги, как и все такого рода модели, которые строятся на базе эмпирических данных, неизбежно фрагментарна и в принципе не способна прописать конкретные механизмы, с помощью которых может происходить категоризация и извлечение смысла. Современные теоретические модели в сфере искусственного интеллекта в основном рассматривают различные варианты нейронных сетей и машинного обучения. Несмотря на многочисленные попытки адаптировать эти модели к реалиям конструкции мозга, до создания правдоподобной и потенциально тестируемой модели пока еще очень далеко.

Отметим в заключение, что модель Кироги – лишь теория, к тому же в значительной степени умозрительная. В этой модели активность концептуальных нейронов МВК может обеспечивать лишь «костяк», или набросок памяти эпизода. Ее полнота должна, согласно модели, зависеть от поддерживающей активности многих других областей коры, которые, можно допустить, как раз и участвовали в извлечении смысла при формировании эпизодической памяти. Важно поэтому рассмотреть, как происходит воспроизведение памяти.

Информация, хранящаяся в памяти, может использоваться только при ее воспроизведении. Как мы видели выше, след памяти, или энграмма, имеет распределенный характер, охватывая разные области коры и подкорки [Wheeler et al., 2013]. Это относится как к эпизодической, так и к семантической памяти. В распределенной сети каждый регион может нести уникальную информацию о закодированном эпизоде или объекте (например, сенсорная, аффективная, пространственная информация). При воспроизведе-

дении происходит реактивация нейронного ансамбля, хранящего энграмму. Многие исследования показывают, что нейронные ансамбли, активируемые при воспроизведении следа памяти, в значительной степени совпадают с нейронными ансамблями, активируемыми при запоминании [Richards, Frankland, 2013; Palacio, Cardenas, 2019]. Как из этой распределенной достаточно широко-масштабной активности возникает смысл воспоминания? Извлекаемый смысл не сводится к сумме сенсорной, аффективной или пространственной информации. В случае эпизодической памяти в зависимости от повода и цели воспоминания возникает либо цельная картина эпизода, либо из нее избирательно извлекается нужная деталь.

Сходным образом в случае семантической памяти возникает либо цельное представление об объекте или понятии, либо извлекаются определенные его свойства. Одна из теорий вообще отвергает существование семантической памяти. По данной теории есть только эпизодическая память, а семантика – это эмерджентная сущность, возникающая при извлечении и абстрагировании следов эпизодической памяти [Kwantes, 2005]. Результаты недавнего фМРТ-исследования свидетельствуют о том, что извлечение информации из семантической и эпизодической памяти поддерживается общими нейронными процессами [Vatansever et al., 2021]. В целом как коннекционистские, так и распределенные модели семантической памяти совместимы с представлением, согласно которому семантическая структура памяти является эмерджентной сущностью [Jones et al., 2014].

Для сохранения информации в долговременной памяти она должна пройти через кратковременную память, объем которой является, таким образом, фактором, лимитирующим скорость запоминания [Nikolić, Singer, 2007]. Консолидация (упрочение), или исчезновение (забывание) следов памяти зависит от многих процессов, включая субъективно оцениваемую важность информации и количество повторных ее предъявлений или извлечений [Dudai, 2003]. По современным представлениям, сон играет важную роль в консолидации долговременной памяти [Ruch et al., 2012]. Считается, что информация кодируется в мозге в виде энграммы, хотя до сих пор нет единого мнения по поводу биологического субстрата энграммы. Долговременная память, в отличие от кратковременной, зависит от синтеза новых белков, участвующих в механизмах синаптической передачи, что обеспечивается метилированием и

деметилованием соответствующих участков генома нейронов [Costa-Mattioli, Sonenberg, 2008]. Предполагается, что паттерн индуцированных и репрессированных генов в нейронах головного мозга после интенсивного обучающего события обеспечивает молекулярную основу для долговременной памяти о событии [Duke et al., 2017].

Идея о том, что долговременная память кодируется в виде структурных изменений в синапсах, облегчающих последующее прохождение возбуждения через синапс, была предложена еще Рамон-и-Кахалем [Ramon y Cajal, 1894]. Позднее Дональд Хебб развил эту идею в теорию, согласно которой основой долговременной памяти являются нейронные ансамбли, кодирующие информацию в виде модификации синаптических связей между нейронами ансамбля [Hebb, 1949]. Современные теории опираются на теорию Хебба. Использование современных методов, таких как оптогенетическая стимуляция специфических предварительно помеченных нейронов в сочетании с фармакологическими и гено-инженерными способами, позволяющими транзитивно блокировать или стимулировать отдельные стадии синтеза белков и считывания информации с ДНК, позволило манипулировать компонентами энграмм в специфических областях мозга и показать, что клеточные ансамбли, помеченные с помощью этих методов, необходимы и достаточны для воспроизведения следа памяти [Tonegawa et al., 2015].

Эксперименты, естественно, являются инвазивными и проводятся, как правило, на крысах с использованием простых моделей обучения с отрицательным или положительным подкреплением. Соответствующие виды памяти можно отнести к эмоциональной эпизодической памяти. Выявляемые в этих экспериментах клеточные ансамбли включают помимо нейронов гиппокампа нейроны областей мозга, ответственных за отрицательное (amygdala) и положительное (nucleus accumbens) подкрепление [Ibid]. Несмотря на специфичность используемых методов, они не могут позволить ответить на некоторые вопросы.

Например, если блокирование определенных нейронов включает соответствующую функцию, значит, эти нейроны необходимы для выполнения данной функции. Однако если стимуляция нейронов восстанавливает указанную функцию, это, строго говоря, не доказывает, что они достаточны для выполнения данной функции, так как стимуляция нейронов может произвести боль-

шое количество эффектов, не регистрируемых в эксперименте, и, таким образом, быть только косвенно ответственной за восстановление функции. Другой вопрос касается содержания кодируемой информации. Пока нет никакой возможности извлечь содержание информации из паттерна изменений в ансамбле нейронов. Если связи между нейронами ансамбля являются субстратом хранения информации, представляется необходимым изучить структуру и функции данных связей. Эта задача пока не решена даже для таких простых моделей памяти, как модель пассивного избегания у крыс [Ibid].

5.2. Внимание

Внимание можно считать еще одним управляющим механизмом, позволяющим выбрать из множества объектов или мыслей то, что является наиболее релевантным в контексте внешней, или ментальной ситуации, и игнорировать все остальное [James, 1890]. Внимание способствует концентрации когнитивных ресурсов для достижения конкретной цели и во многом определяет содержание нашего сознания. Изучение внимания имеет длинную историю, в течение которой накоплено большое количество эмпирических фактов и сформулировано множество теорий. Подавляющее большинство эмпирических данных, на которых строились теории внимания, было получено в поведенческих экспериментах на здоровых испытуемых или пациентах с повреждениями мозга. Внимание участвует в избирательной направленности нашей умственной жизни. Природа этой избирательности – один из основных моментов разногласий между существующими теориями внимания. Некоторые теории рассматривают избирательность внимания как результат ограниченной способности мозга обрабатывать сложные свойства множественных стимулов или направлений мысли. Другие не связывают избирательность внимания с ограниченными возможностями мозга, а объясняют ее необходимостью поддержания единого образа действий или мыслей и взвешиванием сенсорной информации в соответствии с ожидаемой [Mole, 2017].

Наиболее ранней и влиятельной теорией первой категории была теория «бутылочного горлышка», предложенная Дональдом Бродбентом [Broadbent, 1958]. По этой теории внимание подобно бутылочному горлышку, соединяющему две разные сенсорные

системы. До него лежит система, способная одновременно обрабатывать большое количество простых свойств сенсорных стимулов, но неспособная извлекать семантическую информацию и идентифицировать стимулы (система предвнимания). После него находится система с ограниченной емкостью, в которой и происходит идентификация стимулов. Следствием является то, что мы можем идентифицировать только те объекты, на которые обращаем внимание, но можем неосознанно воспринимать простые черты всех остальных объектов в окружении. Например, человек может слышать шум разговора в соседней комнате, но сможет понять его смысл, только если начнет прислушиваться. Уязвимым местом этой теории является то, что при отсутствии восприятия семантических свойств объектов за пределами внимания трудно объяснить, почему мы обращаем внимание на одни, а не на другие объекты.

Поэтому была предложена теория позднего выбора, согласно которой практически все свойства объектов (включая семантические) уже воспринимаются неосознанно системой предвнимания, но только после прохождения через «бутылочное горлышко» внимания информация поступает в сознание и рабочую память [Deutsch J.A., Deutsch D., 1963]. Вариант этой теории развивают в более поздних работах приверженцы теории сознания как глобального рабочего пространства. По их мнению, проекция информации в рабочую память и делает ее доступной для сознания [Dehaene et al., 2006; Prinz, 2012].

Одним из следствий теории позднего выбора является то, что стимулы, на которые мы не обращаем внимания, тем не менее неосознанно воспринимаются (включая, хотя бы частично, их семантические свойства). Это было неоднократно подтверждено в экспериментах с праймингом, когда не воспринимаемые сознательно стимулы вызывают изменение поведенческих показателей [Tipper, Driver, 1988]. Другое следствие теории позднего выбора, согласно которому внимание лишь определяет, что попадает в сознание, а что нет, но не влияет на активность ранних уровней сенсорного восприятия, не было подтверждено экспериментами. Напротив, как было показано, внимание влияет на активность не только высших отделов, представленных в коре, но и более ранних уровней обработки стимула в латеральных колленчатых телах [O'Connor et al., 2002]. По убеждению некоторых теоретиков, дебаты между сторонниками раннего и позднего выбора являются бесплодными, поскольку как те, так и другие представляют сен-

сорное восприятие в виде последовательного линейного процесса, в то время как в реальности имеет место сложная параллельная организация [Allport, 1993].

Другое направление связывает внимание с интеграцией информации (feature integration theory). Данное направление пытается использовать внимание как решение так называемой проблемы связывания (the binding problem). Сенсорная информация поступает в виде множества отдельных черт (формы, цвета, движения, запахи, звуки), принадлежащих разным объектам и кодируемых в мозге различными модулями. Как эти разрозненные черты объединяются в единую картину внешнего мира? Если внимание выхватывает из множества объектов какой-то один, то это позволяет связать все черты, воспринимаемые в данной части пространства, как черты конкретного объекта и таким образом решить проблему связывания [Treisman, 1999]. Данная проблема, однако, критиковалась впоследствии философами как псевдопроблема [Plate, 2007].

Теории «бутылочного горлышка» в целом критиковались многими. Уже в 1970-х гг. Ульрик Нейсер с соавторами показали, что при соответствующей тренировке испытуемых можно обучить выполнять одновременно две задачи, требующие внимания [Neisser, 1976]. По его предположению, присущая вниманию селекция информации может быть связана не с наличием «бутылочного горлышка» в сенсорном восприятии, а с координацией поведения для достижения поставленной цели. То же может иметь место и в отношении процесса мышления, где всегда есть необходимость выбора из множества источников информации и множества способов манипулирования выбранной информацией для достижения цели последовательного мышления [Wu, 2014]. Внимание по этой интерпретации играет гораздо более фундаментальную роль, чем просто участие в сенсорном восприятии.

Внимание – это механизм для достижения поставленных целей. Даже произвольное внимание отражает готовность действовать в отношении стимулов, которые привлекают внимание [Ibid]. С этой точки зрения селективность внимания понимается как выбор способа действий из множества доступных в данный момент вариантов и подавления всех остальных потенциально возможных способов. Еще более общая трактовка внимания приписывает ему функции выбора в любой ситуации, где такой выбор необходим, будь то ситуация «бутылочного горлышка» или ситуа-

ция выбора курса действий или хода размышлений [Watzl, 2017]. По теории Ватцла внимание позволяет существование в сознании структуры приоритетов, без которой был бы невозможен личный взгляд на мир и, стало быть, невозможно существование сознания, которое по определению является субъективным.

Другая интерпретация внимания, не связанная с идеей «бутылочного горлышка», развивается приверженцами байесовского подхода к пониманию когнитивных процессов [Friston et al., 2006; Hohwy, 2013; Clark, 2017]. В рамках этого подхода все когнитивные процессы в целом, включая сенсорное восприятие, мысль и действие, понимаются как иерархически организованная система байесовского тестирования гипотез. На каждом уровне системы гипотезы, пришедшие из более высокого уровня, тестируются на основе информации, пришедшей из ниже расположенного уровня, и ошибка рассогласования пересылается обратно на верхний уровень для исправления гипотезы. Содержание восприятия определяется гипотезой с наименьшей ошибкой рассогласования [Hohwy, 2013; Clark, 2017].

Функция сенсорного восприятия, таким образом, не в том, чтобы предоставить информацию о внешнем мире, а в том, чтобы определить, насколько ошибочны наши гипотезы о внешнем мире. То есть мы видим не то, что, например, на улице начался дождь, а то, что погода более дождливая, чем мы ожидали. По этой теории внимание устанавливает «веса» для ошибок рассогласования в соответствии с ожидаемой точностью восприятия [Hohwy, 2012]. «Точность» тут воспринимается в смысле, контрастирующим ее с «аккуратностью».

Аккуратность понимается как мера рассогласования между значением параметра, закодированным в сигнале, поступающем от органов чувств, и действительным значением этого параметра, а точность – как мера случайных флуктуаций сигнала при постоянстве действительного значения параметра. В этом смысле измерительный инструмент может быть очень точным, но не очень аккуратным. Эмпирические данные показывают, что при принятии решений о содержании восприятия точность берется в расчет. Например, в условиях густого тумана мы можем больше доверять слуху, чем зрению, и это связано с перераспределением внимания, т. е. в приписывании «весов» информации, получаемой из разных каналов. Спорным, однако, является утверждение, что внимание – это нечто большее, чем оптимизация точности в

иерархической байесовской системе. Такое понимание отождествляет все формы внимания с произвольным вниманием, которое понимается как степень уверенности в содержании восприятия [Ransom et al., 2017].

Популярное направление в теориях внимания – уже многократно рассмотренная нами модель соревнования модулей. Простой механизм параллельной гонки соревнующихся часто дополняется элементами взаимной борьбы или процессами управления сверху, которые могут вносить элемент «предвзятости» и опосредуют таким образом механизм целенаправленного внимания [Bundesen, Nabekost, 2008; Reynolds, Desimone, 2000]. Варианты этой модели можно совмещать как с теориями «бутылочного горлышка», так и с теориями внимания для достижения цели.

Давнюю традицию имеют теории, использующие в отношении внимания метафору прожектора, направленного в определенную точку пространства. Есть причины думать, что расположение в пространстве источников сенсорных стимулов играет важную роль в механизмах управления вниманием, однако роль эта сложнее, чем можно думать, опираясь на метафору прожектора. Например, как известно, внимание распределяется в пространстве не на основе простых пространственных координат, когда все его точки имеют одинаковую вероятность попасть в фокус внимания, а на основе распределения в нем объектов. Так, Джон Дункан показал, что внимание легче переключается между двумя местоположениями, которые попадают в границы одного объекта, чем между равноудаленными местоположениями, разделенными границей объекта [Duncan, 1984].

Помимо расположения в пространстве на распределение внимания влияет не только размер, цвет и форма фигуры, но и ее смысл (например, буква это или цифра) [Kravitz, Behrmann, 2011]. По некоторым теориям воспринимаемые объекты кодируются в мозге в виде «карты значимости» (saliency maps), в которой каждому объекту приписываются не только его объективные свойства, включая пространственное расположение, но и субъективные, такие как эмоциональная валентность и мотивационная значимость [Pessoa, 2013]. Эта теория позволяет, в частности, объяснить механизмы непроизвольного внимания [Henry, 2017]. В целом, хотя роль пространственных факторов в распределении внимания трудно отрицать, метафора прожектора в ее простейшей форме описывает далеко не все аспекты управления вниманием.

Расширением рассмотренных выше теорий внимания, как механизма для осуществления действий, является теория, согласно которой управление вниманием интимно связано с управлением моторными функциями. В экспериментах на макаках выявлено, что стимуляция областей двигательной коры, управляющих движениями глаз (frontal eye field), влияет на восприимчивость к стимулам областей зрительной коры [Armstrong, Moore, 2007]. Показано также, что распределение внимания по определенным частям объекта зависит от того, какой рукой испытуемый может схватить этот объект [Deubel, Schneider, 2004]. Хотя эти теории хорошо объясняют некоторые лабораторные эксперименты, их трудно распространить на другие проявления внимания, не связанные прямо с пространственными характеристиками объектов и с моторным контролем соответствующих видов движения.

Многие теории связывают внимание с рабочей памятью и объясняют избирательность внимания ограниченным объемом этой памяти. Избирательность внимания в данном случае обеспечивается конкурсным отбором, определяющим, какая информация получает доступ к рабочей памяти [Knudsen, 2007]. Когнитивные процессы верхнего уровня могут усиливать или ослаблять интенсивность импульсации в тех или иных конкурирующих сенсорных каналах, обеспечивая таким образом механизм произвольного внимания, а мотивационные системы автоматически усиливают редкие или биологически значимые стимулы, обеспечивая механизм непроизвольного внимания [Pattyn et al., 2008]. Эти нисходящие и восходящие механизмы представлены в мозге в виде пространственных карт, стимуляция которых усиливает импульсацию в соответствующих каналах и облегчает восприятие соответствующих стимулов. Когнитивные процессы верхнего уровня представлены пространственными картами, включающими дорсолатеральную префронтальную кору (ДЛПК) и переднее поле глаза (frontal eye field), а также области теменной коры. Экзогенные (восходящие) механизмы представлены областями первичной зрительной коры и верхнего бугорка (superior colliculi) [Zhaoping, 2016].

Можно отдельно выделить теорию внимания, предложенную в 1990 г. Познером и Петерсоном [Posner, Petersen, 1990] и активно развивавшуюся в последующие годы [Petersen, Posner, 2012]. Теория описывает три сети, участвующие в регуляции внимания: пробуждающую сеть (alerting network), включающую стволую систему активации (arousal system) и структуры правого полушария,

связанные с поддержанием бдительности; сеть ориентации (orienting network), включающую прежде всего области теменной коры; и исполнительную сеть (executive network), включающую передние области поясной коры [Posner, Petersen, 1990]. Открытие сетей покая послужило поводом для модификации теории Познера [Petersen, Posner, 2012]. В настоящее время описано несколько сетей, так или иначе участвующих в процессах внимания, а именно – дорзальная и вентральная сети внимания, центральная исполнительная сеть (central executive network) и сеть значимости (salience network) [Seeley et al., 2007]. Эти работы подчеркивают важность взаимодействия между различными, часто далеко друг от друга расположенными областями коры для координации процессов внимания.

Таким образом, все описанные теории и существующие эмпирические данные позволяют объяснить некоторые особенности внимания, наблюдаемые в поведенческих лабораторных экспериментах, но пока нет даже намека на эмпирически обоснованную модель, которая бы детально описывала воплощение механизмов внимания в активности мозга. Как конкретно запускаются волевая концентрация внимания или произвольная его регуляция – остается тайной.

Важным вопросом является связь внимания с сознанием. На одном конце спектра мнений лежит взгляд, согласно которому объекты, на которые мы не обращаем внимания, вообще не попадают в сознание [Cohen et al., 2012; Carruthers, 2015]. Это мнение основано на большом объеме экспериментальных данных, свидетельствующих о том, что испытуемые обычно игнорируют предметы, на которые они не обращали внимания [Rensink, 2002]. Те же самые данные, однако, не доказывают необходимость внимания для сознательного восприятия, и есть данные, показывающие, что сознание и внимание не всегда идут рука об руку [Koch, Tsuchiya, 2007]. В целом, однако, вопрос о соотношении внимания и сознания остается предметом дебатов [Mole, 2017].

5.3. Мышление

Способность мыслить, находить причинно-следственные связи и закономерности в наблюдаемом мире, безусловно, является коренной чертой человека, отличающей его от всех остальных живых существ на Земле. Связанные с мышлением процессы и способности изучают с разных сторон разные дисциплины, включая

философию, лингвистику, антропологию, компьютерные науки, психологию, психиатрию и нейронауки. Тема эта, конечно же, слишком обширна, чтобы пытаться осветить ее в настоящей книге со всех сторон, и мы лишь кратко рассмотрим современные данные о связи процессов мышления с активностью мозга. Эти данные преимущественно относятся к механизмам дедуктивного мышления, т. е. способности вывести частное заключение из общего посыла. Классический пример: (1) все люди смертны; (2) Сократ – человек; (3) значит, Сократ смертен. В рассуждениях подобного рода предполагается, что первые два посыла абсолютно верны, и вывод, который из них можно сделать, однозначен. То есть ответ в данном случае имеет бинарную природу – «верно» или «неверно».

В противоположность приведенному, индуктивное мышление представляет собой вывод общей закономерности из наблюдения частных примеров. Например, если мы видим множество разных птиц и все они могут летать, мы можем прийти к заключению, что все птицы летают. Это заключение, однако, имеет вероятностный характер и может быть ошибочным или требующим уточнения. Правильнее будет сказать, например, что большинство птиц летают, но есть некоторые виды птиц, которые не могут летать. Поэтому, если мы встречаемся с новым видом птиц, мы можем с высокой (но не стопроцентной) вероятностью предположить, что они могут летать.

Надо сказать, что в когнитивной психологии пока нет общепринятой теории о природе ментальных процессов, лежащих в основе дедуктивного мышления. Предложено как минимум две альтернативные теории. По теории ментальной модели (ТММ) дедуктивное мышление – это невербальный процесс конструкции и манипуляции пространственной репрезентацией исходных постулатов проблемы [Johnson-Laird, 2001]. Странники теории формальных правил (ТФП) считают, что дедукция связана с извлечением и использованием формальных правил, наподобие грамматических правил построения фразы [Rips, 1994]. Существуют эмпирические данные в пользу как той, так и другой модели, и истинная природа ментальных процессов, лежащих в основе дедуктивного мышления, остается предметом дискуссий [Ibid].

Как это часто бывает в современной когнитивной психологии, надежды на разрешение спора возлагались на исследование активности мозга при решении дедуктивных задач. Предполагалось, что если верна ТММ, то активация при решении дедуктивных задач

должна наблюдаться в затылочно-теменных областях, участвующих в обработке визуально-пространственной информации, если же верна ТФП, должна «светиться» левая нижняя лобная извилина и другие области левого полушария, участвующие в процессах распознавания грамматических конструкций.

Начиная с конца 1990-х гг. эти гипотезы проверялись в работах Гоела, но накопленные данные не позволили ответить на вопрос, какая из двух теорий верна. В разных работах наблюдалась активация то в затылочно-теменных областях, то в областях, связанных с лингвистическими/синтактическими процессами, то вообще в других областях мозга [Goel, 2007]. Данные различия пытались объяснить различиями в методах и тем, что испытуемые могут использовать разные стратегии решения задач [Roberts, Newton, 2005]. Метаанализ этих работ показал, однако, что достоверная активация во всех видах задач выявляется в левой лобно-затылочной системе и базальных ганглиях [Prado et al., 2011]. По мнению авторов, это согласуется с гипотезой о роли левого полушария в интерпретации поступающей информации [Roser, Gazzaniga, 2006], однако противоречит идее, по которой дедуктивное мышление поддерживается единой системой мозга, опирающейся либо на зрительно-пространственные, либо на основанные на правилах механизмы. Дополнительно выявлялась активация, которая не сопровождала все виды задач, но зависела от характера дедуктивной задачи.

Пропозициональные суждения, которые не анализируют внутреннюю структуру простых высказываний, а лишь учитывают, с помощью каких союзов и в каком порядке они сочленяются в сложные (конструкции типа «если..., то...»), активировали задние теменные, прецентральные и медиальные лобные области левого полушария. Реляционные суждения (типа «А больше, чем В») активировали задние теменные области обеих полушарий и правую среднюю лобную извилину. Категориальные суждения (типа «все капитаны курят трубку») активировали нижнюю лобную извилину и базальные ганглии в левом полушарии [Prado et al., 2011]. Прадо с соавторами делают вывод, что в мозге нет системы, которая участвовала бы во всех трех видах суждений. Только некоторые виды суждений, в частности, реляционные, но не пропозициональные и категориальные, могут вовлекать визуально-пространственные корковые механизмы. Поэтому унитарные теории, такие как ТММ и ТФП, не согласуются с этими данными.

В теориях типа ТММ и ТФП дедуктивные суждения рассматриваются в рамках бинарной логики – «верно» или «неверно». В последние 15–20 лет все более популярными становятся вероятностные модели человеческого мышления, которые относят к «новой парадигме» [Manktelow, 2012]. Она имеет две отличительные особенности, рассматривая мышление, во-первых, в целом как вероятностный процесс; во-вторых, – как двойственный процесс, использующий систему 1 и систему 2 (быструю и медленную системы в теории Канемана) [Kahneman, 2011].

Вероятностная природа человеческих суждений обосновывается тем, что в обычной жизни люди практически никогда не пользуются формальными правилами дедуктивной логики. Стандартом являются так называемые энтимематические рассуждения, которые основываются на как бы очевидном и потому умалчиваемом знании. Например, классический силлогизм про Сократа может звучать: «Сократ смертен, потому что он человек». Первый посыл силлогизма («все люди смертны») опускается, как самоочевидный. Такие суждения, однако, могут быть ошибочными. Например, узнав, что некое существо по имени Клуша является птицей, мы можем решить, что Клуша может летать, так как в подавляющем большинстве случаев птицы могут летать и, соответственно, вероятность того, что Клуша летает, очень высока. Может оказаться, однако, что Клуша – это курица или страус. Странники новой парадигмы считают, что стандартные дедуктивные задачи не энтимематического характера, используемые в лабораторных исследованиях, не отражают практику человеческого мышления, имеющего вероятностный характер. По их мнению, дедуктивное мышление не отличается в принципе от индуктивного, которое по определению является вероятностным [Oaksford, Chater, 2007].

Вторая особенность теорий новой парадигмы – утверждение о двойственной природе мышления [Stanovich, 2011]. В теории Канемана система 1 – это имплицитный (неосознанный), вероятностный и основанный на общем знании способ быстрой генерации ответов. Система 2 работает сознательно и медленно, используя рабочую память и аналитические процессы [Kahneman, 2011]. По убеждению одних, аналитические процессы тоже имеют вероятностную природу [Oaksford, Chater, 2009], другие думают, что

они используют стандартную бинарную логику [Stanovich, 2011]. В любом случае считается, что дедуктивное мышление – это результат взаимодействия между системами 1 и 2.

Работу системы 1 Канеман иллюстрирует следующим примером. Если человек видит сочетание слов «банан» и «рвота», у него непроизвольно генерируется серия реакций, включая неприятные зрительные и звуковые образы, некоторое удивление от сочетания этих слов и, возможно, конструкция целой истории, связывающей рвоту с поеданием бананов. Все это происходит неосознанно и очень быстро, и мы осознаем лишь общую реакцию на эти слова. Примером работы системы 2 является, например, последовательное вычитание одного числа из другого, которое происходит полностью сознательно и использует правила арифметики. Взаимодействие систем можно проиллюстрировать на примере задачи, в которой первый предмет стоит на 1000 рублей дороже второго, а вместе они стоят 1100 рублей. Спрашивается, сколько стоит второй предмет. Быстрый спонтанный ответ системы 1 будет 100 рублей, что очевидно неправильно, так как суммарная стоимость в этом случае будет 1200. Тогда спонтанный ответ системы 1 должен быть подавлен и правильный ответ получается после медленного сознательного расчета с помощью системы 2. Это, однако, возможно, только если человек знает правила арифметики.

В случае дедуктивных логических задач сознательное применение правил логики может быть недоступно обычному человеку, если он не изучал эти правила. При заданных условиях, однако, возможно использование логической интуиции. Показано, что люди способны неосознанно выявлять несоответствие между спонтанным ответом системы 1 и правильным ответом [De Neys, 2014]. Поэтому в дедуктивном мышлении работа системы 2 тоже не всегда может быть осознанной [Oaksford, 2015]. На уровне сознания человек видит лишь быстрый ответ системы 1, сопровождаемый чувством правильности или неправильности этого ответа, которое может зависеть от его оценки системой 2 [Thompson et al., 2011]. Пример Канемана с бананом показывает, что на бессознательном уровне системой 1 может совершаться множество сложных операций, включая оценку вероятности возникновения рвоты после поедания бананов и эмоциональную реакцию, сопровождающую генерацию каузальной модели, связывающей бананы с рвотой. Предполагается, что генерация и использование этой информа-

ции в рассуждении аналогично действию бессознательных механизмов в сенсорном восприятии и регуляции движений [Oaksford, 2015] и организовано по принципу байесовских процессов [Friston, 2008].

Таким образом, согласно этим теориям, большая часть мышления происходит на подсознательном уровне на основе информации, генерируемой системой 1, и с использованием процессов байесовского типа. При этом содержание рассуждения (семантика), а не его форма (синтаксис) имеет первостепенное значение. Большинство же нейровизуализационных исследований дедуктивного мышления, как отмечает Оксфорд, построены так, что содержание рассуждения «выносится за скобки», так как дедуктивные трайлы контрастируются с контрольными трайлами с аналогичным содержанием [Oaksford, 2015]. Кроме того, связанная с содержанием активация в мозге может сильно варьировать как между испытуемыми, так и при повторных измерениях у одного и того же испытуемого. В результате, по мнению Оксфорда, нейровизуализационные исследования не способны выявить активность мозга, связанную с работой системы 1 [Ibid].

Другая проблема с нейровизуализационными исследованиями заключается в интерпретации функционального смысла выявленной активации какой-либо области мозга. Обычная практика – приписывать этой активации функциональное значение на основе ранее выявленных функциональных коррелятов этой области мозга. Однако, как пишет Оксфорд, ранее выявленные корреляты имеют лишь историческое значение и правильнее было бы интерпретировать функциональное значение выявленной активации исходя из теоретической модели изучаемого когнитивного процесса [Ibid]. Не всегда, впрочем, такая модель имеется, или может быть несколько конкурирующих моделей.

При изменении модели изменяется и интерпретация выявленной активации. Кроме того, если приписывать активации функциональное значение исходя из модели текущего исследуемого процесса, то при сопоставлении результатов разных исследований, как правило, оказывается, что одна и та же область мозга может принимать участие в большом количестве разных функциональных процессов, связь которых друг с другом далеко не всегда очевидна. Например, если при решении абстрактных дедуктивных задач выявляется активация медиальной префронтальной коры

(МПК), то, исходя из стандартной интерпретации этого когнитивного процесса, можно предполагать, что МПК координирует процессы извлечения из памяти и использования абстрактных логических правил. Но, исходя из теории Канемана и считая, что эти задачи обычно решаются не с помощью абстрактных логических правил, а с использованием аналогий, нужно думать, что МПК в этом случае координирует процессы извлечения из памяти и сравнения аналогичных суждений.

По мнению Оксфорда, маловероятно, что в человеческом мозге в процессе эволюции мог сформироваться специальный модуль для использования абстрактных правил формальной логики, которая является культурным феноменом, зародившемся около 2 тыс. л.н., и которая нужна для аргументации с оппонентами и изложения мысли на бумаге. Человеку естественней и проще мыслить конкретными образами. Конструкции типа «если..., то...» универсально применяются при описании реальных событий в мире и доступны не только человеку. Домашние животные, например, быстро понимают, что если хозяин их «застукает» за совершением запрещенного поступка, то они будут наказаны. Поэтому интуитивная система 1 имеет наработанные механизмы и легка в использовании.

В целом, из обзора исследований мозговых коррелятов мышления Оксфорд заключает, что имеется лишь очень слабая связь между функциями и активностью областей мозга, которые предположительно участвуют в их осуществлении [Oaksford, 2015]. По мнению автора, описанные в метаанализе Прадо с соавторами [Prado et al., 2011] исследования в принципе не могли выявить структуры мозга, участвующие в процессах мышления. Он допускает, что в подавляющем большинстве случаев спонтанное мышление происходит неосознанно и только его результат поступает в сознание, которое принимает решение согласиться с этим результатом или отбросить его. Таким образом, обнаруженная в метаанализе Прадо с соавторами активация в левой лобно-затылочной системе и базальных ганглиях может быть связана не с самим процессом мышления, а с осознанием результата и подготовкой вербального ответа [Ibid].

По классической теории Пиаже когнитивное развитие детей проходит четыре стадии: от рождения до 2 лет – сенсомоторная стадия; от 2 до 7 лет – преоперативная стадия, во время которой

дети не способны логически мыслить; от 7 до 12 лет – конкретная оперативная стадия, во время которой дети начинают использовать логическое мышление в некоторых логико-математических сферах (например, числовая категоризация); наконец, после 12 лет мышление не ограничено конкретными объектами и может использоваться для абстрактных суждений [Piaget, 1983]. Современные теоретики считают, что Пиаже недооценивал досознательное логическое знание, уже имеющееся у младенцев, и переоценивал логические способности старших детей, подростков и взрослых, часто допускающих систематические ошибки даже в очень простых логических задачах [Houdé, 2000; Kahneman, 2011]. Эти ошибки обычно имеют место, когда дети, подростки и взрослые полагаются на доминантные нелогические интуиции и вводящие в заблуждение эвристические стратегии, а не на логические алгоритмы. Способность преодолеть такие ошибки напрямую зависит от способности подавить эти интуитивные формы мышления [Houdé, Borst, 2014].

Соответственно, в настоящее время теория дискретных фаз развития Пиаже заменена моделью, в которой в любой период времени различные когнитивные стратегии разного уровня сложности перекрываются и взаимодействуют друг с другом по принципам нелинейной динамической системы [Siegler, 1999]. Согласно этой теоретической схеме, прогрессирующая способность подавлять нерелевантные или вводящие в заблуждение стратегии, которая связывается с развитием префронтальной коры, и активировать использование логики определяет развитие детей и переход от одной стадии Пиаже к другой [Houdé, Borst, 2014]. Чаще всего как дети, так и взрослые спонтанно полагаются на интуитивные эвристические решения, но взрослые в большей степени способны преодолеть их. В серии исследований Хауде и Борст показали, что, по крайней мере в определенных типах задач, ошибки возникают не из-за дефицита логического мышления, а из-за неспособности подавить доминантный интуитивный ответ. Эту неспособность можно уменьшить путем соответствующей тренировки, и, согласно полученным нейровизуализационным данным, доминантная активация мозга сдвигается от задних чувствительных зон коры к передним, связанным с исполнительным контролем [Houdé, Borst, 2015]. Интерпретируя результаты своих исследований, авторы предполагают, что левая и правая нижняя лобные извилины участвуют в торможении неправильных ответов.

Важно подчеркнуть, что в силу специфики нейровизуализационных экспериментов (необходимость «привязать» активность мозга ко времени определенного психологического события) исследования мышления практически всегда проводятся в форме предъявления задач и ответа на поставленные в эксперименте вопросы. Это позволяет разбить процесс решения задачи на отдельные этапы и связать наблюдаемую активацию мозга с когнитивными процессами, предположительно сопровождающими каждый этап. В обычной жизни процесс мышления, как правило, не ограничен жесткими рамками вопросов и ответов и может проходить нелинейно, перескакивая с одной стадии на другую и возвращаясь при необходимости к ранее отвергнутым вариантам.

Использование «связанного с событиями» экспериментального дизайна ведет, по мнению Папо, к тому, что выявляемая активация мозга связана с результатом, а не с процессом мышления [Папо, 2015]. В мозге нет специализированного модуля для мышления. Существующие в мозге модули обеспечивают быстрое выполнение перцептивно-моторных функций, которое в основном зависит от статистических свойств входного сигнала. В противоположность этому, мышление зависит от внутренне генерируемых сигналов и его временные и пространственные характеристики ничем не ограничены. Как считает Папо, такого рода процессы должны коррелировать с энергетически затратной длиннодистантной коммуникацией между разными частями мозга. Мышление может быть связано с своего рода резонансным режимом мозга, при котором функциональная эффективность достигается с помощью специфических, хотя и нестабильных, пространственно-временных паттернов [Ibid]. Это рассуждение предполагает, что изучение сопровождающих мышление паттернов функциональных связей («коннективности») между областями мозга может оказаться более плодотворным, чем изучение только активации этих областей.

Мы уже рассматривали метаанализ Прадо с соавторами, в котором обобщены результаты нейровизуализационных исследований мышления за первую декаду текущего столетия [Prado et al., 2011]. В 2020 г. был опубликован метаанализ работ, опубликованных во второй декаде [Wang et al., 2020]. Авторы использовали метод оценки вероятности активации (activation likelihood estimation) каждого района мозга на основе данных 38 нейровизуализационных исследований. Кроме того, был проведен анализ функцио-

нальных связей между этими районами. Сравнение периода решения задачи с фоновым периодом установило 702 локуса активации во всех областях мозга с преимущественным скоплением этих локусов в левом полушарии – в нижней, средней и медиальной лобных извилинах, нижней теменной дольке, хвостом ядре и островке. Анализ коннективности выявил одну корковую сеть, включающую нижнюю лобную и переднюю поясную извилины и островок в левом полушарии, и одну корково-подкорковую сеть, включающую связи хвостатых ядер с нижней и средней лобными извилинами, островком, таламусом, верхней теменной долькой и поясной корой левого полушария. Эти результаты в какой-то степени согласуются с результатами предшествующего анализа (в частности, подчеркивают участие левого полушария).

Для выделенных в этом метаанализе структур ранее было показано участие в широком спектре когнитивных процессов, таких как внимание, рабочая память, эмоции, обработка лингвистической информации. В обоих метаанализах подчеркивается роль нижней лобной (особенно левой) извилины, активация которой многократно раскрывалась в задачах, требующих применения грамматических правил [Grodzinsky, Santi, 2008]. В то же время обнаружена активация теменных областей, связанных с пространственным вниманием и манипуляцией информацией в пространственной рабочей памяти [Dehaene et al., 2003; Koenigs et al., 2009]. Эти данные, как и данные первого метаанализа, показывают, что ни теория ментальной модели, ни теория формальных правил не могут по отдельности объяснить выявленный паттерн активации. Процесс дедуктивного мышления связан с активацией и с передачей информации между большим количеством далеко друг от друга расположенных областей мозга, участвующих в широком спектре когнитивных процессов. Еще раз подтверждается наблюдение, что в мозге нет специализированного модуля для мышления. Все вовлеченные области – это «специалисты широкого профиля», участвующие в массе других когнитивных процессов.

Немалое количество работ посвящено исследованию мозговых коррелятов научного мышления, которое можно считать примером индуктивного мышления. В недавнем обзоре этих работ рассматривались три вида интеллектуальной активности: преодоление заблуждений, причинно-следственные рассуждения и генерация гипотез [Nenciović et al., 2019]. Результаты в целом показывают активацию в латеральных префронтальных областях, связанных с

исполнительными функциями, и в средней височной области, участвующей в процессах декларативной памяти. Преодоление заблуждений, которое можно трактовать как подавление спонтанных интуитивных суждений, генерируемых системой 1 в терминологии Канемана, по интерпретации авторов связано с активностью левой и правой нижней лобных извилин, передней части поясной извилины и левой дорсолатеральной коры. Причинно-следственные заключения, если они генерируются в процессе непосредственного восприятия событий (например, наблюдения за столкновением бильярдных шаров), сопровождаются билатеральной активацией затылочно-височных областей, участвующих в зрительном восприятии. Если решение о наличии причинно-следственных связей должно быть принято сознательно, то дополнительно активируются дорсомедиальные и дорсолатеральные области префронтальной коры. Генерация гипотез сопровождалась активацией нижней лобной, дорсолатеральной и парагиппокампальной извилин в левом полушарии, что интерпретируется как вовлечение процессов внимания, а также рабочей и декларативной памяти.

Обратим внимание на нестандартные способы мышления, а именно – креативное мышление и инсайт. Последнее традиционно описывается как субъективное чувство внезапного «озарения», которое может появляться в процессе решения задачи [Davidson, Sternberg, 2003; Cushen, Wiley, 2012; Weisberg, 2014]. Это чувство обычно сопровождается уверенностью в правильности найденного решения и часто действительно указывает на правильный ответ [Irvine, 2015], но иногда «инсайтное» решение может оказаться и неправильным. Данный феномен многократно описан в воспоминаниях выдающихся представителей творческих профессий [Poincaré, 1913] и довольно рано стал компонентом теоретических моделей креативности. Так, по теории Уоллеса процесс возникновения креативной идеи проходит четыре стадии – подготовка (накопление необходимой информации), инкубация (период, когда сознательное внимание не направлено на решение задачи), инсайт (креативная идея «вспыхивает» в сознании) и верификация (креативная идея подвергается проверке) [Wallas, 1926].

Изучение инсайта в лаборатории сопряжено с очевидными трудностями. Особенно это относится к исследованию мозгового обеспечения инсайта, которое требует точной привязки по времени изучаемого психологического события. Вряд ли приходится рассчитывать на то, что испытуемые смогут по заданию в строго

определенное время находить креативные решения и испытывать инсайт. Как выразились Дитрих и Канзо, попытки связать возникновение инсайта с активностью мозга аналогичны попыткам прибить гвоздями желе к стенке [Dietrich, Kanso, 2010]. Вероятно, поэтому существующие данные о фМРТ- и ЭЭГ-коррелятах инсайта достаточно противоречивы.

Значительная часть этих исследований была направлена на проверку одной из нескольких популярных гипотез в области изучения креативности и инсайта. Согласно одной из гипотез, особую роль в инсайте играет правая височно-теменная кора и, в частности, правая верхняя височная извилина (ВВИ) [Bowden, Jung-Beeman, 2003; Jung-Beeman et al., 2004]. Хотя несколько исследований действительно указывает на особую роль правой ВВИ в возникновении инсайта [Jung-Beeman et al., 2004; Qiu et al., 2006; Kounios et al., 2008; Sandkuhler, Bhattacharya, 2008], еще большее количество работ не подтверждает эту гипотезу [Lavric et al., 2000; Danko et al., 2003; Mai et al., 2004; Kounios et al., 2006; Lang et al., 2006; Qiu et al., 2008a,b], а некоторые исследования, наоборот, указывают на особую роль левополушарных областей мозга, в частности, левой ВВИ, как в возникновении инсайта, так и в других стадиях креативного мышления [Qiu et al., 2008a].

В некоторых ЭЭГ-исследованиях инсайта отмечено увеличение мощности альфа-осцилляций в лобных, теменных и височных областях [Kounios et al., 2008; Sandkuhler, Bhattacharya, 2008], что интерпретируется как свидетельство сниженной активности сознания. Это, однако, противоречит данным фМРТ и ПЭТ, указывающим на активацию лобных и височно-теменных зон коры при инсайте [Dietrich, Kanso, 2010], так как увеличение мощности альфа-осцилляций обычно коррелирует со снижением кровотока в соответствующих областях коры [Sadato et al., 1998; Goldman et al., 2002]. В других работах описано сопровождающее инсайт снижение мощности альфа-осцилляций и увеличение мощности низкочастотных (дельта и тэта) и высокочастотных (бета и гамма) осцилляций [Danko et al., 2003; Jung-Beeman et al., 2004; Kounios et al., 2006].

Одна из немногих хорошо воспроизводимых находок при изучении инсайта – участие ВВИ [Jung-Beeman et al., 2004; Qiu et al., 2006, 2008a; Kounios et al., 2008; Sandkuhler, Bhattacharya, 2008], что неудивительно, учитывая роль этой области коры в вербальных процессах. Эта находка, впрочем, относится лишь к одной

экспериментальной парадигме – тесту отдаленных ассоциаций Медника [Mednick, Mednick, 1967]. В целом, существующие данные дают достаточно противоречивую картину активности мозга, сопровождающей возникновение инсайта. Можно думать, что инсайт является лишь кульминацией серии состояний и процессов мозга, оперирующих на различных временных шкалах [Kounios, Jung-Beeman, 2009]. Соответственно, креативность и инсайт не связаны с каким-то одним ментальным процессом или областью мозга [Dietrich, Kanso, 2010].

Таким образом, результаты, полученные в исследованиях мозговых коррелятов мышления, гораздо скромнее, чем результаты исследований, например, зрительного восприятия или представительства в мозге семантики слов. Если в этих последних направлениях получены данные, позволяющие (пусть и с небольшой степенью точности) «предсказать» из активности мозга содержание зрительного образа или смысл слова, то все результаты исследований мышления сводятся к выявлению корреляций между процессами мышления и активностью некоторых областей мозга. Функциональный смысл данных корреляций, как правило, малопонятен, так как выявляемые области мозга участвуют в большом количестве разнообразных когнитивных процессов.

Общая тенденция в последних работах – придавать большое значение неосознаваемым процессам. Это не только быстрые интуитивные решения, генерируемые системой 1 Канемана, но и решения, появляющиеся в форме инсайта, иногда после долгих и бесплодных сознательных поисков ответа. Важно подчеркнуть, что во всех случаях последнее слово в принятии решения и общая, направляющая процесс мышления роль остаются за сознательной системой 2.

Вторая тенденция – подчеркивать вероятностный характер процессов мышления. Значительная часть этого направления идет в рамках амбициозной программы «байесовский мозг», адепты которой пытаются объяснить все когнитивные процессы и все процессы мозга на основе байесовских механизмов. Как уже неоднократно отмечалось, все теории, пытающиеся объяснить, как мозг осуществляет те или иные когнитивные операции, остаются чисто умозрительными. Даже на теоретическом уровне нет модели, которая пыталась бы объяснить, как конкретно осуществление логических операций или генерирование гипотез может быть «закодировано» в активности нейронов. То же относится и к гипо-

тетическим байесовским механизмам, в основе которых лежит генерирование гипотез и вычисление рассогласования между ожидаемыми и реальными результатами их тестирования. Как в мозге генерируются гипотезы, как вычисляется рассогласование и как это рассогласование используется для корректировки гипотез? Мозговые механизмы всех этих процессов остаются тайной.

Если воспринимать поведение как единый процесс, то информация, поступающая из внешнего мира, должна интегрироваться с информацией, поступающей из памяти, и на основе этой информации, с учетом текущих целей и мотивации, должно приниматься решение, которое затем воплощается в поведении. В некоторых случаях все эти процессы происходят быстро и могут оставаться за пределами сознания. В других принятие решения и подготовка к его осуществлению могут занимать много времени, включают мышление и контролируются сознанием. В любом случае это должно требовать координации восприятия, извлечения следов памяти, эмоций, мышления и других сопутствующих процессов. На уровне сознания нам кажется, что интегратором и принимающим решения субъектом является то, что человек определяет словом «Я», но как это происходит в мозге?

Следующий раздел посвящен существующим данным о мозговых механизмах контроля и координации. В когнитивной психологии соответствующие функции принято обозначать термином «исполнительный контроль». В частности, функции интеграции потоков информации приписывают модулю, который называют «центральный исполнитель» (ЦИ, central executive).

5.4. Функции контроля в мозге

Традиционно функции исполнительного контроля связывают с активностью префронтальных областей мозга [Burgess, Stuss, 2017], однако характер этой связи является предметом дискуссий [Alvarez, Emory, 2006]. Разные аспекты исполнительного контроля связывают с дорсолатеральной префронтальной корой, передней частью поясной извилины (ППК) и орбитофронтальной корой (ОФК): ДЛПК участвует в «онлайн-обработке» информации и механизмах рабочей памяти [Barbey et al., 2013]; ППК связывают с большим количеством когнитивных функций, включая регуляцию эмоций, торможение ошибочного ответа и принятие решений; ОФК – ключевая область, участвующая в тормозном контроле им-

пульсивного поведения [Rolls et al., 2008]. По последним данным, мозжечок также участвует в функциях, связанных с исполнительным контролем [Noozian, 2014].

Существует ли «главный координатор» всех этих процессов, и если да, то где он находится? Этот вопрос остается предметом дебатов. По мнению некоторых исследователей, ЦИ не существует и отдельные когнитивные функции, связанные с активностью разных частей мозга, не «грузятся» на один латентный фактор ЦИ [Parkin, 1998]. Алан Бэдли, автор наиболее популярной теории рабочей памяти, считает, что концепция ЦИ плохо определена и ее можно критиковать как еще одну теорию «гомункулуса» в мозге, но пока ее следует оставить как напоминание о важной контрольной функции рабочей памяти [Baddeley, 1996]. Что же из себя представляет рабочая память и где она локализована, также является предметом дискуссий. В недавно опубликованной работе отмечено, что функции и локализация в мозге процессов рабочей памяти перекрываются с функциями и локализацией в мозге процессов внимания [Panichello, Buschman, 2021].

Если внимание контролирует сенсорное восприятие путем увеличения чувствительности сенсорных зон коры к определенным стимулам с помощью нисходящих влияний от префронтальной и теменной коры, то сходные механизмы работают и в отношении репрезентации объектов в рабочей памяти. Запись нейронной активности у обезьян выявила, что сходные области префронтальной коры участвовали в выборе как объекта в рабочей памяти, так и визуального стимула. При выборе элементов памяти они изначально были представлены в независимых подпространствах нейронной активности в префронтальной коре. Выбор элемента приводил к преобразованию его представления в новое подпространство. Похожие трансформации наблюдались и для внимания. Напротив, в теменной и зрительной коре внимание и отбор из рабочей памяти были представлены независимо друг от друга [Ibid]. В целом исследование мозговых коррелятов рабочей памяти показывает, что ее функции связаны с активностью множества систем, участвующих в сенсорном восприятии и управлении действиями, и ее можно считать эмерджентным свойством мозга и сознания [Postle, 2006].

Подавляющее большинство эмпирических данных, на которых строились теории рабочей памяти и ЦИ, было получено в поведенческих экспериментах на здоровых испытуемых и пациентах

с повреждениями мозга. Исследования были в основном сосредоточены на демонстрации функциональной диссоциации разных компонентов рабочей памяти. Данные же о конкретных областях мозга, участвующих в этих процессах, были получены позднее с использованием методов ПЭТ и фМРТ.

Согласно результатам экспериментов, различные управляющие функции (манипулирование и обновление информации, координация двух задач, процессы торможения и переключения) не только задействуют различные лобные области, но также зависят от задних (в основном теменных) областей. Эти результаты согласуются с гипотезой о том, что в мозге управляющие функции связаны с широко распределенной активностью, не ограниченной лобными областями. Функции ЦИ, таким образом, нужно понимать, как процесс взаимодействия между сетью регионов, а не как связь между одним регионом и одним когнитивным процессом более высокого уровня [Collette, van der Linden, 2002]. В последние годы ПЭТ- и фМРТ-исследования корреляций между активностью разных частей мозга привели к открытию так называемых сетей покоя, одной из которых (центральной исполнительной сети) приписывают функции ЦИ [Seeley et al., 2007]. Однако по сути исследования этой сети лишь демонстрируют важность взаимодействия между лобными и теменными областями мозга в осуществлении функций, приписываемых ЦИ.

Таким образом, несмотря на большое количество теорий о механизмах исполнительного контроля, основанных на данных поведенческих экспериментов, изучение активности мозга показывает, что выполнение когнитивных задач, требующих внимания, тормозного контроля, а также координации и интеграции информации из разных источников, сопряжено с активацией широко распределенных областей мозга. Это, в общем-то, не кажется удивительным, однако совершенно не проясняет конкретных механизмов интеграции и контроля, так же как и механизмов представления содержания рабочей памяти в активности мозга.

Завершив анализ данных о мозговых коррелятах сознания и неосознанного, а также базовых когнитивных процессов, включая механизмы управления и контроля, необходимо рассмотреть широко представленную в средствах массовой информации тему свободы воли, которая принципиально важна для решения обсуждаемых нами вопросов.

СВОБОДА ВОЛИ

К теме свободы воли обращались в той или иной форме практически все философы – в Западной традиции начиная с Платона и Аристотеля, а в Восточной – гораздо раньше. В философии обсуждение этой темы связано с онтологическими вопросами устройства мира, детерминизма, природы этики, добра и зла, примата материи (см., например [O'Connor, Franklin, 2020]). По убеждению некоторых философов, наличие свободы воли не требует специальных доказательств по ряду причин. Во-первых, у нас есть интроспективное доказательство свободы по крайней мере тех решений, на которые мы обращаем сознательное внимание [Guillon, 2014; Norgan, 2015; Wayne, 2017]. Во-вторых, наша вера в реальность свободы воли является эпистемологически фундаментальной и не требующей независимой доказательной поддержки, в частности потому, что она лежит в основе моральной ответственности [Swinburne, 2013]. Вместе с тем скептицизм по поводу наличия свободы воли также имеет давние традиции в философии. В значительной степени он подпитывался представлением о детерминированности всех событий в мире. Если все предопределено либо божественной волей, либо физическими законами, то свобода воли отдельного человека – это иллюзия.

Аргументация о детерминированности физического мира во многом потеряла силу после появления квантовой теории. Часто, однако, утверждается, что на макроуровне законами квантовой физики можно пренебречь и руководствоваться классическими детерминистскими законами. Эта идея, когда-то распространенная, сейчас подвергается сомнению даже на уровне фундаментальной биологии. Социальные, биологические и медицинские науки изобилуют чисто статистическими обобщениями, из чего можно сделать следующий вывод: в рамках данных наук любое событие имеет вероятностный характер и неизбежно включает элементы неопределенности. Современная наука, таким образом, решительно не поддерживает идею о том, что все, что мы делаем, предопределено прошлым, а далекое прошлое и вообще находится полностью вне нашего контроля [Roskies, 2014; Ellis, 2016]. Можно думать, что мы подвержены бесчисленному количеству причин-

ных влияний, но сумма этих влияний не определяет то, что мы делаем, а лишь повышает или понижает вероятность того, что мы сделаем то или другое.

Один из аргументов современных скептиков в вопросе о свободе воли состоит в том, что выборы, которые делает человек, определяются его личностью, культурной принадлежностью и образом мыслей, которые, в свою очередь, зависят от генетически заданной предрасположенности и воспитания. Все это, так же как и контекст ситуации, во многом обуславливающий поведение, является внешним по отношению к человеку, а значит, никакой свободы выбора нет [Strawson, 1994b]. Противники данной точки зрения возражают, что перечисленные выше факторы могут лишь объяснить, почему человек делает те или иные выборы, но не определяют их каузально. В каждой конкретной ситуации человек свободен делать или не делать то, к чему его подталкивает предрасположенность и контекст ситуации [Clarke, 2003]. Таким образом, некоторые поступки, совершенные вопреки действию этих факторов, могут выражать больше свободы и ответственности, чем другие [Mele, 2006]. Даже то, что из себя представляет человек в каждый момент его жизненной истории, зависит не только от внешних факторов (генетика, культура, контекст ситуации), но и от того, какие выборы он делал раньше [O'Connor, 2009b].

Еще одна линия аргументации скептиков в вопросе о свободе воли получила импульс после ставших знаменитыми экспериментов Бенджамина Либета в 1980-х гг. [Libet, 1985]. В классическом варианте эксперимента Либета испытуемых просили произвольно шевелить пальцем в свободно выбранный момент времени. Перед их глазами находилась стрелка циферблата, по которой они определяли и запоминали момент времени, в который они решили совершить движение. Параллельная запись ЭЭГ показала, что за 350 мс до отмеченного испытуемым момента в области двигательной коры регистрируется потенциал готовности. Согласно наиболее популярной интерпретации этих экспериментов, решение, воспринимаемое нами как сознательное и свободное, на самом деле принимает мозг до того, как оно становится известным сознанию.

На основе приведенных и похожих экспериментов многие заключили, что ощущение сознательной воли – это иллюзия [Wegner, 2002]. Были придуманы теории, объясняющие механизм данной иллюзии. Сам Либет считал, что сознание может все-таки играть активную роль, а именно – оно может (если успеет) наложить вето

на принятое мозгом решение [Libet, 2002]. Легко понять, почему его эксперименты именно в такой интерпретации стали широко известны. Они в точности соответствуют материалистической концепции, согласно которой сознание – это лишь иллюзия и эпифеномен активности мозга. По другой интерпретации, часто не отличающейся от первой, решение принимает подсознание и лишь потом происходит его осознание. При такой интерпретации свобода воли также считается иллюзией, так как мы не контролируем неосознаваемые мотивы, которые влияют на принятие решения до того, как оно становится осознанным.

Интерпретацию экспериментов Либета и аналогичных экспериментов его последователей многократно критиковали прежде всего потому, что ситуация, в которой испытуемому дается задание «свободно» выбрать и запомнить момент движения, отличается от естественных ситуаций свободного выбора. В течение всего эксперимента испытуемый должен помнить эту инструкцию и осуществлять интроспективный контроль своих ощущений, решая, в какой момент у него возникло «свободное» желание совершить движение, наблюдая при этом за стрелкой и запоминая, в каком положении она находилась в момент появления желания.

Спонтанная электрическая активность мозга, регистрируемая с помощью ЭЭГ, состоит из нескончаемой последовательности потенциалов, по форме не отличающихся от потенциала готовности. Само решение отождествить некоторый всплеск электрической активности с готовностью совершить движение носит статистический характер и в значительной степени является спекулятивным. В экспериментах с вызванными потенциалами для их выявления проводят усреднение большого количества повторяющихся предъявлений стимула, чтобы по возможности элиминировать электрическую активность, не привязанную по фазе к моменту предъявления стимула.

Неизбежная неточность в измерении как времени потенциала готовности, так и времени принятия решения в эксперименте Либета может легко объяснить разницу в 350 мс просто ошибкой измерения. Неизвестно также, каким когнитивным процессам соответствует регистрируемый на поверхности черепа потенциал готовности. Многие указывают, в частности, что зарегистрированная Либетом негативная волна может отражать неспецифическую мозговую активность, связанную с постоянным вовлечением испытуемого в задачу, но не является маркером подготовки к движе-

нию [Haggard, 2005]. Показано, например, что «потенциал готовности» регистрируется не только когда испытуемый решает совершить движение, но и когда он решает не совершать движения, при этом в том случае, когда он решает совершить движение, не выявляется латерализации этого потенциала в зависимости от того, какой рукой он собирается сделать движение [Trevena, Miller, 2010]. Все это говорит о том, что «потенциал готовности» не является маркером принятия решения совершить движение и ставит под сомнение те далеко идущие выводы, которые многие сделали из экспериментов Либета [Mele, 2009; O'Connor, 2009b; Nahmias, 2014].

Эксперименты Либета дали толчок большой серии исследований роли неосознанного и активности мозга в принятии сознательных решений. В этих экспериментах активность мозга регистрировали с помощью не только ЭЭГ, но и фМРТ. Серию таких работ провела группа немецких исследователей под руководством Джона-Дилана Хайнеса в 2007 г. В качестве примера можно разобрать одну из первых их работ, опубликованную в престижном журнале *Nature Neuroscience* и процитированную к настоящему времени уже 2000 раз [Soon et al., 2008].

Экспериментальная парадигма, использованная в этой работе, является вариантом эксперимента Либета с некоторыми вариациями, отчасти обусловленными спецификой фМРТ-эксперимента, отчасти желанием уточнить некоторые, оставшиеся неясными, моменты. Лежа в сканере, испытуемые (всего 36 человек) видели на экране сменяющуюся (раз в 500 мс) череду букв и в свободно выбранный момент, когда они «почувствовали побуждение сделать это», нажимали левую или правую кнопку, стараясь запомнить, какая буква была на экране в момент принятия решения. После этого им показывали четыре буквы, и они должны были выбрать ту, которая была на экране в момент X. Затем все повторялось несколько раз. Среднее время от начала трайла до нажатия кнопки было 21.6 с, левая и правая рука выбирались одинаково часто и последующий опрос показал, что, по мнению испытуемых, в среднем в 88.6 % случаев сознательное решение принималось за 1 с до нажатия кнопки.

Анализ фМРТ-данных проводился с использованием методов декодирования. Для каждого момента времени в период, предшествующий нажатию клавиши, вычислялось количество информации, которую можно извлечь из паттернов активности в различных областях мозга для предсказания сделанного выбора (нажатие

клавиши левой или правой рукой). В резюме статьи авторы пишут, что, по их данным, решение закодировано в активности префронтальной и теменной коры уже за 10 с до того, как оно появляется в сознании. По их предположению, это отражает работу сети высокоуровневых зон управления, которые начинают готовить предстоящее решение задолго до того, как оно входит в сознание. На рис. 2 в статье представлены графики распределения вероятностей сделанного выбора в разные моменты времени, рассчитанные исходя из активности разных областей мозга.

Как и можно было ожидать, наибольшей предсказательной силой обладала активность двигательной коры, зарегистрированная после момента, отмеченного испытуемыми как время сознательного решения. В этом случае предсказание было верным в 75 % случаев (напомню, что при случайных совпадениях уровень правильных предсказаний был бы 50 %). В периоды, предшествующие времени сознательного решения, предсказания на основе активности двигательной коры мало отличались от уровня случайных совпадений. Однако предсказания на основе активности некоторых других областей мозга в этот период времени были достоверно выше уровня случайных совпадений. Максимальная точность предсказания (около 60 %) найдена для латерального лобного полюса (*lateral frontopolar cortex*), несколько меньше (около 55 %) для медиального лобного полюса (*medial frontopolar cortex*) и задней части поясной извилины. На мой взгляд, для такого уровня предсказаний (достоверно, но ненамного выше уровня случайных совпадений) выражение «закодировано в активности префронтальной и теменной коры» выглядит слишком сильным. Можно говорить о том, что в период, предшествующий отмеченному испытуемыми как время принятия решения, в выявленных областях мозга происходят какие-то подготовительные процессы, наличие которых говорит не о том, какое решение будет принято, а о том, что вероятность одного решения (из двух возможных) немного выше, чем вероятность другого.

Интересен набор выявленных структур, которые, как показано в других исследованиях, участвуют в процессах именно сознательного выбора, таких как исполнительный контроль, свободный выбор и интроспективный анализ намерений [Sakai, Passingham, 2003; Vogt, Laureys, 2005]. Экспериментальная задача в этом исследовании – такая же нереалистичная, как и в оригинальном исследовании Либета. Очевидно, что по условию задачи испытуе-

мы должны сделать два не связанных друг с другом выбора. Они должны выбрать, какой рукой нажать клавишу и в какой момент это сделать (т. е. фактически выбрать букву из потока сменяющихся на экране букв). Поскольку в инструкции подчеркивалось, что нажатие клавиши должно произойти в момент предъявления выбранной буквы, то скорее всего испытуемые сначала готовили движение, а потом выбирали букву. Соответственно, то, что в работе фигурирует как время сознательного решения, – на самом деле время выбора буквы на экране, а выбор руки, с помощью которой движение осуществляется, происходил до этого.

Экспериментальное задание – очень нетривиальное и далекое от ситуации действительно свободных решений. Испытуемых инструктировали, что они свободны решить, какой рукой нажать клавишу, но как только появится осведомленность о принятом решении, они должны запомнить букву на экране. Такая формулировка уже подразумевает, что решение должно как бы откуда-то прийти в сознание. Испытуемый должен прислушиваться к своим ощущениям и понять, в какой момент он «чувствует побуждение» нажать клавишу правой или левой рукой. Сама задача – прислушиваться к приходящим из подсознания ощущениям – должна способствовать активации подсознания. Кажется правдоподобным поэтому, что предварительные стадии принятия решения происходят за пределами сознания. Это абсолютно не значит, однако, что в данном случае сознание не участвует в выборе решения. Все такого рода задачи начинаются с того, что испытуемый получает и осознает инструкцию и принимает сознательное решение ей следовать. Далее он сознательно дает себе задание делать то или другое (например, прислушиваться к своим ощущениям), и в выполнении этого задания вполне могут быть стадии, не контролируемые сознанием. Так же как во время прогулки, когда мы решаем повернуть на перекрестке налево или направо, мы не контролируем сознательно сокращение каждой мышцы, необходимое для выполнения этого решения.

Три года спустя та же исследовательская группа Хайнеса опубликовала результаты повторного исследования с аналогичной экспериментальной парадигмой [Bode et al., 2011]. В этот раз был использован более мощный сканер, позволивший получить лучшее пространственное и временное разрешение фМРТ-сигнала. Исследование было ограничено лишь областью лобного полюса,

который показал лучшую предсказательную силу в предыдущем исследовании. Анализ фМРТ-данных показал, что в среднем по всем испытуемым точность предсказания выбора на основе активности левого латерального лобного полюса увеличивалась по мере приближения момента осознания выбора, достигая максимума (57 %) за 1 с до него. Отметим опять, что максимальная точность предсказания ненамного превышает уровень случайных совпадений (50 %).

Самый интересный результат – выявленная динамика точности предсказания по отношению к моменту осознания решения. Точность предсказания начинает достоверно превышать уровень случайных совпадений за 9 с до принятия решения и увеличивается вплоть до последней секунды. Одновременно увеличивается однородность паттерна активации внутри левого латерального лобного полюса. Затем в момент принятия решения и сразу после него точность предсказания резко падает ниже уровня достоверности и остается на низком уровне вплоть до нажатия клавиши. Эта динамика хорошо укладывается в интерпретацию, согласно которой активность мозга в данном случае отражает работу подсознательных механизмов, цель которых – подготовить решение для передачи его в сознание, где, можно думать, решение окончательно и принимается.

Если бы решение принималось в подсознании, а в сознании возникала бы лишь иллюзия его принятия, то не было бы привязки точности предсказания к моменту осознания решения. По интерпретации авторов, в течение подсознательной фазы намерение медленно «эволюционирует» по направлению к окончательному варианту, и как только пересекается порог стабильности эволюционирующего паттерна, принимается сознательное решение, после которого паттерны активации теряют предсказательную силу [Ibid]. В конце статьи авторы предостерегают, что, хотя фМРТ и предоставляет уникальные возможности выявления паттернов активации мозга, конкретный механизм принятия решений не может быть выявлен только на основе фМРТ-данных.

В дальнейших работах авторы предполагают, что выявленная ими активация фронтальных областей коры отражает процессы сохранения в памяти отложенных намерений. Росто-медиальная префронтальная кора связана с сохранением намерений в условиях отвлечения на какую-то постороннюю активность, а росто-ла-

теральная – независимо от наличия отвлекающей деятельности [Momennejad, Haynes, 2013]. В еще одной работе авторы показывают, что аналогичные эффекты можно наблюдать не только для простых двигательных выборов (правая или левая рука), но и для абстрактных действий – «свободное» решение складывать или вычитать цифры [Soon et al., 2013]. В этом случае «предсказывающая» активация выявлена в медиальной префронтальной и затылочной коре, входящей в так называемую дефолтную сеть мозга (ДСМ). Максимальная точность предсказания (менее 60 %) была отмечена за 4–2 с до сознательного решения и падала ниже уровня достоверности после него. Активность ДСМ также резко падала после принятия сознательного решения.

Нужно отметить, что экспериментальные ситуации наподобие описанных выше трудно отнести к ситуациям, в которых выбор имеет хоть какое-то реальное значение. Не говоря даже о том, что участие в эксперименте не является, как правило, чем-то важным для испытуемых, предложенные в этих экспериментах варианты выбора (нажать кнопку правой или левой рукой, или в какой момент времени это сделать) во всех смыслах совершенно равнозначны. В повседневной жизни мы, как правило, не контролируем сознательно свое поведение в ситуациях, не являющихся важными для нас. Сознательный контроль, однако, всегда включается, если ситуация становится неопределенной или потенциально опасной. На пустой или хорошо знакомой дороге опытный водитель может вести увлекательную беседу с пассажиром, но он моментально прекращает ее и обращает все внимание на дорогу, как только обстановка неожиданно изменяется.

Все рассуждения о наличии или отсутствии свободы воли обычно относятся к ситуациям, в которых выбор имеет большие и важные для человека последствия, например, в вопросах морального выбора. Есть ли основания распространять на эти ситуации результаты экспериментов, в которых испытуемый, выполняя инструкцию экспериментатора, выбирает, какой рукой нажать кнопку? Легко представить, что, когда человек принимает трудное решение, взвешивая сознательно «за» и «против» разных вариантов, каждый из которых имеет серьезные мотивационные аспекты и важные последствия, он в большей степени будет руководствоваться сознанием, чем когда ему нужно прислушиваться к неважным и совершенно равнозначным побуждениям нажать кнопку левой или правой рукой [Mele, 2013].

В большом количестве реальных жизненных ситуаций каждый шаг оперативного поведения, безусловно, не контролируется сознанием. Часто поведение в таких ситуациях является автоматическим и регулируется механизмами процедурного знания. Иногда наши поступки бывают неконтролируемыми и импульсивными, однако задним числом мы, как правило, это понимаем и часто сожалеем об этом. Очевидно, что сознательный контроль оперативного поведения значительно его замедляет, в силу чего участие сознания в принятии решений часто сдвинуто во времени по отношению к реализации принятого решения. При принятии важных решений мы можем долго продумывать разные варианты и «проигрывать» наши ответы на возможные повороты ситуации, а потом в реальном поведении действовать на основе этих заготовок уже практически без участия сознания – так же, как опытный наездник не контролирует каждое движение скакуна и иногда «отпускает поводья», сохраняя при этом контроль над общим направлением движения.

Как было отмечено выше, сознание – это не просто осознание неосознанного, а качественно другой вид ментальности, часто связанный с другими структурами мозга. Каждый из возможных видов ментальности выполняет свои функции и их вряд ли можно противопоставлять друг другу. В принятии «сознательных» решений неизбежно участвуют мотивы и информация, пришедшие из подсознания, а принятые сознательно решения могут интернализироваться и «уходить» в подсознание, чтобы потом участвовать в оперативной организации поведения без прямого участия сознания. Традиция противопоставлять сознание и подсознание как две антагонистические силы пришла во многом из учения Фрейда, которое не выдержало проверки временем.

Современные данные, в том числе и результаты исследований свободы воли, показывают, что сознание и подсознание сосуществуют и тесно друг с другом связаны. В любой момент времени большинство ментальных процессов не осознается (можно говорить, что они протекают в подсознании), но результаты этой подсознательной черновой работы, если они достойны или нуждаются во внимании и оценке, поступают в сознание. Одна из функций подсознания – переработка больших объемов информации и сжатие их до приемлемого уровня, другая – работа с биологическими мотивациями, более древними в эволюционном смысле. Роль со-

знания состоит в принятии решений на основе информационной выжимки, пришедшей из подсознания, с учетом как мотивационных драйвов, так и сознательных целей и обязанностей.

Комментируя эксперименты Либета и подобные им, Людер Дик пишет, что сознательное свободное решение в этих экспериментах принимается еще до его начала, а потом черновая работа подготовки конкретного движения передается подсознанию. Удивительно, однако, то, что даже в таких простых случаях сознание включается в последние 200 мс, чтобы проконтролировать пришедшее из подсознания решение и остановить его в случае необходимости [Deecke, 2012]. Таким образом, свободное сознательное принятие решений можно считать специфически человеческой особенностью, лежащей в основе всех тех достижений, которые отличают человека от остального животного мира.

Анализируя теории о соотношении сознательных и неосознанных процессов, автор имплицитивного ассоциативного теста [Greenwald et al., 1998] делит их на три категории. К первой он относит теории, утверждающие примат подсознания, практически полностью определяющего наше поведение и содержание сознания. К этой категории прежде всего принадлежит теория Фрейда. Теории второй категории рассматривают сознание и подсознание как две параллельные независимые силы. Наконец, третья точка зрения, к которой склоняется и автор, рассматривает неосознанное как стадию обработки информации, имеющую место до того, как на нее обращается сознательное внимание и выполняющую таким образом подчиненные функции [Greenwald, 1997; Greenwald, Banaji, 2017].

Гринвальд пишет далее, что эмпирические данные трудно с легкостью уложить в любую из трех категорий, однако теории третьей категории могут быть полезной отправной точкой. В то же время хорошо известный вариант этих теорий, рассматривающий обработку информации и принятие решений как многостадийный последовательный процесс, не может предложить никакой модели того, как неосознанные процессы могут переходить в сложные сознательные суждения. Не стоит забывать и идею, согласно которой человеческое сознание возникло прежде всего как функция коммуникации. Действительно, пожалуй, единственная человеческая активность, осуществление которой невозможно без участия сознания, – передача информации другим людям. Для того чтобы передать свои мысли, чувства и впечатления другим людям, их

обязательно нужно пропустить через сознание. Чаще всего, но не обязательно, это происходит в форме слов. Развитие этой функции позволило аккумуляцию знаний не только в обществе современников, но и в ряду поколений и привело в конечном счете к появлению всех достижений, отличающих человека от остального живого мира.

В конце этой главы подчеркну еще раз, что, несмотря на тенденцию отождествлять неосознанные процессы с активностью мозга, дискуссии о роли сознания и подсознания в принятии решений не касаются, по сути, вопроса о роли мозга в данных процессах. Поскольку у нас нет прямого доступа к содержанию подсознания, мы можем судить о нем лишь косвенно на основе поведения, что затрудняет проверку предсказательной силы содержания подсознания в отношении поведения. Имплицитный ассоциативный тест позволяет в некоторой степени преодолеть это затруднение, однако не надо забывать, что в таком случае одно поведение предсказывается на основе другого. Тем ни менее исследования с использованием ИАТ показывают, что для некоторых видов поведения неосознаваемые мотивы являются лучшим предсказателем, чем эксплицитные интроспективные оценки [Egloff, Schmukle, 2002]. В других случаях сознательные намерения лучше предсказывают поведение [Webb, Sheeran, 2006].

В популярной литературе иногда встречаются утверждения, что активность мозга, измеренная, например, с помощью фМРТ, лучше предсказывает поведение, чем сознательные намерения самого человека, оцененные на основе его словесного отчета [Wolpert, 2010]. Эти утверждения обычно основаны на слишком смелой интерпретации эмпирических данных, из которых они выводятся. В качестве примера можно рассмотреть исследование группы Либермана, опубликованное в престижном журнале и прославившее поводом для статьи Волперта [Falk et al., 2010].

Исследование было проведено на 20 испытуемых, которых сначала спросили, как часто за последнюю неделю они пользовались солнцезащитным кремом и собираются ли они его использовать в следующую неделю. Затем в фМРТ-сканере им показали убедительные тексты и изображения о пользе солнцезащитного крема. Далее испытуемые еще раз ответили на вопрос, собираются ли они использовать солнцезащитный крем в следующую неделю, и при расставании им вручили в качестве подарка тюбик такого крема. Через неделю к ним обратились по электронной почте с

просьбой указать, сколько раз за прошедшую неделю они его использовали. Изменение намерений в результате просмотра текстов о пользе солнцезащитного крема оценивали разницей в количестве дней, в которые испытуемые намеревались использовать крем на следующей неделе в соответствии с отчетами, заполненными до и после просмотра, а изменение поведения – разницей в количестве дней, в которые испытуемые реально использовали крем в предшествующую неделю на основе отчета, заполненного перед сканированием и через неделю после него. Эти две переменные коррелировали друг с другом положительно, но недостоверно ($r = 0.17$), активность же МПК во время просмотра достоверно коррелировала с изменением поведения ($r = 0.49$) [Ibid].

Авторы делают вывод, что активность МПК при прослушивании текстов лучше предсказывает изменение поведения в результате убеждения, чем самоотчеты испытуемых. По их мнению, данный факт указывает на то, что реальный психологический эффект убеждения испытуемыми может не осознаваться. Но можно придумать и другие объяснения. Например, современные люди, привыкшие к тому, что их постоянно бомбардируют рекламой всяческих продуктов, не хотят казаться олухами, которые на глазах у экспериментатора изменили свое мнение в результате просмотра рекламных роликов. Поэтому, заполняя опросник сразу после просмотра, некоторые испытуемые могут стремиться показать взгляды, близкие к тем, которые они показывали до просмотра. В повседневном же поведении, когда их никто не контролирует, они могут поддаться услышанным аргументам. А может быть, кто-то из них просто решил использовать подаренный тюбик – «не пропадать же добру». Это вполне может объяснить низкую корреляцию между самоотчетами о намерениях и реальным поведением.

Психологи, работающие с опросниками, хорошо знают, что люди могут преднамеренно или непреднамеренно исказить информацию о своих поведении и качествах не потому, что у них нет сознательного доступа к информации, а потому, что они хотят представить себя в определенном свете. Вполне возможно, тем не менее, что поведение, которое в основном определяется неосознанными драйвами, содержание которых мы не контролируем, может лучше коррелировать с активностью мозга, чем с отчетом. Обычно, однако, это такое поведение, которое человек не считает особо важным (мазаться или нет защитным кремом, например).

В случае действительно важных решений человек, как правило, сознательно продумывает свои мотивы и действия и их возможные последствия.

В целом предсказательная сила активности мозга в отношении поведения несравненно меньше, чем предсказательная сила эксплицитных или имплицитных оценок содержания ментальности. Показано, например, что ретестовая надежность фМРТ-данных (т. е. воспроизводимость результатов при повторных экспериментах) гораздо ниже, чем для данных, полученных с помощью опросников или ИАТ [Elliott et al., 2020]. О чем говорит низкая ретестовая надежность в этом случае? Если повторяется один и тот же эксперимент с предъявлением одних и тех же стимулов и выполнением одного и того же задания, но сопровождающая выполнение задания активность мозга варьирует от эксперимента к эксперименту, это говорит о том, что связь между активностью мозга и поведением невелика.

Известно также, что в психиатрии изменения активности мозга не считаются патогномоничными признаками и не используются при постановке диагнозов. Психиатры до сих пор предпочитают полагаться преимущественно на традиционные методы наблюдения и опроса пациентов и их родственников. Если такие масштабные и заметные даже неспециалистам поломки ментальности, как психические болезни, не сопровождаются сопоставимыми изменениями в работе мозга, то трудно согласиться с утверждением, что активность мозга является источником ментальной жизни. Экспериментальные данные с использованием ЭЭГ или фМРТ лишь показывают, что активность некоторых областей мозга коррелирует с некоторыми когнитивными процессами. Нет, однако, даже намека на понимание конкретного механизма, с помощью которого мозг мог бы «принимать решения» или порождать содержание сознания или подсознания.

ПРЕДВАРИТЕЛЬНЫЙ ИТОГ

Кратко обобщив рассмотренный в предыдущих главах материал, попытаемся еще раз ответить на поставленный вопрос о соотношении ментального и активности мозга. В рамках когнитивной психологии и нейронаук можно выделить три предметные области, изучаемые разными методами и относящиеся, по-видимому, к разным явлениям. Назовем их: 1) поведение, 2) ментальность, 3) мозг. Первая, наиболее эмпирически доступная область – видимое поведение человека. Бихевиористы в свое время считали, что это – единственная доступная для научного исследования сфера. Разработано много методов, позволяющих количественно оценить разные аспекты поведения человека и животных.

Вторая область – то, к чему у нас часто нет прямого доступа: содержание ментальных процессов. Его можно разбить на сознательную и неосознанную части, но разбиение во многом будет искусственным, так как обычно не представляется возможным провести осмысленную границу между содержанием сознания и подсознания – они могут переходить друг в друга, и наличие какого-то содержания в сознании или подсознании зависит от многих факторов, таких, например, как направленность внимания. Разделение на сознание и подсознание важно по методическим причинам – о содержании сознания мы можем судить по самоотчетам испытуемого или по собственным интроспективным наблюдениям, а содержание подсознания можем оценивать лишь по косвенным поведенческим признакам. Существующие данные показывают, что большая часть ментальных процессов происходит за пределами сознания.

Сознание – это что-то вроде панели управления. В современном автомобиле, например, большая часть процессов регулируется автоматически, и только самая существенная информация выводится на панель управления, чтобы помочь водителю в управлении автомобилем. Организм человека – гораздо более совершенная машина, которая часто может двигаться «на автопилоте», вообще без участия «водителя». Ресурсы сознания ограничены, поэтому вся трудоемкая «низовая» работа происходит за пределами сознания и лишь ее результат может быть представлен на

рассмотрение сознания. Если принимается какое-то сознательное решение, требующее воплощения в поведении, то детали этого воплощения также перепоручаются подсознанию. Нет никакой причины, однако, рассматривать сознание и подсознание как две радикально различные или даже враждебные по отношению друг к другу силы.

Данные о функциях вентрального и дорзального зрительных трактов показывают, что содержание сознательной и подсознательной (используемой для управления поведением без участия сознания) ментальности может быть связано с разным нервным субстратом. Это ничего не говорит о том, как то и другое «кодируется» в активности, соответственно, вентрального и дорзального путей, но показывает, что (по крайней мере в случае зрительного восприятия) содержание сознания не может рассматриваться лишь как эпифеномен активности мозга, обеспечивающей управление поведением. Координация поведения на основе зрительной информации (например, как взять чашку, стоящую на подносе, и переставить ее на середину стола) происходит за пределами сознания и коррелирует с активностью дорзального пути, а понимание смысла зрительной информации (например, выделение из потока зрительных стимулов чашки, подноса и стола как отдельных объектов) является осознанным и коррелирует с активностью вентрального пути. Можно думать поэтому, что осознанное восприятие и категоризация зрительных стимулов должны иметь функциональное значение, не сводимое к регуляции моторного поведения.

Наконец, третья область – активность мозга, которая во времена бихевиористов была практически недоступна для изучения, а теперь активно исследуется с помощью неинвазивных методов, таких как фМРТ, ПЭТ, ЭЭГ, МЭГ и прочих, включая разные методы транскраниальной стимуляции и торможения мозга. Вопрос о том, как указанные три области связаны друг с другом, и является предметом дебатов в философии и когнитивных науках. Наше субъективное восприятие и вся история человеческой цивилизации согласуется с представлением, согласно которому содержание ментальности определяет поведение.

Можно спорить, в какой степени каждый конкретный поведенческий акт определяется сознанием или подсознанием, или в какой степени содержание сознания определяется содержанием подсознания – в контексте настоящего обсуждения это, собственно, не имеет значения. По методическим причинам легче просле-

живать связь между поведением и содержанием сознания, и весь наш субъективный опыт и история человечества показывают наличие и эффективность каузальной связи между содержанием сознания и поведением. Да, в некоторых случаях нам не удастся осуществить то, что мы сознательно планировали, или, наоборот, мы совершаем поступки, которых от себя не ожидали и о которых впоследствии часто сожалеем. Иногда это связано с действием неучтенных внешних факторов, иногда с недооценкой собственных подсознательных мотивов, но, как правило, мы впоследствии можем это понять и учесть при планировании дальнейших действий. В целом у большинства здоровых людей сохраняется уверенность в эффективности сознательного контроля собственного поведения. Если бы не было каузальной связи между содержанием сознания и поведением, мы бы воспринимали собственное поведение как что-то от нас независящее. Такое, однако, наблюдается лишь в редких случаях, обычно связанных с тяжелой психопатологией.

Адепты редукционизма утверждают, что субъективное чувство эффективности сознательного контроля – это лишь иллюзия, формируемая путем фабрикации фальшивого нарратива, который задним числом увязывает концы с концами и представляет события в нужном свете [Dennett, 1991]. Например, момент сознательного принятия решения проецируется назад во времени и искусственно привязывается к моменту, когда мозг уже принял решение, но сознание об этом еще ничего не знало [Libet, 2002]. Причины и следствия поступков подтасовываются или ложно интерпретируются, чтобы создать иллюзию эффективности сознательного контроля. Трудно отрицать, что в некоторых случаях люди могут не отдавать себе отчета в истинных мотивах некоторых своих поступков и задним числом объяснять их сознательным решением. Это, однако, относится к взаимодействию между сознанием и подсознанием, а не между ментальностью и мозгом. Стронникам редукционизма удобно отождествлять содержание подсознания с активностью мозга, однако оснований для этого не больше, чем для отождествления с активностью мозга содержания сознания. Когда при прогнозе поведения кроме содержания сознания можно учесть и содержание подсознания, точность предсказания повышается [Egloff, Schmukle, 2002], хотя она вряд ли может во всех случаях быть стопроцентной, если учесть вариацию внешних факторов и не всегда предсказуемое взаимодействие этих факторов с содержанием сознания и подсознания.

В соответствии с интерпретацией физикалистов нужно думать, что фальшивый нарратив конструируется мозгом, поскольку, по этой интерпретации, кроме мозга, ничего и нет. Закономерный вопрос – кого тогда нужно обманывать? Получается, что мозг не только совершает всю реальную работу управления поведением, но и дополнительно берет на себя заботу конструирования фальшивого нарратива, чтобы обмануть самого себя. В физикалистских концепциях сознание либо вообще не существует (иллюзия), либо является эпифеноменом, не обладающим каузальной силой. Мы уже обсуждали логическую непоследовательность концепции иллюзионизма. Кто или что является субъектом, у которого возникает иллюзия сознания? По логике материалистического монизма субъектом должен быть мозг. Значит, у мозга возникает иллюзия, что он обладает сознанием, и, чтобы поддерживать эту иллюзию, он постоянно конструирует нарратив, согласно которому у иллюзорного сознания есть каузальная сила, хотя на самом деле она есть лишь у самого мозга, а не у возникающей в мозгу иллюзии.

Если сознание – не иллюзия, а эпифеномен, то оно должно автоматически сопровождать работу мозга, как тень, отбрасываемая предметом. Для чего в данном случае должен создаваться нарратив, создающий иллюзию каузальной силы сознания? Это похоже на трюк фокусника, который проецирует тень своих рук на экран и создает у зрителей иллюзию, что тень представляет собой каких-то зверюшек, совершающих некие действия, хотя на самом деле все действия совершают за экраном руки фокусника. Если, по логике материалистического монизма, человек является продуктом естественного отбора наиболее приспособленных, то необходимым и достаточным было бы лишь появление эффективного, обладающего каузальной силой мозга. Возникновение сознания, а тем более специальных механизмов для конструирования вводящего в заблуждение нарратива о каузальной эффективности этого сознания, трудно объяснить.

Обсудим теперь вопрос: в какой степени представленные выше эмпирические данные и теории мозгового обеспечения ментальных процессов согласуются с физикализмом и постулируемой им эпифеноменологической природой сознания. Напомним, что, согласно этой концепции, каузальной силой обладает лишь мозг, а не сознание. Поскольку, как мы обсудили выше, у нас есть (иллюзорные по физикалистской интерпретации) субъективные впечат-

ления о каузальной эффективности сознания, то для доказательства их иллюзорности демонстрации наличия корреляций между нервной активностью и содержанием сознания недостаточно, так как подобные корреляции можно объяснить и в рамках дуалистических представлений. Физикалистские нейронные теории должны объяснять природу и механизм этих корреляций и показать, что постулируемый нервный субстрат идентичен, или, по крайней мере, понятным образом соответствует феноменологии сознания [Metzinger, 2000].

Рассмотренные нами эмпирические данные показывают, что связь между активностью мозга и феноменологией сознания имеет статистический характер и выявляется лишь в виде корреляций, к тому же не очень больших. Ни эти эмпирические данные, ни сформулированные теории сознания не позволяют объяснить природу и механизм этих корреляций. Всегда можно думать, конечно, что это положение дел преходящее и объясняется лишь недостаточностью нашего методического арсенала выявления активности мозга. Пройдет время, появятся новые методы и удастся показать, что некий нервный субстрат идентичен, или, по крайней мере, понятным образом соответствует феноменологии сознания.

К сожалению, уже накопленные данные о характере активности мозга, сопровождающей когнитивные процессы, позволяют в этом усомниться, так как показывают принципиальное различие между природой нервной активности и ментальности. Связь активности мозга с феноменологией ментальной активности носит статистический характер. Мозг представляет сенсорную информацию в вероятностных терминах. Например, на уровне отдельных нейронов или вокселей в фМРТ-образе мозга направление движения объекта представлено в виде функции распределения вероятностей, в то время как содержание чувственного образа не имеет вероятностного характера. Для извлечения из активности мозга информации об ориентации зрительных образов нужно применить изошренный математический аппарат. Никому пока не удалось показать наличие в мозге «машинерии» для производства подобных вычислений. Кроме того, с эпифеноменологической точки зрения, если бы мозг эти вычисления как-то проводил, их результат должен был проявиться в активности какой-то части мозга, и эта активность должна была уже иметь не вероятностный, а бинарный характер, соответствующий содержанию чувственного образа. До сих пор никому не удалось найти в мозге то место, где

появляется результат подобных вычислений, если не считать концептуальные нейроны гиппокампа, но их активность, как мы уже говорили, отражает не содержание сознания, а лишь тот факт, что воспринимаемый объект опознан.

За последние годы опубликовано множество работ по «кодированию» сенсорной информации в мозге. В этих работах, которые обычно используют методы фМРТ или ЭЭГ, часто утверждается, что регистрация и анализ активности мозга позволяет «прочитать» содержание сознания в момент восприятия внешнего мира. Довольно часто, как показано в статье Ли с соавторами [Li et al., 2021], в них есть методические погрешности, которые искусственно вызывают появление корреляций между активностью мозга и типом предъявляемых стимулов. Но даже если использованные методы получения данных и их анализа безупречны, данные лишь показывают, что из активности мозга можно извлечь информацию о семантических признаках воспринимаемого объекта, что не удивительно, так как даже с позиции дуализма мозг – это аппарат, предназначенный для взаимодействия ментальной сферы с материальным внешним миром, и как таковой он должен собирать и представлять в активности нейронов всю информацию, необходимую для распознавания смысла воспринимаемых объектов. Полученные данные, однако, ничего не говорят о том, как эта информация извлекается в мозге и как она используется для распознавания смысла.

Вариативная природа активности мозга хорошо видна, когда данные нейрофизиологического эксперимента не просто усредняются, а анализируются индивидуально. При общепринятом подходе к анализу данных вариативность рассматривается как шум и задачей считается извлечение из этого шума «полезной» информации. Стандартом является применение статистических методов, позволяющих выявить «достоверные» эффекты. Таковым считается эффект, если вероятность его случайного появления ниже какого-то общепринятого порога (обычно 5 %). Хорошо известно, что как психологические, так и нейрональные эффекты никогда не воспроизводятся один к одному у разных людей. Поэтому принято собирать данные на выборке испытуемых. Есть рекомендации статистиков, как такие выборки должны формироваться, но на деле мало кто следует этим рекомендациям, и выборки, как правило, формируются по принципу удобства. К тому же подавляющее большинство данных в области психологии и нейронаук получе-

но на студентах преимущественно психологических факультетов, просто потому, что исследователям, которые обычно работают в университетах, проще пригласить в эксперимент собственных студентов, чем искать испытуемых на стороне. Часто студенты вынуждены участвовать в эксперименте, так как иначе их не допустят к экзамену, или за участие им обещают какие-то послабления в учебной программе. Вопрос о том, в какой степени полученные на таких испытуемых данные можно распространять на всю популяцию, нередко поднимается, но серьезной проверки в обозримом будущем вряд ли можно ожидать.

Но состав выборки в данном случае не главный вопрос. «Достоверность» выявляемых эффектов зависит от величины эффекта, размера выборки и уровня вариации изучаемого показателя. При относительно небольших размерах выборки (от 10 до 50 человек), которые доминируют в нейровизуализационных исследованиях, некий эффект может быть «достоверным» не только потому, что он есть у всех испытуемых, но и в тех случаях, когда он сильно выражен у некоторых и практически отсутствует у остальных. Статистики, опять же, рекомендуют в каждом случае анализировать распределение эффекта в выборке, но далеко не всегда это делается. В тех случаях, когда внимание обращается на эффекты, наблюдаемые у каждого отдельного испытуемого в процессе одного и того же эксперимента, видно преобладание различий между испытуемыми в активности мозга над элементами сходства. Так, Базиле с соавторами в серии работ показывают, что при выполнении разных когнитивных задач паттерн активации мозга, выявляемый с помощью локализации источников электрической активности, настолько различен у разных испытуемых, что не удастся найти ни одной области, которая была бы специфична для когнитивной задачи и была бы общей для всех испытуемых [Basile et al., 2021]. Наблюдения дополняются результатами метаанализа экспериментальных фМРТ-данных, показывающего их низкую ретестовую надежность [Elliott et al., 2020]. Это говорит о различии активности мозга, сопровождающей выполнение одного и того же задания, не только у разных испытуемых, но и у одного и того же испытуемого при повторных экспериментах.

Конечно, можно думать, что и содержание ментальности в процессе одного и того же эксперимента может различаться как у разных испытуемых, так и у одного и того же человека при повторных тестированиях. Испытуемые могут думать о посторонних

вещах, или быть в разных эмоциональных состояниях. Однако эта вариация не зависит от содержания экспериментального задания и в принципе может контролироваться, например, путем опроса испытуемых после эксперимента [Knyazev et al., 2012]. Связанное с содержанием экспериментального задания содержание ментальности должно мало отличаться для большинства испытуемых в одном эксперименте или при его повторениях, иначе они не смогли бы его успешно выполнять. Это, таким образом, еще одно доказательство того, что содержание ментальности гораздо более определено и лучше связано с поведением, чем активность мозга.

Вторая характеристика активности мозга, сопровождающей ментальные операции, – ее распределенный характер. Чем сложнее когнитивный процесс, тем большее количество областей мозга активируется. Они, как правило, расположены далеко друг от друга (например, лобные и теменные области) и участвуют во многих достаточно разнородных когнитивных процессах, т. е. они неспецифичны. Метаанализ значительного количества экспериментов с записью фМРТ в процессе выполнения разнообразных когнитивных задач (1138 задач по 11 когнитивным доменам: выполнение действия, наблюдение действия, подавление действия, внимание, слуховое восприятие, зрительное восприятие, эмоции, лингвистическая семантика, рассуждение, эксплицитная семантическая память и рабочая память) показал, что большинство областей мозга активируется в большинстве случаев [Anderson, Penner-Wilger, 2013]. Используя шкалу от 0 (область участвует лишь в задачах из одного когнитивного домена) до 1 (область участвует во всех когнитивных доменах), авторы вывели среднюю вовлеченность для 78 больших областей мозга – она составила 0.7. По их убеждению, локальные нейронные цепи не очень избирательны и обычно участвуют во множестве задач из разных когнитивных доменов [Ibid].

Но, может быть, специфику надо искать не в паттернах активации, а в паттернах функциональных связей между отдельными областями мозга? В последние 20 лет стало популярным изучение функциональных связей между областями мозга на основе статистической взаимозависимости их активности в течение какого-то времени. Это направление привело к открытию так называемых сетей покоя; каждая из них представляет набор областей мозга, активность которых коррелирует друг с другом даже в состоянии покоя [Biswal et al., 1995]. Накоплено немало эмпирических данных, показывающих связь активности указанных сетей с определенными

ми когнитивными процессами или эмоциональными состояниями, но связь эта опять же имеет статистический характер. Каждый устойчивый пространственный паттерн коактивации сопряжен с большим количеством схожих, но все-таки достаточно разнообразных процессов и состояний.

В уже упомянутом выше метаанализе Андерсена и Пеннер-Уилгера замечено, что для функциональных связей показатель вовлеченности меньше, чем для активации, т. е. при выполнении задач из разных когнитивных доменов одна и та же область мозга может взаимодействовать с разным набором других областей [Anderson, Penner-Wilger, 2013]. Однако часто наблюдается похожесть паттернов коактивации при выполнении задач, казалось бы, очень непохожих друг на друга (например, движение пальцев и счет в уме). По предположению авторов, это свидетельствует о том, что в процессе эволюции появление высших когнитивных функций, таких как язык и мышление (включая математическое), не сопровождалось появлением новых специализированных модулей, а использовало модули, исходно предназначенные для других целей [Ibid]. Если это действительно так, то можно сделать следующее заключение: движущей силой эволюции когнитивных функций человека было не изменение мозга, а предшествовавшее ему появление потребности в определенной когнитивной функции. Эта потребность удовлетворялась путем рекомбинации связей уже существовавших в мозге модулей. В любом случае, изучение паттернов коннективности, так же как и паттернов активации, не дает надежды на то, что их удастся связать с семантикой, т. е. специфическим содержанием ментальности.

В области моделирования когнитивных процессов рассмотренные нами ранее вычислительные теории разума (КВТР и КонТ), пожалуй, дальше всего продвинулись в плане постулирования конкретных, поддающихся математическому описанию механизмов, но они моделируют, собственно, ментальные механизмы, а не активность мозга. Даже если забыть про все проблемы, с которыми эти теории пока не могут справиться при моделировании ментальных процессов на основе алгоритмов, попытки аппроксимировать КВТР и КонТ модели (например, разные варианты нейронных сетей) на активность мозга очень далеки от того, что известно о реальной природе нервной активности. И, что важнее всего, все попытки объяснить природу человеческой ментальности на основе алгоритмов игнорируют содержание мысли. Поскольку алго-

ритм – это лишь формальные правила манипулирования символами, содержание которых не имеет значения для работы системы и, соответственно, не учитывается, то все вычислительные теории разума подчеркивают примат синтаксиса над семантикой, что противоречит нашему субъективному опыту, показывающему каузальную силу именно содержания мыслей.

Таким образом, приходится признать, что на сегодняшний день активность мозга и содержание ментальности остаются отдельными сущностями. Они, безусловно, коррелируют друг с другом, но, как одно возникает из другого – остается тайной. Содержание ментальности (включая ее сознательную и неосознаваемую части) коррелирует с поведением, рассматриваемым широко как ориентация в окружающем мире, принятие решений и их осуществление, и, как нам кажется, каузально с ним связано. Активность мозга также коррелирует с поведением, но в тех случаях, когда у нас есть доступ к содержанию ментальности, можно видеть, что поведение сильнее коррелирует с содержанием ментальности, чем с активностью мозга. И, что самое главное, несмотря на большое количество теоретических моделей и немалое количество эмпирических данных, пока нет понимания того, как мозг может реально осуществлять координацию и интеграцию всех необходимых процессов и принятие решений.

У Декарта функции контроля и управления отданы нематериальной душе. В современных физикалистских теориях данные функции приписывают каким-то процессам в мозге, однако суть этих процессов и их локализация остаются предметом споров. Некоторые теории, наподобие теории множественных набросков Деннетта [Dennett, 1991], в своей основе отрицают существование центрального регулятора и утверждают, что видимый порядок возникает спонтанно из свободной конкуренции модулей. Другие приписывают функции контроля процессам внимания, или рабочей памяти, или специальным функциональным сетям, охватывающим как фронтальные, так и парietальные области коры мозга. Нет ни одной теории, которая могла бы правдоподобно объяснить природу и механизмы интеграции как сознательных, так и неосознанных ментальных процессов в то, что на субъективном уровне мы воспринимаем как собственную личность. Исследование природы и мозговых коррелятов «Я» (the self) – пожалуй, одна из наиболее противоречивых областей психологии и нейробиологии, в которой мнения варьируют от полного отрицания реальности «Я»

[Hood, 2012] до приравнивания его к активности небольшого набора срединных областей коры [Northoff et al., 2006], или констатации сложной и составной организации этого конструкта [Gallagher, 2013].

Конечно, всегда можно думать, что завтра найдут что-то такое, что все поставит на свои места и позволит полностью объяснить, как мозг продуцирует ментальность и управляет поведением, но это вряд ли можно рассматривать как серьезный аргумент, и на уровне сегодняшнего знания уже можно сделать какие-то выводы.

1. Постулаты «сознание – это иллюзия» и «сознание – это эпифеномен активности мозга» не согласуются прежде всего с наличием в мозге структур и процессов, специфически связанных с сознанием. Это и существование отдельных дорзального (для неосознаваемого поведения) и вентрального (для сознания) зрительных и слуховых путей, и наличие концептуальных нейронов в МВК, которые активируются только в ответ на осознанное узнавание объекта, и тот факт, что предсказательная сила активности мозга в отношении поведения нарастает вплоть до принятия сознательного решения и затем резко падает. Заметим, что эти процессы не отражают один к одному содержания сознания, как того требовала бы концепция эпифеноменализма, но показывают, что сознание имеет самостоятельное значение, отдельное от неосознаваемого контроля поведения.

2. На современном уровне знания нет возможности объяснить, как процессы, происходящие в мозге, могут порождать содержание сознания; как осуществляется категоризация сенсорных образов; как извлекается смысл и происходит узнавание образа; как принимается сознательное решение. Безусловно, имеется немалое количество гипотез по поводу возможных мозговых механизмов, которые могли бы обеспечивать указанные процессы, но эти гипотезы (помимо отсутствия подтверждающих их эмпирических данных) сталкиваются с рядом принципиальных трудностей. Работы по изучению кодирования информации при зрительном восприятии показывают, что по активности нейронов на конечных уровнях зрительного анализатора можно с некоторой точностью предсказать идентичность воспринимаемого объекта. В большинстве опубликованных работ она достоверно выше уровня случайных совпадений, но не превышает 75 %. При этом отдельные нейроны кодируют не идентичность стимула, а некоторое его свойство, и сила ответа нейрона распределена между похожи-

ми свойствами по закону Гауса [Chang, Tsao, 2017]. Какие механизмы в мозге обеспечивают это кодирование – пока остается предметом спекуляций, так же как и то, как мозг извлекает смысл из данной информации.

Рассмотрим знакомый многим пример. Представьте, что вы приехали в чужой город и прогуливаетесь по площади, наполненной толпами туристов. Вы рассеянно посматриваете по сторонам, и вдруг в какой-то момент с удивлением узнаете в толпе своего старого знакомого, которого никак не ожидали встретить в этом месте. Оpozнание происходит за доли секунды, хотя в этом случае предсказывающее кодирование вряд ли может помочь. Распознаванию лиц посвящено множество исследований, многие из которых мы уже рассмотрели выше. Считается, что лица обладают особым статусом в иерархии визуальных стимулов в силу их значимости в эволюции человека, как существа социального. Соответственно, в зрительном анализаторе есть области, специализирующиеся именно на восприятии лиц, например, лицевая область в веретенообразной извилине [Grill-Spector et al., 2004]. Как мы видели выше, извлечение признаков, по которому можно идентифицировать лицо, происходит в зрительном анализаторе за первые 100–200 мс, а уже через 300 мс концептуальные нейроны МВК регистрируют сознательное опознание лица [Quian Quiroga, 2016]. Как по извлеченным признакам (механизмы их извлечения в мозге пока также остаются тайной) за 100–200 мс происходит опознание?

Сейчас достаточно широкое распространение получили компьютерные приложения, позволяющие распознавать лица. Эффективность этих приложений пока в целом уступает эффективности человека, но они постоянно совершенствуются. Наиболее эффективные приложения используют нейронные сети глубокого обучения, как, например, созданная Facebook система DeepFace, которая распознает человеческие лица на цифровых изображениях. Она использует девятислойную нейронную сеть с более чем 120 млн соединений и была обучена на 4 млн изображений, загруженных пользователями Facebook. Прирост эффективности этой системы по сравнению с аналогичными был достигнут благодаря возможности использовать огромные объемы внешних данных в модели обучения [Simonite, 2014]. DeepFace обрабатывает изображения лиц в два этапа. Сначала она корректирует угол лица так, чтобы человек на снимке смотрел вперед, используя трехмерную

модель «среднего» лица, смотрящего вперед. Затем начинается глубокое обучение, когда моделируемая нейронная сеть разрабатывает численное описание переориентированного лица и сравнивает это описание с оцифрованными изображениями, хранящимися в базе данных. Если DeepFace обнаруживает достаточно похожие описания двух разных изображений, она решает, что они должны отображать одно и то же лицо [Simonite, 2014].

По существующим гипотезам что-то подобное оцифровке изображения происходит и в мозге [Chang, Tsao, 2017]. Следующий этап – сравнение изображения с изображениями, хранящимися в «базе данных», т. е. в этом случае – в семантической памяти. Можно только спекулировать по поводу того, сколько в среднем изображений разных лиц хранится в памяти человека. Это количество может быть достаточно большим, хотя и наверняка меньшим, чем в базе данных DeepFace. Если для ориентира опереться на гипотезу Дунбара, то можно предположить, что в среднем современный человек способен (исходя из средних размеров человеческого мозга) поддерживать стабильные отношения со 150 другими людьми [Purves, 2008], т. е. он должен помнить лица этих людей. По современным представлениям изображение каждого лица хранится в виде энграммы памяти, т. е. в виде распределенного ансамбля связанных друг с другом нейронов. Сколько нейронов может входить в такой ансамбль и в каком виде хранится информация об идентичности лица – пока остается тайной. Давайте для удобства считать, что ансамбль, кодирующий идентичность одного лица, состоит из 200 нейронов и что информация хранится в том же виде, в каком она закодирована в активности нейронов НВИ при восприятии лица. Во всяком случае, 200 нейронов потребовалось в работе [Chang, Tsao, 2017], чтобы декодировать идентичность лица с 75%-й надежностью.

Считается, что при воспроизведении следа памяти происходит реактивация нейронного ансамбля, хранящего энграмму, и что нейронные ансамбли, активируемые при воспроизведении следа памяти, в значительной степени перекрываются с нейронными ансамблями, активируемыми при восприятии и запоминании [Richards, Frankland, 2013; Palacio, Cardenas, 2019]. Как происходит сверка информации, извлеченной из памяти, с информацией, кодируемой зрительным анализатором, мы не знаем. Давайте представим, что в мозге есть какой-то компаратор, который тоже содержит 200 нейронов, каждый из которых одновременно получает

сигналы от соответствующих нейронов зрительного анализатора и активированной энграммы памяти. При этом сила его ответа должна быть обратно пропорциональна величине рассогласования входных сигналов. Чтобы это выполнялось, потребуется хотя бы один вставочный нейрон. Далее величина рассогласования должна как-то суммироваться по всем 200 нейронам компаратора и временно храниться в каком-то отдельном модуле до окончания сравнения со всеми 150 лицами. Затем выбирается лицо с наименьшим показателем рассогласования и уровень этого показателя сравнивается с некоторым (не обязательно жестко заданным) порогом. Если уровень рассогласования ниже порога, воспринимаемое лицо идентифицируется с соответствующим следом памяти и в сознание поступает информация об успешном опознании.

Чтобы сравнить одну энграмму с набором входных стимулов, компаратор или какой-то другой модуль в мозге должен как-то активировать эту энграмму, что потребует как минимум послать один сигнал к каждому из 200 нейронов энграммы. Нейроны энграммы могут ответить пачками импульсов, частота которых будет кодировать выраженность соответствующей характеристики у кодируемого лица. Эта частота может варьировать, например, от 0 до 50 спайков в секунду [Chang, Tsao, 2017]. Примем среднее за 25. Таким образом, считывание информации одной энграммы потребует 200 синаптических событий на ее входе и некоторое количество, кратное 200, на выходе. Количество выходных спайков зависит от времени, выделяемого на считывание одной энграммы. Понятно, что чем меньше это время, тем хуже будет надежность оценки для медленных ритмов. Допустим, что оцениваются лишь те черты, которые выражены выше среднего уровня, т. е. нейроны, импульсирующие с частотой ниже 25 Гц, просто игнорируются. Чтобы зафиксировать импульсацию с частотой 25 Гц, нужно как минимум 40 мс. Тогда для извлечения информации из 150 энграмм, если производить эту операцию последовательно, потребуется 6 с, что неприемлемо. Можно допустить, что извлечение информации из всех 150 энграмм производится параллельно, но для этого нужно иметь 150 компараторов (каждый из 200 нейронов).

Понятно, что все это лишь фантазии, и я привожу их здесь лишь для иллюстрации сложности технической проблемы реализации вычислений в активности мозга. Как отмечают Галлистел и Кинг [Gallistel, King, 2009], нейробиологи не знают с уверенностью ответов даже на самые простые вопросы о том, как выполня-

ются вычисления в нервной ткани. Нет единого мнения о том, что из себя представляют даже примитивные вычислительные операции, из которых строятся все другие вычисления, хотя этот вопрос является первостепенным для инженеров, конструирующих вычислительные машины. Также нет единого мнения по такому основному предварительному вопросу: каков механизм кодирования информации в последовательности импульсов нейрона. Реализация вычислительных операций над символами принципиально зависит от формы, в которой информация в этих символах закодирована. Если нет единого мнения о том, как значения простых переменных закодированы в цепочке импульсов, тем более нет надежных знаний о том, как вычислительные механизмы в мозге объединяют информацию из двух или более источников. Альтернатива состоит в том, что элементарные вычислительные операции выполняются внутри отдельных нейронов либо на молекулярном уровне, либо субклеточными структурами, подобными дендритным шипикам [Koch, 1999], однако это ставит вопрос о том, как результаты этих элементарных вычислений передаются и объединяются с результатами других вычислений.

Считается несомненным, что потенциал действия – это механизм, с помощью которого информация передается на «большие» расстояния (более 10 мк). Однако важно понимать, насколько медленна эта передача по сравнению с тем, с чем привыкли работать создатели вычислительных машин. Чтобы пройти 100 мк (типичное расстояние, разделяющее нейроны в локальной цепи, такой как колонка в коре), потенциалу действия требуется порядка 10–100 мс, что в 100 млн раз больше, чем время прохождения такого же расстояния в цепях компьютера [Gallistel, King, 2009]. Если к этому добавить время синаптической задержки (0.5–1.0 мс) [Byrne, 2020], то становится понятным, что реализацию описанного выше гипотетического механизма распознавания лиц невозможно уложить в те временные отрезки, за которые все реально происходит.

Невозможно понять, как работающие так медленно нейроны могут выполнить такие сложные операции с той скоростью, с какой это реально происходит [Gallistel, King, 2009]. Мозг может, например, обрабатывать до стадии распознавания и анализа потоки разрозненных изображений («пикник», «свадьба», «бейсбол» и т. д.), приходящих по одному изображению каждые 100 мс [Potter et al., 2004]. В настоящее время мы не знаем ни одного алгоритма, способного распознавать и анализировать содержание подобных

сложных изображений хотя бы приблизительно так эффективно, как это делает человек, даже если компьютеру дать любое нужное ему время. Для выполнения этой задачи (восприятие образа, его распознавание и анализ) любой отдаленно адекватный алгоритм требует доступа к большому количеству разнообразных источников информации и выполняет большое количество высокоуровневых вычислений, каждое из которых включает большое количество элементарных вычислений. Как может такое медленное устройство, как мозг, обрабатывать 10 сцен за секунду – большая загадка.

Современные нейросетевые модели представляют собой модели распознавания или категоризации. Они моделируют память с адресацией по содержанию. В памяти ввод части ранее испытанного входного вектора запускает паттерн активности, характерной для категории, к которой вектор и принадлежит. Переход в этот характерный паттерн активности считается реализацией распознавания или категоризации. Данный переход в стабильный паттерн активности (так называемое состояние аттрактора) является примером вычислительного механизма, который фундаментально зависит от повторяющейся (рекуррентной) передачи сигнала между тысячами нейронных элементов. Никто не знает, как система, основанная на этом механизме, может классифицировать изображения со скоростью 10 единиц в секунду.

Стандартный ответ на вопрос, как мозг может так быстро выполнять такие сложные вычисления, – это «массовый параллелизм», при котором проблема разбивается на бесчисленное количество очень простых вычислений, выполняемых одновременно множеством разных микромодулей. Однако на современном уровне развития когнитивной нейробиологии такой ответ преимущественно является прикрытием незнания. Неизвестно, в какой степени хотя бы в принципе можно разбить приведенную здесь и многие другие вычислительные задачи на элементы, выполняемые одновременно. Еще меньше известно, как такое разбиение может происходить в мозге, как части задачи могут переадресовываться на соответствующие микромодули и как результаты, полученные от всех этих микромодулей, могут быть сопоставлены и интегрированы для получения ответа, причем все это с такой скоростью, чтобы обеспечить пропускную способность в 10 изображений в секунду.

Специалисты компьютерных наук, изучая массово параллельные вычисления в течение десятилетий, уже пришли к понима-

нию, что, за исключением некоторых специфических случаев, их реализовать непросто. Значит, если массовый параллелизм – это ответ, то возникает вопрос, какова архитектура мозга, которая позволяет реализовать массовые параллельные вычисления. Вопрос в настоящее время остается без ответа. Есть еще следующая загадка: современный компьютер выполняет элементарные вычислительные операции в миллиард или триллион раз быстрее, чем могут работать механизмы, основанные на аксональной и синаптической передаче. Выполнение вычислений параллельно, а не последовательно обменивает ресурсы пространственные (количество вычислительных механизмов, работающих над различными частями задачи в любой момент времени) на временные (количество различных частей проблемы, которые один механизм может обработать за единицу времени).

Поскольку временные ресурсы современного компьютера на много порядков больше, почему он не может превзойти мозг, выполняя такое же количество последовательных операций за меньшее время, чем мозг может сделать их все одновременно параллельно? Галлистел и Кинг заключают, что либо мы серьезно заблуждаемся относительно скорости, с которой нервная ткань может выполнять элементарные вычислительные операции, либо мозг «знает» более эффективные алгоритмы, минимизирующие необходимое количество операций [Gallistel, King, 2009].

Вспомним два варианта эпифеноменализма – редукционизм и слабый эмерджентизм – и попробуем оценить соответствие выводов, вытекающих из этих философских позиций, рассмотренным выше, а также философские определения редукционизма. «Основной вопрос редукции заключается в том, могут ли свойства, концепции, объяснения или методы из одной научной области (обычно на более высоких уровнях организации) быть выведены или объяснены с помощью свойств, концепций, объяснений или методов из другой области науки (обычно на более низких уровнях организации)» [Brigandt, Love, 2017, p. 1]. «Утверждение, что x сводится к y , обычно означает, что x есть не что иное, как y ». «Если кто-то утверждает, что ментальное сводится к физическому... он подразумевает, что... ментальное может быть возвращено к физическому» [van Riel, van Gulick, 2019, p. 1].

Рассматривая недавно возникшее направление в моделировании ментальных процессов на основе формализма квантовой теории (так называемое квантовое познание), мы видели, что мен-

тальные процессы и процессы в мозге подчиняются разным законам. Классическая теория вероятности и булева логика хорошо описывают активность мозга, но для описания когнитивных процессов лучше подходит формализм квантовой теории и небулева логика [Wang et al., 2013]. То есть «свойства, концепции, объяснения и методы» в области изучения ментальных процессов не могут быть «выведены или объяснены» с помощью свойств, концепций, объяснений и методов из области нейробиологии. Из имеющихся у нас данных мы не можем сказать, что x (ментальная сфера) есть не что иное, как y (активность мозга), и что «ментальное может быть возвращено к физическому». Можно заключить поэтому, что на основе имеющихся в настоящее время данных редукционизм вряд ли можно рассматривать как жизнеспособную опцию.

Слабый эмерджентизм базируется на трех постулатах: (1) физический монизм, (2) тезис о системных свойствах и (3) тезис синхронного детерминизма [Bedau, 1997; Stephan, 2002]. Первый постулат можно не рассматривать, потому что он принимается просто как аксиома. Второй постулат (свойство является системным, если им обладает система, но не обладает ни одна из ее частей) вполне применим к рассмотренным нами данным о свойствах ментальной сферы и активности мозга. Третий постулат является ключевым. В приложении к содержанию сознания этот тезис означает, что (1) разные содержания сознания должны иметь разную физическую основу (т. е. соответствовать разным паттернам мозговой активности), (2) одно и то же содержание сознания должно иметь одну и ту же физическую основу как при повторных его появлениях у одного и того же субъекта, так и у разных субъектов.

Для проверки этих требований нужно ставить специальные эксперименты, которые в существующей литературе мне не удалось найти. Однако ряд имеющихся данных позволяет сомневаться в их выполнимости по крайней мере для некоторых видов семантики сознания. Например, как мы видели, данные о локализации нейронных ансамблей, кодирующих семантическую память, показывают, что один и тот же смысл у разных людей и у одного и того же человека в разные моменты времени может быть закодирован с помощью разных конфигураций нейронных сетей. Тем не менее субъективно человек воспринимает его как один и тот же смысл, который не отличается от понимания этого смысла други-

ми людьми. Если бы это было не так, социальное поведение было бы невозможным.

Строго говоря, суть заключается в том, что одно и то же содержание сознания может иметь разную физическую основу при повторных его появлениях у одного и того же субъекта и у разных субъектов. Это подтверждается и рассмотренными выше данными о меж- и внутрииндивидуальной вариативности активности мозга, записанной с помощью ЭЭГ или фМРТ. Можно думать поэтому, что единственная потенциально жизнеспособная разновидность эмерджентизма – его сильный вариант, при котором эмерджентная сущность обладает нисходящей каузальной силой. По мнению Эллиса, нисходящая причинно-следственная связь от разума к миру лежит в основе всего человеческого планирования и действий и любой физической эксперимент – это пример интеллектуальной нисходящей причинно-следственной связи [Ellis, 2008].

Далее мы перейдем к рассмотрению видов психических процессов, которые связаны с ценностными ориентациями, и увидим, что, хотя их истоки часто можно проследить в животном мире, у человека они имеют другой характер и выполняют, вероятно, другие функции. Связь этих процессов с активностью мозга в целом еще менее понятна, чем для когнитивных процессов.

ЦЕННОСТИ

Пожалуй, наиболее характерной чертой суждений человека в обычной жизни является их оценочный характер. Имеются в виду, например, суждения о собственном поведении и поведении других людей. Редко описание какого-либо поступка человека не сопровождается его оценкой – хорошо это или плохо. Также можно выделить суждения о внешних ситуациях – об изменении погоды или финансовой ситуации в мире. Мы не просто фиксируем, что начался проливной дождь, или что упал индекс Джонса, а моментально оцениваем, хорошо это или плохо. О чем говорит тенденция классифицировать все события на хорошие и плохие? На чем основывается данная классификация и выполняет ли она какую-то функцию в ментальной жизни человека?

Прежде всего нужно отметить, что оценка по принципу «хорошо/плохо» не является обязательным компонентом восприятия мира. Немалое количество объектов и событий мы воспринимаем как нейтральные. Вряд ли можно всерьез говорить, что компьютер воспринимает мир, но современные «интеллектуальные» системы совершают очень сложные «ментальные» операции и могут собирать большое количество информации о внешнем мире. Все это не сопровождается классификацией объектов на хорошие и плохие, если такая опция преднамеренно не заложена в программу устройства. С другой стороны, различение «хорошего» и «плохого» является непременным атрибутом любого живого существа. Конечно, мы не можем спросить обезьяну и тем более бактерию – что они считают хорошим или плохим, но в большинстве случаев это можно легко понять по их поведению – к хорошему они стремятся приблизиться, а от плохого отодвинуться как можно дальше.

Соответствующее поведение выделяют в категории приближения (approach) и отступления/избегания (withdrawal/avoidance) [Greenberg, 2017]. Ребенок, фактически сразу после появления на свет, тоже четко различает и демонстрирует родителям – что он считает для себя хорошим или плохим. Можно заключить, таким образом, что восприятие мира через призму «хорошее/плохое» – это обязательный атрибут любого живого существа. В основе такого восприятия лежит стремление к самосохранению. Хорошо

то, что этому способствует, плохо то, что угрожает. Без такого стремления поддержание жизни как неравновесного упорядоченного состояния невозможно, так как спонтанные процессы, подчиняющиеся второму закону термодинамики, самопроизвольно разрушают любую сложность. Объяснение появления такого «стремления» в живых системах, по мнению Кауфмана, требует открытия «четвертого закона термодинамики» [Kauffman, 2004].

В простейшем случае целью является лишь самосохранение. Наличие стремления к самосохранению необходимо и достаточно для того, чтобы могли работать механизмы естественного отбора, сравнимого с финишной чертой в соревновании бегунов. Кто обогнал соперников и первым пересек финиш, тот победил. Но чтобы это произошло, нужно, чтобы все бегуны хотели бежать. У более развитых биологических видов самосохранение может отходить на второй план, а на первый выходит сохранение собственного потомства. Наконец, у самых «продвинутых» существ (в первую очередь у человека) приоритетом может быть сохранение и благополучие социальной группы, члены которой не обязательно являются кровными родственниками. В этом случае оценки «хорошего» и «плохого» могут сильно меняться, так как то, что хорошо для собственного сохранения, не обязательно хорошо для группы.

8.1. Поведение приближения и отступления/избегания

Внешне восприятие мира через призму «хорошее/плохое» может выражаться в поведении приближения и отступления/избегания. Концепцию поведения приближения и отступления впервые ввел в 1939 г. Теодор Шнейрла [Schneirla, 1939]. Хотя эту концепцию считают лишь «прагматической гипотезой» [Bunge, 1980], ее универсальная применимость к поведению животных на всех уровнях организационной сложности указывает на ее законоподобный характер [Greenberg, 2017]. Двухфазные процессы (приближение и отступление) характерны для животных на всех филогических уровнях и прослеживаются до эволюции адаптивных сенсомоторных систем у низших животных, таких как светочувствительные одиночные клетки у дождевых червей. Стимулы низкой интенсивности вызывают реакции приближения, а высокоинтенсивные – отступления. Это верно на протяжении всей истории жизни простых организмов и на ранних стадиях развития более сложных животных.

Концепция поведения приближения и отступления широко использовалась в психологических исследованиях, прежде всего в психологии развития и в психологии индивидуальных различий, однако Гринберг отмечает две особенности использования этой концепции. Во-первых, несмотря на ее широкое использование, особенно в психологии развития, мало кто отдает должное Шнейрле как автору этой концепции, чаще она воспринимается как общее знание, не требующее доказательств. Во-вторых, если термин «поведение приближения», так же как и его наполнение, сохранились практически неизменными, поведение отступления стали преимущественно называть поведением избегания (avoidance).

Шнейрла считал, что поведение приближения неправильно противопоставлять поведению избегания, так как они возникают на разных стадиях эволюции и индивидуального развития. Приближение и отступление – более ранние формы поведения, а их более поздние аналоги – поисковое поведение (seeking) и избегание [Schneirla, 1959]. Элиот, цитируя Шнейрла, пишет, что выбор между поведением приближения и избегания является наиболее важным адаптивным выбором, который организмы должны были делать в ходе эволюции в прошлом, и вполне понятно, почему эта адаптивная функция сохранилась у всех представителей мира живых существ [Elliot, 2006]. Действительно, исходным условием существования живого организма является его самосохранение и способность найти необходимые для поддержания жизни ресурсы. Особи, не обладающие такими способностями, должны были элиминироваться на ранних этапах эволюции. Поэтому тенденция классифицировать все стимулы внешней среды на «хорошие» (потенциальные источники награды – поведение приближения) и «плохие» (потенциальные угрозы – поведение избегания) должна лежать в основе мотиваций любого живого существа.

У относительно простых животных сенсорные органы «заточены» на восприятие именно потенциально полезных или вредных особенностей окружения. У лягушки, например, зрительный анализатор устроен так, что она может воспринимать лишь движущиеся объекты, потенциально являющиеся добычей или хищником [Ewert, 2004]. К категории хорошего должна относиться не только пища, так как у особи, способной поддержать собственное существование, но не способной оставить потомство, нет перспективы сохранить свою линию в будущих поколениях. Соответственно, стимулы, исходящие от противоположного пола, а также,

у млекопитающих и птиц, от потомства, тоже попадают в категорию «хорошее». В категорию «плохое» входит все, что непосредственно угрожает жизни. У более сложно организованных животных поведенческим ответом на стимулы угрозы не всегда является бегство (flight) или избегание (avoidance). Это может быть и ответная атака (fight) и застывание на месте (freezing).

Нейрональный субстрат поведения приближения и избегания хорошо изучен в экспериментах на животных. В классических экспериментах Джеймса Олдса и Питера Милнера крысы с живленными в определенные точки мозга электродами могли до изнеможения нажимать педаль, подающую в эти точки стимуляцию [Olds, Milner, 1954]. Эти точки располагались по ходу путей (медиальный переднемозговой пучок) и в центрах дофаминергической системы, нейроны которой лежат в черной субстанции ствола мозга и посылают аксоны в разные части мозга. Наиболее важное переключающее ядро (nucleus accumbens) расположено в базальных ганглиях переднего мозга. Эту систему сейчас принято называть системой вознаграждения. Электростимуляция других точек (центральное серое вещество в среднем мозге, медиальный и латеральный гипоталамус) вызывает поведение бегства [Schmitt et al., 1981]. В системе избегания важную роль играет также миндалины – центр негативных эмоций. Существует также гипотеза, связывающая поведение приближения с активностью левого полушария, а поведение избегания – правого [Barnard et al., 2015].

Системы приближения и избегания относят к базовым биологическим мотивационным системам, лежащим в основе регуляции поведения в том числе и человека. Эта концепция приобрела большую популярность в сфере исследования темперамента. Разные ее варианты можно найти в теориях Джеффри Грея [Gray, 1970], Роберта Клонингера [Cloninger, 1988] и Джерома Кагана [Kagan et al., 1987]. Согласно этим теориям, чувствительность к стимулам награды и наказания лежит в основе индивидуальных различий поведения людей. В теориях Грея и Клонингера мотивационные системы приближения и избегания в мозге связываются с активностью дофаминергической и серотонинергической систем соответственно. Иногда из них делают вывод, что все поведение человека фактически сводится к стремлению получить награду и избежать наказания [Cott, 2004]. В основе такого взгляда лежит представление о человеке как о рациональной машине, поведение которой обусловлено исключительно стремлением максимизиро-

вать прибыль и минимизировать потери. Такое представление широко распространено в научных областях, занимающихся моделированием поведения человека, таких как экономическая теория игр. Торстейн Веблен писал в свое время, что, по гедонистической концепции, человек – это молниеносный счетчик удовольствий и страданий, который колеблется, как однородная глобула желаний счастья, под воздействием стимулов, перемещающих его по местности, но оставляющих его нетронутым [Veblen, 1899]. Интересно, что моделирование поведения рационального агента, стремящегося максимизировать прибыль и минимизировать потери, лучше предсказывает поведение животных, чем поведение человека [Ross, 2019].

8.2. Эмоции

С темой мотивационных систем награды и наказания тесно связана тема эмоций. Мы не будем подробно останавливаться на этой воистину необъятной и противоречивой теме (достаточно сказать, что до сих пор нет общепринятых взглядов ни на определение понятия «эмоция», ни на классификацию эмоций) и лишь затронем некоторые ее аспекты, связанные с дихотомией «хорошее/плохое». Начнем с эволюционных нейробиологических теорий эмоций, развивавшихся Джеймсом Папецем [Papez, 1995] и Полом Маклином [MacLean, 1952]. По этим теориям эмоции млекопитающих связаны с активностью ряда структур, включая гипоталамус, миндалину, гиппокамп, поясную извилину, перегородку и некоторые другие. Набор этих структур был впервые описан как единая система Папецем, позже его стали называть кругом Папеца. Позже Маклин предложил термин «лимбическая система». Лимбическая система анатомически связана с обонятельным анализатором и гораздо более развита у млекопитающих, чем у рептилий. Основанное на эмоциональной памяти и эмоциональном реагировании, поведение млекопитающих гораздо более гибкое, чем у пресмыкающихся, у них оно основано на врожденном наборе стандартных реакций.

Пол Маклин выдвинул гипотезу, согласно которой в мозге человека можно выделить стволую часть, идентичную мозгу пресмыкающихся, лимбическую систему (Paleomammalian complex), аналогичную таковой у всех млекопитающих, и неокортекс (Neomammalian complex), более всего развитый у человека [MacLean, 1990]. По этой гипотезе поведенческие паттерны, соответствую-

щие активности этих частей мозга, можно увидеть и в поведении современного человека. В частности, эмоциональное поведение связано с активностью лимбической системы. Эта гипотеза впоследствии критиковалась эволюционными нейробиологами на основании того, что, по современным представлениям, новые части мозга, такие как лимбическая система и неокортекс, в процессе эволюции не добавлялись к прежде существовавшим частям, а просто увеличивались в размерах из «заготовок», имевшихся уже и в мозге рептилий [Striedter, 2005]. Несмотря на эту критику, основную идею Маклина о том, что эмоции современного человека можно рассматривать как развитие базовых механизмов реагирования, лежащих в основе поведения приближения и избегания, разделяют многие [Panksepp, 2002].

Эмоции чаще всего описывают как дискретные и устойчивые реакции на внутренние или внешние события, которые имеют особое значение для организма. Они непродолжительны и состоят из скоординированного набора реакций, которые могут включать физиологические, поведенческие и нервные механизмы [Fox, 2008]. Следуя традиции, родоначальником которой был Дарвин [Darwin, 1998], принято считать, что базовые эмоции являются результатом эволюции; они принципиально одинаковы у человека и родственных или близких животных и обеспечивают поведенческие решения древних и повторяющихся проблем, с которыми сталкивались предки [Ekman, 1992; Panksepp, 1998]. Нет единого мнения по поводу того, какие эмоции нужно считать базовыми.

Один из наиболее известных исследователей нейробиологии эмоций у животных Яак Панксепп на основе своих экспериментов с электрической и фармакологической стимуляцией выделял семь видов эмоционального поведения, включая поиск, страх, ярость, похоть, заботу, панику и игру [Panksepp, 2005]. Каролл Изард определил десять основных эмоций (страх, гнев, стыд, презрение, отвращение, вина, дистресс, интерес, удивление и радость), которые нельзя свести к более простым, но которые могут сочетаться, приводя к другим, более сложным эмоциям [Izard, 2011].

Томкинс предположил, что существует девять основных, или врожденных аффектов, контролируемых унаследованными программами: интерес, удовольствие, удивление, страх, гнев, дистресс, стыд, презрение и отвращение. Их мотивационная сила исходит из того, что они ассоциированы с чувствами удовольствия или боли [Tomkins, 2008].

Пол Экман, следуя примеру Дарвина, классифицировал типы эмоциональных выражений человеческих лиц и предложил выделять шесть базовых эмоций, которые универсально распознаются представителями всех человеческих культур и включают счастье, печаль, отвращение, страх, удивление и гнев [Ekman, 1982]. Представленные Экманом доказательства универсальности базовых эмоций подвергались критике по методологическим и концептуальным причинам. Методологически утверждалось, что эксперименты ошибочны, поскольку они используют парадигму принудительного выбора, которая искусственно способствует консенсусу, и полагаются на экологически нереалистичные стимулы, такие как изображения лиц актеров, имитирующих эмоции [Russell, 1994]. Концептуально утверждалось, что гипотеза о том, что естественный отбор будет благоприятствовать появлению универсальных выражений лица, неправдоподобна, поскольку в конфликтных ситуациях не в интересах человека дать наблюдателям знать, какие эмоции он испытывает (например, показывать страх при конфликте с противником). И даже если бы существовала универсальность распознавания и генерации эмоциональных выражений, альтернативные объяснения, такие как обучение, могли бы объяснить эти данные [Fridlund, 1994].

В противовес теории базовых эмоциональных выражений Экмана, Фридлунд выдвигает поведенческую экологическую гипотезу, заменяя понятие выражения эмоций понятием демонстрации, производимой в зависимости от аудитории, когда ожидается, что эта демонстрация принесет выгоду [Ibid]. Например, то, что Экман описал бы как гневное, грустное или счастливое лицо, поведенческие экологи описывают, соответственно, как демонстрацию готовности к атаке, демонстрацию просьбы о помощи и демонстрацию готовности к поддержке.

Проблемой теории базовых эмоций является то, что характерные для них профили ответов не были убедительно продемонстрированы ни на уровне экспрессивных ответов или вегетативных и нейронных реакций, ни на уровне предустановленных или усвоенных действий [Barrett, 2006; Lindquist et al., 2012]. Анализ семантики человеческого языка, безусловно, показывает гораздо большее количество разнообразных эмоций и их оттенков. Так, Коуэн и Келтнер, используя статистические методы анализа эмоциональных состояний, вызванных короткими видеороликами, выделили 27 разновидностей эмоциональных переживаний: вос-

хищение, обожание, эстетическая оценка, веселье, гнев, тревога, трепет, неловкость, скука, спокойствие, замешательство, тяга, отвращение, сопереживание боли, восхищение, возбуждение, страх, ужас, интерес, радость, ностальгия, облегчение, романтика, грусть, удовлетворение, сексуальное желание и удивление. Эти данные свидетельствуют о том, что базовые эмоции – скорее абстракции, и, по крайней мере у человека, дискретность отдельных эмоциональных состояний можно поставить под сомнение. Выдающийся немецкий психолог Вильгельм Вундт предлагал альтернативный взгляд на эмоции. По его убеждению, эмоциональные состояния представляют собой непрерывные ряды, которые можно аппроксимировать размерностями приятности и интенсивности [Schlosberg, 1941]. Действительно, факторный анализ описаний эмоциональных состояний выявляет два главных латентных фактора – валентность (уровень негативности/позитивности) и уровень возбуждения (arousal) [Russell, 2003].

Предположение Дарвина о принципиально одинаковых у человека и животных базовых эмоциях не разделялось бихевиористами, которые относили такие утверждения к проявлениям антропоморфизма и предлагали описывать поведение животных исключительно в терминах «стимул – реакция» [Watson, 1930]. Позднее исследователи эмоций предлагали сравнивать некоторые типичные проявления эмоциональных состояний и их функциональное значение у животных и человека [Oatley, Jenkins, 1996]. Один из подходов – изучение когнитивного предубеждения (cognitive bias) в восприятии стимула в зависимости от текущего эмоционального состояния, аналогично тому, как человек в зависимости от настроения может воспринимать стакан наполовину пустым или полным.

Когнитивные предубеждения были обнаружены у многих видов, включая крыс, собак, макак-резусов, овец, цыплят, скворцов и пчел [Rygula et al., 2012]. Панксепп использовал вокализацию животных как индикатор их эмоционального состояния, аналогично словесным отчетам людей [Panksepp, 2005]. Наконец, сравнение паттернов активности мозга, сопровождающей эмоциональные состояния, также указывает на их принципиальную схожесть у человека и животных [Ibid]. Критики подобных подходов считают, однако, что интерпретация экспериментальных данных, полученных на животных, в значительной степени зависит от результатов экспериментов, полученных на людях, и неизбежно содержит элементы антропоморфизма [Paul et al., 2005]. Социальные конструк-

тивисты отрицают универсальность эмоций. По их мнению, стиль социальной жизни определяет, как на основе базовых эмоций развиваются сложные эмоциональные состояния человека [Elfenbein, Ambady, 2002].

Предложено большое количество теорий эмоций. Рассмотрим кратко лишь некоторые, наиболее популярные. Одной из первых естественно-научных теорий была теория Джеймса – Ланге, которая постулировала, что наше ощущение телесных изменений по мере их возникновения является эмоцией [James, 1884]. По убеждению Джеймса, здравый смысл ошибается в отношении причинно-следственной связи эмоций и телесных изменений. Мы сожалеем, потому что плачем, злимся, потому что нападаем на кого-то, боимся, потому что дрожим, а не плачем, нападаем или дрожим, потому что сожалеем, злимся или боимся [James, 1884]. Предположение Джеймса, согласно которому эмоции не вызывают своих проявлений, а, скорее, возникают из них, казалось многим проблематичным. Если эмоции не имеют каузальной силы в отношении телесных проявлений и поведения, зачем они вообще нужны [Arnold, 1960]? Кстати, этот спор во многом повторяет рассмотренные нами ранее дебаты в отношении каузальной силы сознания и/или активности мозга. Кроме того, теория не могла адекватно объяснить различия между эмоциями. Висцеральные реакции, характерные для различных эмоций, таких как страх и гнев, неразличимы и поэтому не могут быть тем, что позволяет нам различать эмоции [Cannon, 1929]. Уолтер Кэнон и Филип Бард предложили свою теорию, по которой висцеральные реакции опосредуются активностью гипоталамуса, а эмоциональные чувства – дорзального таламуса [Ibid].

Хотя последующие исследования не пришли к консенсусу в отношении того, имеют ли эмоции существенно разные телесные профили на вегетативном, экспрессивном или нервном уровнях (см., например [Clark-Polner et al., 2016]), очевидно, что изменения активности мозга или автономной нервной системы и чувства, сопровождающие эти изменения, могут лишь частично привести нас к адекватной таксономии эмоций, и, кроме того, теория Джеймса – Ланге не может объяснить роль эмоций в целенаправленном поведении [Scarantino, de Sousa, 2021]. Целенаправленность, или преднамеренность (*intentionality*), эмоций (они всегда связаны с каким-то объектом внешнего или внутреннего мира) предполагает участие когнитивных оценочных процессов или суждений в их

возникновении. Мой гнев на кого-то возникает как следствие моего суждения о том, что этот человек меня обидел [Nussbaum, 2001]. Данное направление мысли развивалось в когнитивистских теориях эмоций.

Согласно теории Магды Арнольд, возникновение эмоции связано с когнитивной оценкой (appraisal) значения ситуации для индивида в терминах «хорошо/плохо». Эмоция – это испытываемое влечение к чему-то, что воспринимается как хорошее (благоприятное, полезное), или отторжение от чего-то, что воспринимается как плохое (вредное) [Arnold, 1960]. Совершенно очевидно, что один и тот же стимул может вызывать разные эмоции у разных людей или у одного и того же человека в разное время, следовательно, не стимулы как таковые вызывают эмоции, а оценка этих стимулов.

Лазарус предложил шесть структурных размерностей оценки (appraisal), включая (1) соответствие цели, (2) конгруэнтность цели, (3) вовлеченность эго, (4) поощрение или наказание, (5) потенциал преодоления и (6) будущая перспектива [Lazarus, 1991]. Теорию Лазаруса называют когнитивно-реляционно-мотивационной, поскольку она утверждает, что эмоция – это сложное состояние АВ, с (оценкой) А как причиной и В как комбинацией тенденции к действию, физиологического изменения и субъективного аффекта [Lazarus, 1991]. При этом оценка является не только причиной эмоции, но и ее частью. Шерер с соавторами [Scherer et al., 2001] выделили шестнадцать размерностей оценки, которые можно сгруппировать в четыре класса: оценка релевантности, оценка последствий, оценка потенциала выживания и оценка нормативной значимости. Один из вариантов оценочных теорий (belief and desire theory of emotions) возник в рамках междисциплинарного направления, сочетающего идеи аффективной науки и информатики [Reisenzein, 2009]. По этой теории эмоции вызываются врожденными механизмами, эволюционная функция которых заключается в сравнении вновь приобретенных убеждений с существующими желаниями и убеждениями, тем самым отслеживая и обновляя центральную репрезентативную систему человека (систему убеждений/желаний).

В некоторой оппозиции когнитивистским теориям стоят перцептивные теории эмоций, которые можно разделить на строгие и слабые варианты [Brady, 2013]. Строгие версии описывают эмоции как формы перцепции, слабые – хотя и подчеркивают сход-

ство эмоций с перцепцией, признают и важные различия. Нео-джеймсианская теория Принца является примером строгой перцептивной теории эмоций [Prinz, 2004]. Сенсорное восприятие, очевидно, оснащено системами, предназначенными для восприятия сигналов зрения, обоняния, осязания, слуха и вкуса. Следуя по стопам нейробиологической работы Дамасио [Damasio, 2003], Принц предполагает, что эмоции также могут зависеть от специальной системы внутри соматосенсорной системы. Таким образом, эмоции буквально представляют собой восприятие телесных изменений на висцеральном, гормональном или скелетно-мышечном уровнях или в форме изменений в соматосенсорных областях мозга [Prinz, 2004].

Слабые варианты перцептивных теорий эмоций полагают, что эмоции аналогичны сенсорному восприятию или проприоцепции, но, помимо этого, они являются прямым восприятием формальных объектов, а не только телесных изменений. Так, по мнению Робертса, «эмоции – это своего рода восприятие» в форме конструкций [Roberts, 2003]. Конструкты («впечатления – то, как вещи кажутся субъекту») основаны на беспокойстве, поскольку сформированы из желаний и нежеланий субъекта. Например, боязнь отца, что огонь причинит вред его дочери, является толкованием огня как опасного, обусловленным нежеланием отца, чтобы с его дочерью случилось что-то плохое.

Тапполет [Tappolet, 2016] предполагает, что эмоции – это перцептивный опыт оценочных свойств (ценностей), таких как опасность (страх) или пренебрежение (гнев). Оценочное восприятие, как и чувственное, неконцептуально по своей природе и когнитивно недоступно. Это дает ответ на вопрос: почему существа, не обладающие концепциями, такие как животные и младенцы, могут иметь эмоции, и объясняет эмоциональную «неподатливость», которую можно приравнять к визуальной иллюзии. Как мы визуальное воспринимаем опущенный в воду карандаш как изогнутый, так и эмоционально воспринимаем прозрачную платформу над Гранд-Каньоном как опасную, хотя понимаем ее неопасность. Критики, однако, возражают, что зрительная иллюзия легко преодолевается при нашей уверенности в ее несоответствии действительности, а эмоциональная иллюзия с трудом поддается переоценке и зачастую побуждает нас действовать, исходя из нее (например, отказаться проходить над Гранд-Каньоном по прозрачной платформе) [Salmela, 2011].

Согласно мотивационным теориям, фундаментальным аспектом эмоций является то, что они мотивируют человека совершать те или иные действия. Деонна и Терони предлагают рассматривать эмоцию как отношение к объекту, вызвавшему эмоцию, и готовность совершить применительно к нему определенные действия. По их теории (*attitudinal theory of emotions*) эмоциональные состояния – это, по сути, переживания ощущения готовности тела к действию. Например, боязнь собаки равносильна «переживанию того, что собака опасна», если оно сопровождается «переживанием того, что тело готовится» к преодолению данной опасности. Точно так же гнев является «переживанием оскорбления» в той степени, в какой оно сопровождается «переживанием готовности тела к ответным действиям» [Deonna, Teroni, 2015]. Таким образом, эмоции – это ощущаемые установки готовности к действию, не сводимые к неэмоциональным установкам и специфичные для каждой эмоции.

Отправной точкой мотивационной теории эмоций Скарантино [Scarantino, 2014] является убежденность в несводимости эмоций не только к суждениям и восприятиям, но и к чувствам. Эмоции должны пониматься вместо этого как особые виды «центральных побудительных состояний» или «поведенческих программ», которые определяются тем, что они делают, а не тем, как они переживаются. И что они делают, так это обеспечивают «общее направление поведения, выборочно усиливая согласованные наборы поведенческих вариантов» [Gallistel, 1980].

Избирательное потенцирование может привести к возникновению чувств, но они не являются необходимыми для самого потенцирования, которое состоит из изменений вероятностей вариантов поведения. Идея о том, что эмоции – это поведенческие программы, которые вызывают приоритетные импульсы к действию, может быть объединена с историей о том, как некоторые из таких программ эволюционировали для решения фундаментальных жизненных задач, приведших к тому, что Скарантино [Scarantino, 2015] назвал «новой теорией базовых эмоций». Критики этих теорий указывали на то, что некоторые эмоции вообще не мотивируют к действию. Во-первых, горе и депрессия влекут за собой общее снижение готовности к действию. Во-вторых, неясно, какие тенденции к действиям могут вызвать эмоции, обращенные в прошлое, такие как раскаяние и сожаление, потому что они сосредоточены на том, что уже произошло и что нельзя изменить.

В-третьих, такие эмоции, как радость, включают избирательное усиление довольно широкого диапазона вариантов поведения, поэтому неясно, склонность к каким действиям может быть связана с ними. В-четвертых, похоже, что тенденции к одному и тому же действию могут быть связаны с разными эмоциями, при условии, что эти тенденции описаны на достаточно абстрактном уровне анализа [Prinz, 2004; Tappolet, 2016].

Поскольку теории базовых эмоций, когнитивистские и перцептивные теории, а также мотивационные теории эмоций в значительной степени потерпели неудачу, радикально новое направление возникло в рамках психологического конструкционизма. Конструкционисты предложили полностью отказаться от модели «латентных переменных», характерной для теории базовых эмоций, заменив ее моделью «эмерджентных переменных», согласно которой эмоции не являются причиной появления выражений лица, вегетативных изменений и поведенческих действий, а скорее возникают из них [Barrett, Russell, 2015]. В частности, как указывают психологи-конструкционисты, не существует однозначного соответствия между гневом, страхом, счастьем, грустью и т. д. и любыми нейробиологическими, физиологическими, экспрессивными, поведенческими или феноменологическими реакциями, и различные реакции, якобы свидетельствующие о возникновении базовых эмоций, даже не сильно коррелируют друг с другом. Эта изменчивость ставит под сомнение саму идею о том, что эмоции имеют онтологический статус как причинные сущности и что они существуют в мозге или теле и вызывают изменения в сенсорных, перцептивных, моторных и физиологических процессах [Barrett, 2005].

По убеждению конструкционистов, эмоции складываются на лету гибкими способами с использованием строительных блоков, не относящихся к эмоциям, примерно так, как приготовленные блюда состоят из ингредиентов, которые не считаются специфическими для них, и могут использоваться по альтернативным рецептам. Один из ингредиентов, из которых состоят эмоции, – это основной аффект, представляющий собой нейрофизиологическое состояние, доступное для сознания в виде простого, нерелективного чувства, являющегося смесью гедонистических (удовольствие/неудовольствие) и активационных (сонный/возбужденный) значений [Russell, 2003]. Психологические конструкционисты подчеркивают, что мы всегда находимся в том или ином

состоянии основного аффекта, который выступает в роли барометра, информирующего нас о нашем «отношении» к потоку событий. Показания барометра – это чувства, понимаемые как смесь удовольствия/неудовольствия и активации/деактивации. Показания могут быть представлены в виде точек в пространстве осей активации/деактивации и удовольствия/неудовольствия [Russell, 1980]. Появились попытки интеграции конструкционизма с когнитивистскими теориями [LeDoux, 2017; Moors, 2017]. Социальный конструкционизм подчеркивает элементы различия между культурами как в эмоциональной лексике, так и в диагностических характеристиках эмоций [Mesquita, Frijda, 1992].

Еще одно направление в изучении эмоций (энактивистские теории эмоций) развивается в рамках междисциплинарной программы энактивизма, обусловленной неудовлетворенностью тем, как когнитивные процессы изучаются в когнитивной науке [Di Paolo, Thompson, 2014; Gallagher, 2017]. В частности, для теории эмоций актуальны две энактивистские темы. Первая – это акцент на активной роли человека в его отношениях с внешним миром, который не является независимой от наблюдателя данностью, а скорее активно формируется путем «осмысления». Эта «смыслообразующая» деятельность лежит в основе познания, как его понимают энактивисты, и доступна всем живым существам, какими бы простыми они ни были, в той степени, в которой они являются автономными и адаптивными системами [Thompson, 2007].

Вторая тема – это акцент на воплощенном (embodied), встроенном (embedded) и расширенном (extended) характере когнитивных процессов. В то время как традиционная когнитивная наука и нейробиология сосредоточены на мозге изолированно от остального тела и от окружающей среды, энактивисты утверждают, что мы не сможем понять ментальные процессы, если пренебрегаем взаимодействиями между мозгом, телом и окружающей средой, поскольку они динамически разворачиваются во времени. Энактивизм, таким образом, как бы объединяет разные направления (когнитивистские, перцептивные и мотивационные теории) в единой концептуальной схеме и отчасти пересекается с конструкционизмом, особенно в отношении акцента на эмоциях как на эмерджентных и гибких сущностях.

Однако, в отличие от конструкционистов, описывающих эмоцию в виде сложной композитной конструкции, энактивисты полагают, что «осмысление» – примитивный феномен, доступный

любому живому существу от бактерий до человека, когда они оценивают окружающую среду с точки зрения того, способствует ли она их выживанию, и действуют так, чтобы повысить свою жизнеспособность, как это делает бактерия, уплывающая от вредного вещества [Colombetti, 2014]. Другая особенность энактивистских теорий – отрицание репрезентативного характера эмоций. Например, страх не означает, что есть реальная опасность, а гнев не означает, что было действительно нанесено оскорбление [Hutto, 2012]. Содержанием страха является не опасность, а представление определенной ситуации как чего-то, от чего следует убежать, вместе с побуждением совершить это действие – содержание, которое по сути является воплощенным, поскольку включает в себя телесную подготовку к бегству.

Представление о нерепрезентативном характере эмоций трудно согласовать, однако, с нормативной практикой сверять содержание эмоций с когнитивной оценкой ситуации. Как только мы понимаем, что чей-то страх побудил его избежать определенного положения дел, мы задаемся вопросом, является ли то, что мотивировало избегание, действительно опасным. Другими словами, мы относимся к эмоциям как к уместным или неуместным по отношению к обстоятельствам, их вызвавшим [Hufendiek, 2018].

Подводя итог вышеприведенному обзору, можно отметить, что трудно выделить положения в отношении природы и функций эмоций, с которыми бы согласились все (или большинство) исследователей. Попробуем сделать несколько заключений.

(1) Можно утверждать, что эмоциональные эпизоды включают, по крайней мере в прототипических случаях, набор физиологических, феноменологических, экспрессивных, поведенческих и ментальных компонентов, являющихся диагностическими для эмоций, и в некоторой степени коррелируют друг с другом. Величина этих корреляций продолжает оставаться центральной темой теоретических дебатов – модели с латентными переменными (модели базовых эмоций) предполагают найти сильные корреляции, тогда как модели эмоций в качестве эмерджентных сущностей ожидают найти слабые корреляции. Эмпирические данные больше согласуются с последними.

(2) Физической основой эмоций является мозг, но нет нейронных цепей, которые бы однозначно соответствовали общепринятым типам эмоций, а мозг функционирует не изолированно, а воплощен (*embodied*) и встроен (*embedded*) в окружающую среду.

(3) Эмоции обычно связаны с сознательными переживаниями.

(4) Роль эволюции в формировании эмоций продолжает быть предметом дебатов. Исследователи продолжают спорить о том, достаточно ли эмпирических доказательств существования базовых эмоций. Некоторые считают, что роль эволюции ограничивается формированием универсальных адаптаций, таких как основной аффект и способность категорировать события на хорошие и плохие, что в совокупности приводит к возникновению эмоций.

(5) Эмоции больше не считаются структурно противоположными разуму.

(6) Эмоции обычно связаны с оценкой значимости ситуации и эта оценка варьирует от примитивных до сложных форм обработки информации.

(7) Эмоции обычно коррелируют с изменениями мотивации что-либо делать.

Таким образом, можно думать, что человеческие эмоции формируются на основе базового мотивационного механизма, исторически прослеживаемого до самых ранних живых существ. Его главная функция состоит в быстрой оценке ситуации в терминах «хорошо/плохо» и запуске основанной на этой оценке поведенческой программы. Человеческие эмоции, однако, не сводятся к этим базовым механизмам и включают большой набор сложных когнитивных процессов. Субъективно испытываемое эмоциональное состояние является эмерджентной сущностью, не сводимой к участвующим в его возникновении компонентам. В мозге нет специализированных модулей для разных типов эмоций. Так же как высшие когнитивные функции задействуют области мозга, изначально для них не предназначенные, так и эмоции связаны с активностью областей мозга, первоначальной функцией которых было нечто другое (например, обонятельный мозг млекопитающих).

8.3. Мораль

«Крошка сын к отцу пришел, и спросила кроха: Что такое хорошо и что такое плохо?» [Маяковский, 1955]. Цитата из классика хорошо иллюстрирует важность оценочных суждений в человеческой культуре. Следующий за вопросом сына ответ отца показывает, что предлагаемая ему схема разделения на хорошее и плохое предполагает следование культурным стереотипам, например, быть аккуратным, смелым (поскольку это мальчик, для девочки, вероятно, подчеркивались бы другие аспекты), слушаться стар-

ших и т. д. Это демонстрирует существенное отличие основанного на оценочных суждениях поведения человека от поведения животных.

Внешней средой для человека является социальная среда. Благополучие человека зависит от его адекватного взаимодействия с этой средой и от благополучия других людей, составляющих его непосредственное окружение. Оно зависит также от процессов в более широких социальных слоях, включающих сообщество, к которому он принадлежит, государство и, в конечном счете, весь мир. Поэтому расчет «выгоды» становится невероятно сложным. При таком расчете часто вступают в противоречие эгоистические сиюминутные интересы конкретного человека и интересы других людей и общества в целом; в перспективе они могут сильно повлиять на благополучие человека.

Нахождение оптимального баланса в подобных расчетах – задача нетривиальная. Можно думать, что она была бы непосильной для большинства людей, если бы у них не было достаточно четкого свода правил, с помощью которых они находят решение в большинстве случаев. Этот свод правил (моральный кодекс) позволяет в типичных ситуациях отличить «хороший» поступок от «плохого». Заметим, что «хорошее» и «плохое» в таком случае далеко не совпадают с элементарной дихотомией (пища, секс, дети – хорошо – приближение; угроза жизни – плохо – избегание), а в целом – скорее противоречат ей. Преследовать исключительно эгоистические цели обычно считается морально неприемлемым, а жертвовать собственными интересами (в предельном случае и жизнью) в интересах других людей, наоборот, заслуживает поощрения. Является ли свод моральных правил универсальным для людей, живущих в разных культурах и находящихся в разных ситуациях? Откуда берутся эти правила и как их наличие согласуется с эволюционной концепцией происхождения человека? Как мораль «закодирована» в мозге? Все эти вопросы мы обсудим ниже.

8.3.1. *Философия морали*

В философии природу ценностных ориентаций и суждений изучает аксиология; она включает этику (философия морали), эстетику (философия красоты и искусства) и философию религии [Flew, 1979]. Метаэтика – раздел этики, изучающий смысл моральных суждений (таких как «хорошо/плохо», «правильно/неправильно»), а также их онтологический и эпистемологический статус [Garner, Rosen, 1967].

По одному из определений, мораль – это набор психологических адаптаций, позволяющих эгоистичным людям пользоваться плодами сотрудничества [Greene, 2013]. По другому, по сути похожему, моральные системы – это взаимосвязанные наборы ценностей, добродетелей, норм, практик, идентичностей, институтов, технологий и развитых психологических механизмов, которые работают вместе для подавления или регулирования личных интересов и создания кооперативных обществ [Haidt, 2011].

Мораль нужно отличать от других сводов правил, существующих в любом сообществе. Этикет, например, иногда рассматривается как часть морали, применительно к нормам, считающимся менее серьезными, чем те виды норм поведения, которые имеют большее значение для морали. Гоббс выражает такую точку зрения, когда он использует термин «малая мораль» для описания порядочности поведения, например, как один мужчина должен приветствовать другого, или может ли человек полоскать рот или ковырять в зубах перед компанией, и отличает их от тех качеств людей, которые касаются их совместной жизни в мире и единстве [Hobbes, 1994]. Когда этикет включается в состав морали, практически невозможно выбрать свод правил, который был бы принят всеми разумными существами [Gert, Gert, 2020].

Закон отличается от морали наличием четких письменных правил, наказаний и должностных лиц, толкующих законы и применяющих наказания. Поведение, регулируемое моралью и законом, часто в значительной степени совпадает, и законы часто оцениваются – и изменяются – по моральным соображениям [Dworkin, 1986]. Несмотря на то, что мораль группы или общества может происходить из их религии, даже в этом случае мораль и религия – не одно и то же. Мораль – это только руководство к поведению, тогда как религия – всегда нечто большее. Хотя часто наблюдается значительное совпадение действий, запрещенных или требуемых религией, и действий, запрещенных или требуемых моралью, религии могут запрещать или требовать больше, чем запрещено или требуется моралью, и могут рекомендовать какие-либо виды поведения, запрещенные моралью.

Как и в случае с законом, некоторые религиозные обряды и предписания подвергаются критике по моральным соображениям, например, практика дискриминации по признаку расы, пола или сексуальной ориентации [Gert, Gert, 2020]. Руководства к поведе-

нию, считающиеся моральными, обычно включают избегание и предотвращение причинения вреда другим [Frankena, 1980] и, возможно, определенную норму честности [Strawson, 1961], но все они, как правило, включают и другие темы, которые могут считаться даже более важными. Например, вопросы, связанные с религиозной практикой и заповедями или с обычаями и традициями (такие как чистота и святость) могут быть более важными, чем предотвращение вреда, так же как признание авторитета определенных людей и подчеркивание лояльности по отношению к группе.

Приравнивание внутригрупповой лояльности к морали, наблюдаемое в некоторых обществах, позволяет сравнительным и эволюционным психологам, таким как Франс де Ваал [De Waal, 1996], утверждать, что мораль есть и у животных. Хотя все общества включают в мораль больше, чем просто заботу о минимизации вреда людям, эта черта, в отличие от чистоты и святости или признания авторитета и подчеркивания верности группе, присутствует в моральном кодексе любого общества. Кроме этого, трудно найти общее содержание, разделяемое всеми системами морали, так же как и общее обоснование верности этих систем. Одни могут апеллировать к религии, другие – к традициям, третьи – к рациональной природе человека [Gert, Gert, 2020]. Поэтому, по утверждению моральных релятивистов, было бы ошибкой использовать термин «мораль» для обозначения универсального кодекса поведения, который при определенных условиях будет одобрен всеми разумными людьми [Prinz, 2007]. Другие считают, что существуют культурные различия в относительном весе, придаваемом, например, соображениям справедливости и межличностной ответственности [Wong, 2014], и фундаментальные разногласия в рейтинге различных видов вреда и пользы, а также относительно того, кто защищен моралью [Gert, 2005].

В оценке смысла моральных суждений можно выделить два противоположных направления – когнитивизм и некогнитивизм. Когнитивисты доказывают, что моральные суждения описывают объективные свойства явлений, т. е. они могут быть верны или неверны [van Roojen, 2018]. В рамках этого направления можно выделить моральный реализм, этический субъективизм и теорию ошибок. Моральный реализм утверждает, что моральные суждения отражают объективные свойства мира. Он имеет две разновидности – этический натурализм и этический ненатурализм.

Согласно убеждениям натуралистов, этические свойства явлений выводятся из их неэтических свойств, доступных эмпирическому изучению. Примером могут быть утилитаристские теории, относящиеся к категории правильных те действий, которые увеличивают суммарный уровень счастья и благополучия всех участников. Классическим примером утилитаристского решения является дилемма трамвая. Трамвай приближается к развилке, на одной ветви которой находятся пять рабочих, ремонтирующих пути, а на другой только один рабочий. Вы стоите рядом с рычагом для перевода стрелок. Если их не перевести, трамвай задавит пять человек, если перевести – только одного. Нужно ли переводить стрелки? Утилитаристский ответ – да, так как суммарное благополучие пяти человек больше, чем одного. В более кровавом варианте (дилемма моста) для остановки трамвая нужно собственными руками столкнуть с моста под его колеса толстяка. Этический натурализм утверждает, что существуют объективные и нередуцируемые к неэтическим особенностям свойства (например, «доброта»), которые люди способны понимать интуитивно. Это направление лежит в основе аналитической философии, одним из создателей которой был Джордж Эдвард Мур. Философ считал натуралистическим заблуждением (*naturalistic fallacy*) стремление объяснить моральные свойства, такие как доброта, естественными свойствами, такими как приятность или желательность [Moore, 1903].

По мнению этических субъективистов, «верность» или «неверность» моральных суждений определяется либо отношением каждого отдельного человека, либо договоренностью между людьми в рамках общества. Большинство субъективистов, таким образом, являются релятивистами, т. е. «верность» или «неверность» моральных суждений они считают относительной. Некоторые, однако, оставаясь субъективистами, доказывают универсальность моральных норм, привлекая для этого либо фигуру гипотетического идеального наблюдателя (*ideal observer theory*), который, будучи человеком рациональным и знающим, мог бы оценить верность или неверность моральных суждений [Firth, 1952], либо Бога, определяющего верность или неверность моральных суждений (для человека предметом морального выбора в этом случае является подчинение/неподчинение воле Бога) [Murphy, 2019]. Теория ошибок представляет собой вариант морального нигилизма, утверждающего, что все моральные сентенции ложны [Mackie, 1977].

Согласно теории некогнитивизма, моральные суждения не могут быть истинными или ложными, потому что они не отражают смысл явлений. Существуют следующие разновидности некогнитивизма: квазиреализм и универсальный прескрипционизм. Эмотивизм утверждает, что этические сентенции просто выражают эмоциональное одобрение или неодобрение некоторых поступков [Stevenson, 1937]. Квазиреализм также считает этические сентенции проекцией эмоций, но эти проекции придают им видимость реальных утверждений [Blackburn, 1993]. Наконец, универсальный прескрипционизм признает, что этические сентенции действуют как универсальные предписания делать или не делать что-то [Hare, 1952].

Онтология морали пытается ответить на вопрос о природе моральных суждений. Если не рассматривать моральный нигилизм, отрицающий существование морали, то в плане онтологии можно разделить все существующие позиции на универсализм, признающий универсальность моральных принципов, и релятивизм, доказывающий их относительность. Исторически представление об универсальности моральных принципов можно проследить от семи законов Ноя в Талмуде до Универсальной декларации прав человека. Кант в «*Основах метафизики морали*» выводит универсальные моральные принципы из стандарта рациональности, который он называет категорическим императивом [Johnson, Cureton, 2021]. Моральные релятивисты, напротив, предполагают, что все моральные суждения берут свое начало либо в социальных, либо в индивидуальных стандартах, и не существует единого стандарта, по которому можно оценить истинность морального суждения.

Эпистемология морали пытается ответить на вопрос: возможно ли нравственное познание и можно ли доказать верность моральных суждений. Большинство моральных эпистемологий (кроме морального скептицизма), включая эмпиризм и моральный рационализм, постулируют, что моральное познание возможно. Одни считают, что моральное знание добывается путем логических рассуждений, другие – что оно может постигаться лишь с помощью интуиции [Haidt, 2001; Bucciarelli et al., 2008]. По теории, опирающейся на аналогию между моралью и языком, предложенной Хомским, у людей есть врожденная способность генерировать в процессе индивидуального развития имплицитную систему моральных ценностей, так же как они способны генерировать

грамматическую лингвистическую структуру. Так же как ребенок без специального обучения быстро начинает генерировать грамматические конструкции, он начинает генерировать сложные моральные суждения, которые более или менее универсальны у людей разных культур [Mikhail, 2005]. Эта модель подчеркивает универсальность и интуитивность моральных суждений.

Степень участия эмоций в моральных суждениях может варьировать в зависимости от ситуации. В описанных выше дилеммах трамвая и моста 89 % участников считали возможным перевести стрелки, чтобы пожертвовать одним человеком и спасти пятерых, но лишь 11 % были готовы для этого столкнуться собственными руками с моста жертву [Hauser et al., 2007]. Поскольку прагматический результат обоих действий идентичен, разницу авторы объяснили эмоциональной реакцией, вызванной необходимостью вступить с жертвой в прямой физический контакт, а не просто перевести стрелки. «Мы чувствуем, что не должны сталкивать толстяка, но считаем, что лучше спасти пятерых, чем одного – чувство и мысль в данном случае различны» [Edmonds, 2014].

Хайдт считает, например, что моральное суждение сродни эстетическому, когда мгновенное одобрение или неодобрение события или объекта возникает в момент восприятия и не может быть легко отвергнуто последующим рациональным рассуждением. Более того, это рассуждение часто используется для обоснования уже принятого решения [Haidt, 2001]. Иллюстрируя свою «теорию моральных устоев», Хайдт приводит данные исследований о влиянии чувства отвращения на принятие моральных решений. Эта эмоция, которая уникальна для человеческих существ и появляется относительно поздно в развитии (примерно в возрасте от 5 до 8 лет), включает характерное выражение лица, тенденцию отстраняться от объекта отвращения, небольшое снижение температуры тела и частоты сердечных сокращений, а также чувство тошноты и необходимости очиститься. Кроме того, испытывающий отвращение субъект склонен относиться с отвращением ко всему, с чем соприкасается (физически или символически) вызывающий отвращение объект.

Исторически чувство отвращения связано с расизмом, сексизмом, гомофобией и ксенофобией и ссылку на него часто использовали в кампаниях дискриминации и даже геноцида. По утверждению Келли [Kelly, 2011], универсальные телесные проявления отвращения возникли в эволюции, чтобы помочь людям избежать

употребления токсинов и других вредных веществ, а более когнитивное или символическое чувство отвращения и загрязнения развилось, чтобы помочь избежать болезней и паразитов. Эта система позже стала использоваться для совершенно другой цели – обозначения границы между своими и чужими и мотивирования сотрудничества внутри группы и наказания и исключения нарушителей. Исследования показали, что отвращение связано с более суровыми моральными суждениями. Испытуемые, заполнявшие анкеты в вонючих или грязных комнатах, более жестко оценивали аморальные проступки [Schnall et al., 2008]. Этот эффект, однако, может быть неспецифическим, так как усиление аффективного возбуждения любого типа связано с усилением карательной реакции на проступки [Lerner et al., 1998].

По теории Хайдта при вынесении моральных суждений либералы полагаются на оценку двух размерностей (вред/забота и справедливость/взаимность), а консерваторы используют три дополнительных размерности (лояльность к своим, авторитет/уважение и чистота/святость). Среди консерваторов исследования выявили связь между моральными оценками, основанными на измерении чистоты/святости, и переживанием чувства отвращения. Люди с более высокой чувствительностью к отвращению были более консервативны в отношении политических вопросов, таких как однополые браки и аборты [Inbar et al., 2009]. Другие исследователи, однако, критикуют преувеличение роли интуиции и считают, что основой моральных решений являются рассуждение и рефлексия [Blasi, 2009]. Эти примеры показывают, что человеческая система морали использует базовые биологические мотивационные и эмоциональные механизмы, но использует их в существенно другом контексте, так же как в процессе эволюции появление высших когнитивных функций, таких как язык и мышление, не сопровождалось появлением новых специализированных модулей, а использовало модули, исходно предназначенные для других целей [Anderson, Penner-Wilger, 2013].

Теория двойственного процесса пытается найти место в решении моральных проблем и эмоциям, и рассуждению [Greene, 2017]. По этой теории при принятии морального решения первой возникает автоматическая, бессознательная интуитивная реакция, которая сменяется более осторожным рассуждением, результат которого либо согласуем, либо несовместим с первоначальным интуитивным ответом [Haidt, 2001; Cushman et al., 2010]. Эта теория

является, собственно, вариантом теории Канемана [Kahneman, 2011] в приложении к моральным суждениям. Предполагается, что интуиция руководствуется деонтологическими предписаниями (что можно и чего нельзя делать в данной ситуации), а рассуждение оценивает ситуацию с утилитаристских позиций. Споры идут по поводу того, осуществляются ли эти процессы последовательно (сначала интуиция, потом размышление) или параллельно [De Neys, Bialek, 2017].

Для тестирования своей теории Грин использовал эксперимент с записью фМРТ в процессе решения моральных дилемм [Greene et al., 2001]. Было показано, что при решении «личных» моральных дилемм (таких, как столкнуть собственными руками толстяка с моста, чтобы спасти пятерых) бóльшая активация наблюдалась в медиальной префронтальной коре, задней части поясной извилины и угловой извилине, а при решении «неличных» моральных дилемм (перевести стрелку, чтобы поезд задавил одного, а не пятерых) бóльшая активация наблюдалась в латеральной префронтальной коре, связанной с процессами рабочей памяти. Авторы интерпретируют эти данные так, что в первом случае преобладают эмоциональные, а во втором – когнитивные процессы. В 2001 г. эта интерпретация могла с некоторой натяжкой быть принята, но с тех пор накоплено большое количество данных, показывающих, что перечисленные структуры являются компонентами дефолт-системы мозга, участвующей в широком спектре процессов самосознания и социального сознания [Greicius et al., 2003; Raichle, 2015].

Позже Грин заявил, что миндалевидное тело в первую очередь отвечает за эмоциональную реакцию, в то время как вентромедиальная префронтальная кора отвечает за сопоставление рациональной утилитаристской и эмоциональной деонтологической реакций [Greene, 2013]. Характерно деонтологические суждения преимущественно поддерживаются автоматическими эмоциональными реакциями, в то время как характерно консеквенциалистские суждения преимущественно подкрепляются сознательными рассуждениями и связанными с ними процессами когнитивного контроля [Greene, 2014]. Аргументы в пользу теории двойного процесса, основанные на данных нейровизуализации, подвергались критике за использование обратной логики (рассуждение типа «если область Z, активность которой в другом эксперименте коррелировала с когнитивным процессом А, активна, значит имеет место процесс А») [Klein, 2010].

Питер Сингер пытается найти эволюционное обоснование теории двойственного процесса. Он считает, что на протяжении большей части нашей эволюционной истории люди жили небольшими группами, где насилие было повсеместным. Деонтологические суждения, связанные с эмоциональными и интуитивными реакциями, были выработаны людьми в процессе взаимодействий между собой. В прошлом веке наши социальные организации изменились, поэтому такие взаимодействия стали менее частыми. Таким образом, утверждает Сингер, мы должны полагаться на консеквенциалистские суждения, которые лучше подходят для нашего времени, чем деонтологические, которые были полезны для более рудиментарных взаимодействий [Singer, 2005]. Как и большинство теоретических обоснований в эволюционной психологии, эти рассуждения чисто спекулятивные.

Селим Беркер критикует теорию двойственного процесса и выводы, которые из нее делают. По этой теории наша деонтологическая интуиция управляется эмоциями, тогда как консеквенциалистские суждения подразумевают работу мысли. Следовательно, деонтологическая интуиция не имеет никакой нормативной силы, в отличие от консеквенциалистских суждений. Беркер подвергает сомнению утверждение о том, что интуиция, управляемая эмоциями, менее надежна, чем суждения, основанные на разуме. Кроме того, предположение, что деонтологические интуиции включают только эмоциональные процессы, а консеквенциалистские суждения – только работу мысли, не имеет достаточного эмпирического обоснования [Berker, 2009]. Некоторые указывают, что интуитивные ответы в этических вопросах часто оказываются более верными (например [Kass, 1998]).

В целом, можно суммировать, что какими бы путями не достигалось моральное решение (интуицией или рассуждением), в основе лежит противопоставление интересов других людей собственным эгоистическим интересам. Преследовать исключительно последние в ущерб первым считается аморальным во всех системах морали. То есть мораль, по сути, противостоит эгоистическим базовым биологическим мотивациям в пользу интересов других людей. Является ли мораль исключительно человеческой особенностью, или ее можно найти и у животных? Если да, то как мораль могла появиться в процессе естественного отбора, который, казалось бы, должен отбирать лишь сугубых эгоистов? «Закодирована» ли мораль в структуре мозга?

8.3.2. Биология морали

Если говорить о возможных путях эволюции, которые могли привести к появлению морали, то нужно сразу отметить следующее. Распространенное представление о том, что любое качество, которое мы наблюдаем у современного человека (строение частей тела, физиологические процессы, особенности поведения и ментальных процессов), существует лишь потому, что оно стало таким в результате естественного отбора и, значит, давало непосредственные преимущества для выживания, – является заблуждением. По современным представлениям, все новое появлялось в эволюции не благодаря, а, скорее, вопреки естественному отбору.

По терминологии Е.В. Кунина, это могло совершиться лишь когда вид «проходил через бутылочное горлышко», т. е. по каким-то причинам численность вида и, соответственно, очищающая сила естественного отбора резко снижались. Механизм появления нового не определен, и Кунин приписывает это просто случайному стечению обстоятельств. Появившись, новое может либо быстро элиминироваться естественным отбором, если оно уменьшает шансы на выживание и конкуренция достаточно высока, либо сохраняться в латентном состоянии, пока в изменившихся обстоятельствах ему не найдется применение [Кунин, 2012]. Второй возможный путь появления некоторого свойства или процесса – его опосредованность другим процессом, который отбирался естественным отбором. Возможно, например, что способность выносить моральные суждения является побочным продуктом интеллектуальных способностей, которые возникли в процессе эволюции. Исходя из этой позиции, мы склонны выносить моральные суждения потому, что мы интеллектуальные и размышляющие существа, а не потому, что естественный отбор специально дал нам эту способность в качестве адаптации [Machery, Mallon, 2010]. Однако в подавляющем большинстве теорий (которые скорее нужно считать гипотезами) эволюционного происхождения морали рассматриваются лишь механизмы прямого естественного отбора.

Нужно также подчеркнуть, что мораль в приложении к человеку – это всегда моральное суждение. То есть это не просто некоторые эмоции, такие как симпатия, или альтруизм, или виды поведения, а суждения о том, как должно и не должно себя вести в тех или иных социальных обстоятельствах. Определенные эмоции

и поведение имеют значение лишь в той мере, в какой они относятся к появлению такого суждения, но в отсутствие суждения они принадлежат лишь к протоморали [Ibid]. Тем не менее во многих дискуссиях используется более свободное понятие морали, которое относится просто к определенным типам социально положительных эмоций или поведения, таким как альтруизм.

Многие дискуссии о морали и эволюции в основном сосредоточены на проблеме альтруистических чувств и поведения. Однако даже в этом случае нужно отличать психологический альтруизм от биологического альтруизма, который встречается у многих видов животных [Kitcher, 2011]. Психологический альтруизм предполагает заботу о благополучии других и сознательное обеспечение выгоды для них самих, без каких-либо ограничений по типу получаемой выгоды. Напротив, биологический альтруизм не имеет ничего общего с намерениями или мотивами, а относится только к «выгодам», которые повышают репродуктивную успешность других (увеличивая их генетический вклад в будущие поколения) [FitzPatrick, 2021]. Хотя психологический альтруизм отличается от биологического, существует множество возможных объяснений эволюции психологического альтруизма, апеллирующих к тем же факторам, которые могут объяснять происхождение биологического альтруизма, а именно: родственный отбор или теория «инклюзивной приспособленности» [Hamilton, 1964]; давление отбора, ведущее к командной работе [Axelrod, 1984] и косвенной взаимности [Joyce, 2006]; отбор групп [Sober, Wilson, 1998].

Сама идея о том, что биологический альтруизм может возникнуть в результате естественного отбора, может показаться странной. В конце концов, один из способов охарактеризовать естественный отбор – рассмотреть его с точки зрения «эгоистичных генов»: естественный отбор происходит, когда вариант гена (аллель) в данном локусе имеет тенденцию вызывать изменение телесного или поведенческого признака (фенотипический признак) таким образом, что увеличивается относительная частота этого варианта гена в следующем поколении [Dawkins, 2016]. Но как мы можем перейти от «эгоистичных генов», увеличивающих свое представительство в генофонде за счет улучшения репродуктивного успеха их носителей, к таким вещам, как сотрудничество и биологический альтруизм?

Один гипотетический механизм под названием «родственный отбор» подразумевает, что альтруистическое поведение по отношению к близким родственникам увеличивает суммарное выживание клана. Есть и другие, которые не ограничиваются родственниками и объясняются с помощью теории игр и групповых динамических моделей [Okasha, 2020]. Ключевым моментом является то, что любой генетически обусловленный признак может развиваться, если он окажет правильный обратный эффект на гены, влияющие на него. Соответственно, генетически обусловленная предрасположенность к психологическому альтруизму может развиваться, если она способствует распространению генов, которые ее вызывают. Если, предположим, вы посылаете деньги для улучшения системы медицинской помощи детям в Африке, с чисто биологической точки зрения кажется, что такой «неизбирательный» альтруизм не является биологически адаптивным, поэтому может показаться загадочным, как такая черта могла развиваться в результате естественного отбора. Но черта, не адаптивная в настоящее время, могла когда-то быть таковой. В среде, в которой жили наши предки-гоминины, где было мало контактов с посторонними, даже относительно неизбирательный альтруизм имел тенденцию приносить пользу родственникам или потенциальным реципрокаторам и, таким образом, мог быть адаптивным механизмом в целом. Если это так, то неизбирательное альтруистическое поведение, являющееся неадаптивным в нынешнем глобальном мире, может быть выражением некогда адаптивной тенденции, наподобие тому, как пристрастие к жирной пище, бывшее ранее адаптивным, не приносит пользы в сегодняшней экономике быстрого питания [Dawkins, 2006]. Конечно, это только спекуляции, и вполне возможно, что альтруистическая тенденция сама по себе вовсе не является адаптацией, а коренится в ценностях, которые мы развили независимо от эволюционных влияний в результате морального познания в культурном контексте.

Еще одно, кроме альтруизма, чувство, часто обсуждаемое в контексте эволюции морали, – чувство социальной справедливости. Действительно, у людей есть сильное, эмоционально заряженное чувство элементарной справедливости, негодование в отношении обманщиков и желание их наказать; все это находит выражение как в культурных нормах, так и в индивидуальных моральных суждениях. Некоторые из этих психологических черт могут иметь аналоги у животных. Сара Броснан и Франс де Ваал

[Brosnan, de Waal, 2014] утверждают, что неприятие неравенства, т. е. негативные реакции на неравное вознаграждение за аналогичные задачи, широко распространено среди кооперативных видов животных. В простейшем случае животное протестует, когда видит, что компаньон получает более высокую награду за аналогичную задачу, как в известном исследовании этих авторов с обезьянами-капуцинами, хотя аналогичные эффекты теперь наблюдались и у неприматов, таких как собаки и вороны. В более сложном случае шимпанзе иногда негативно реагируют на несправедливость, даже если они сами получают большее вознаграждение.

Эволюционная гипотеза в отношении этой более сложной формы неприятия неравенства заключается в том, что оно развилось у существ с достаточными предсказательными способностями и эмоциональным контролем потому, что сохраняло полезные отношения сотрудничества, которым могло угрожать такое неравенство. Согласно выдвинутой Броснан и де Ваал гипотезе, наблюдаемая у шимпанзе реакция – это то, что позволило развить у людей чувство справедливости, которое направлено не на равенство само по себе, а на продолжение сотрудничества [Ibid].

Однако необходима осторожность при переносе наблюдений за животными на человека. Наши моральные суждения и результирующее поведение нельзя рассматривать просто как последствия таких биологических сил, наравне с кооперативной деятельностью пчел или негодованием, испытываемым обезьянами-капуцинами (или даже шимпанзе) по поводу неравного вознаграждения за равный труд. Когда рациональный агент выносит суждение, будь то в сфере морали или в таких областях, как наука, математика или философия, правильный вопрос заключается в первую очередь не в том, что послужило эволюционной причиной появления такого суждения, а в том, какие причины были у этого человека для его вынесения, так как мы обычно предполагаем, что человек способен к разумному рассуждению и пришел к своему выводу по причинам, которые он приводит (независимо от того, являются ли эти причины и этот вывод в конечном итоге правильными). То есть мы предполагаем, что люди обладают значительной автономией в своем мышлении [FitzPatrick, 2021].

Люди в большей или меньшей степени обладают способностью к рассуждению, которая следует автономным стандартам, соответствующим рассматриваемым предметам, а не является рабским служением эволюционно запрограммированным инстинктам,

отфильтрованным через культурные нормы. Такие рассуждения и результирующее поведение автономны в том смысле, что они не формируются конкретными эволюционно заданными тенденциями, а следуют независимым нормам, соответствующим рассматриваемым вопросам [Nagel, 1979b]. Это трудно отрицать в отношении таких абстрактных предметов, как алгебраическая топология, квантовая теория поля, популяционная биология, модальная метафизика или двенадцатитональная музыкальная композиция, все из которых очевидно включают именно такие автономные приложения человеческого разума.

Даже если за нашей общей тенденцией к определенным видам умственной деятельности или за некоторыми из наших мотивов в этих занятиях стоят эволюционные влияния (например, если конкретный музыкант или поэт мотивирован сочинять музыку или стихи, чтобы произвести впечатление на потенциального полового партнера), это не означает, что детали самой этой деятельности регулируются такими влияниями [Buchanan, Powell, 2019]. Эволюционный биолог, отвергающий предположение об автономности человеческой мысли, рискует пасть жертвой собственной концепции, так как его теория также может рассматриваться как следствие действия эволюции на его мозг, а не отражение реальных законов мира. Поэтому обычно эволюционные биологи распространяют свои гипотезы не на все виды человеческих рассуждений, а избирательно на моральные и теологические суждения.

Уилсон [Wilson, 1998], например, выдвигает дефляционную гипотезу о том, что моральная и теологическая рефлексия – это занятия, «основанные на генетических алгоритмах» (или, на жаргоне эволюционных психологов, они используют возникшие в процессе эволюции специфичные для этой предметной области модули). В то же время, однако, он предполагает, что его собственные биологические и квазифилософские размышления о биологии, морали и религии не таковы: они автономны и поэтому не могут быть объяснены с точки зрения эволюционных инстинктов.

У людей обычно есть причины для своих моральных суждений, и, независимо от того, согласны мы с ними или нет, мы обычно анализируем эти причины, чтобы объяснить, почему они верят в то, что делают. То есть мы склонны относиться к моральным суждениям человека так же, как относимся к его аргументированным математическим, научным или философским суждениям. На-

пример, даже если верно, что эволюционный отбор, благоприятствующий заботе о потомстве, сильно повлиял на наше отношение к детям, из этого не следует, что это полное объяснение того, почему мы считаем наши обязательства по уходу за своими детьми особыми. Здесь мы видим только часть объяснения, в то время как другая часть может быть связана с признанием целесообразности родительской заботы: у нас есть веские причины проявлять особую заботу о наших детях, что должно мотивировать нас, даже если мы не были уже мотивированы инстинктивными чувствами [FitzPatrick, 2021].

Чтобы расширить потенциальный охват модели эволюционного влияния на моральное суждение, часто выдвигают «гипотезу рационализации» в отношении причин, по которым люди приводят моральные рассуждения для убеждений, сформированных на самом деле эволюционными факторами. С этой точки зрения объяснение причин наших моральных убеждений интерпретируется как простая апостериорная рационализация, т. е. наши моральные убеждения вызваны просто эмоциями или «моральными инстинктами», которые являются наследием нашего эволюционного прошлого, и мы изобретаем рационализацию этих убеждений, чтобы попытаться «оправдать» их для себя, не осознавая их истинное происхождение [Haidt, 2001; Leiter 2007].

Идея о том, что некоторые моральные убеждения допускают такое разоблачительное объяснение, не нова. Философы-моралисты давно признали, что люди часто имеют моральные убеждения, основанные на личных интересах или предубеждениях, и рационализируют свои взгляды, изобретая оправдания данной позиции. Например, всего несколько поколений назад многие считали аморальными межрасовые браки, а сегодня многие воспринимают таковыми гомосексуальные браки. Философы склонны считать «гипотезу рационализации» правдоподобной для таких убеждений, потому что предлагаемые в этих случаях обоснования не выдерживают критического анализа и существуют правдоподобные альтернативные объяснения того, почему люди могут придерживаться таких убеждений, например, что они неверно истолковали личное чувство отвращения как восприятие объективной моральной истины [FitzPatrick, 2021]. Но насколько широко дефляционная модель «гипотезы рационализации» применима к моральным суждениям и в какой степени в этих случаях замешаны эволюционные влияния?

С одной стороны, кто-то может отрицать, что допущение об автономности мышления вообще применимо к моральной сфере: нам либо не хватает способностей в сфере морального мышления, либо мы никогда их не применяем. Однако такое утверждение кажется малообоснованным. Почему человеческий интеллект практически не знает границ в других областях, но когда дело доходит до морального мышления, застревает в колее, пробитой для нас эволюцией, рабски следуя предписанным ей образцам мышления? Сам факт человеческого самосознания делает такую картину маловероятной. Как только нам говорят, что наше мышление ограничено эволюционно заданными путями, само наше осознание этих влияний открывает путь для поиска новых возможностей. Если, например, вам скажут, что вы эволюционно запрограммированы в своих моральных суждениях предпочитать свою группу посторонним, то вы, как рефлексивный агент, можете принять во внимание данный факт в своем последующем моральном размышлении и решить, что это предпочтение неоправданно и, таким образом, прийти к новому, более эгалитарному моральному взгляду. Точно так же может быть верно то, что мы обладаем специальными эволюционно заданными механизмами для «считывания» лиц как заслуживающих или не заслуживающих доверия, и часто на этой основе выносим суждения за доли секунды; но это не мешает нам пересматривать такие суждения, например, когда мы, размышляя о поведении кого-то с «заслуживающим доверия лицом», понимаем, что он на самом деле негодяй [FitzPatrick, 2021].

Примечательно, что Джонатан Хайдт, ведущий сторонник гипотезы рационализации в области морали, не занимает крайнюю позицию в рассматриваемом вопросе. Он допускает, что способности рационального мышления в области морали существуют и иногда используются, выделяя моральные размышления философов в качестве вероятных примеров [Haidt, 2001]. По его мнению, однако, такое использование автономных рассуждений в производстве моральных суждений является редкостью; в подавляющем большинстве случаев наши суждения есть результат непосредственной «интуиции», отражающей эмоциональную реакцию. Однако возникает много вопросов по поводу даже такого «умеренного» утверждения.

Даже если наши моральные суждения основаны на «интуициях», может оказаться, что эти интуиции сами основываются на предшествующих актах рефлексии и рассуждений [Pizarro, Bloom,

2003]; или интуиции частично являются когнитивными реакциями на объективные особенности ситуаций, которые могут опередить нашу способность сформулировать их – так же, как шахматный мастер может «увидеть», что определенный ход будет лучшим, до того, как сможет четко сформулировать, почему [Kamm, 2007]. Точно так же рассуждения авторитетных философов и религиозных лидеров могут повлиять на этические представления целых обществ, тем самым влияя на моральные суждения их членов, даже если в конкретном случае они не могут подкрепить свои суждения аргументами. Таким образом, остается неясным, в какой степени человеческое моральное суждение может объясняться апостериорной рационализацией, а в случаях, когда это так, то какую роль играет эволюционное влияние на наши эмоции.

Многие моральные убеждения – например, относительно морального равенства людей разного пола и всех рас и национальностей, или моральных обязательств даже перед будущими поколениями в далеких странах – гораздо легче объяснить как результат размышлений, чем как следствие эволюционных факторов. Трудно представить, как моральное убеждение некоторых людей в том, что мы должны обеднять себя и ограничивать собственное воспроизводство, или даже вообще отказаться от детей, чтобы помочь далеким незнакомцам, может быть объяснено гипотезой апостериорной рационализации эволюционно заданных интуиций.

Поразительный диапазон моральных убеждений как в разных культурах, так и внутри них, сам по себе является свидетельством значительного вклада мышления, выходящего за рамки упражнений с эволюционно заданными «нейрокомпьютерными системами» и «алгоритмами социальных контрактов», возникшими для решения конкретных адаптивных задач [Cosmides, Tooby, 2008]. Даже если действительно существуют такие специальные модули, генерирующие рассуждения об обязанностях и правах, как это утверждают эволюционные психологи, вопрос в том, насколько они объясняют и действительно ли они исключают значительный вклад автономных моральных рассуждений. Таким образом, умеренная эволюционная гипотеза состоит в том, что эволюционные силы могут адекватно объяснять лишь некоторые способности и тенденции, связанные с моральным мышлением, чувством и поведением, и могут объяснять полностью или частично содержание

наших моральных мыслей, чувств и поведения лишь в той мере, в какой они на них влияют (индивидуально или через культурное развитие) [FitzPatrick, 2021].

Если не рассматривать «социальный дарвинизм» Герберта Спенсера [Spencer, 1879], который, в отличие от Дарвина, понимал эволюцию как целенаправленный процесс [Rachels, 1990], то нормативное содержание моральных истин трудно вывести из дарвиновской эволюции путем естественного отбора. Краткий обзор определенных черт, считающихся, вероятно, биологической адаптацией у ряда видов, например, склонность убивать «пасынков» или развратничать, должен убедить нас в том, что даже если подобные склонности также являются адаптациями у людей и присутствуют в нас по аналогичным эволюционным причинам, это вовсе не будет аргументом в пользу того, что мы должны так себя вести.

Независимо от того, почему у человека есть какая-либо черта, перед ним всегда стоит вопрос: правильно ли использовать ее, или лучше от нее отказаться и сопротивляться ей, насколько это возможно [Korsgaard, 2006]? Это предполагает, что этика, как и математика, является автономным предметом в том смысле, что у нее есть свои собственные «внутренние стандарты оправдания и критики», так что ее выводы не могут быть оправданы другими формами исследования, такими как эволюционная биология [Nagel, 1979b]. Как пишет Рейчелс, нормативные этические выводы опираются на этические размышления и аргументы, точно так же, как математические постулаты доказываются посредством математических рассуждений, а не посредством изучения информации о нашем эволюционном прошлом или о том, что происходит в нашем мозге, когда мы участвуем в этих формах рассуждений [Rachels, 1990]. Было бы такой же ошибкой пытаться ответить на этические вопросы, исследуя фМРТ или изучая нашу эволюционную историю, как и пытаться решать математические задачи такими средствами.

Согласно традиционно распространенному убеждению, люди обладают особым моральным статусом, который намного выше статуса животного. Данное убеждение было основано на идее о том, что мы созданы по образу Бога и наделены душой. Но, если теория эволюции верна, то не существует четкого разделения между *Homo sapiens* и остальным живым миром; вместо этого су-

существует континуум эволюционной истории, без каких-либо видов, созданных отдельно, по божественному образу. Это не означает, однако, что мнение об особом человеческом статусе ложно. У него могут быть другие формы поддержки, не зависящие от идей креационистов. Хотя мы и родственники других животных, мы сильно отличаемся даже от наших ближайших родственников-приматов. Чтобы это увидеть, достаточно после посещения зоопарка посетить библиотеку или картинную галерею. Различия между людьми и другими формами жизни на Земле не стираются указанием на то, что люди стали такими, какие они есть, благодаря случайным эволюционным процессам, воздействующим на мозг, подобный мозгу современных обезьян, вместе с тысячелетиями культурной эволюции. Именно эти реальные различия, независимо от того, как они возникли, обосновывают различия в моральном статусе людей и других животных [FitzPatrick, 2021].

Идея морального прогресса занимает видное место в размышлениях ученых об истории и будущем моральной практики. Хотя может быть трудно оценить, добилось ли общество морального прогресса в целом за какой-то период, есть очевидные свидетельства прогресса в определенных сферах. Так, ликвидация работорговли и развитие современного движения за права человека являются убедительными примерами повсеместного морального прогресса. Каковы движущие силы такого морального прогресса (а также морального регресса, когда он происходит), и как эволюционная биология может повлиять на перспективы и стратегии достижения морального прогресса (и предотвращения морального регресса)?

Аллен Бьюкенен и Рассел Пауэлл рассматривают эти вопросы в своих теоретических работах [Buchanan, Powell, 2019; Buchanan, 2020]. Предположим, что тенденция к большей инклюзивности в отношении моральных соображений (т. е. устранение расовых и гендерных барьеров для присвоения полного морального статуса человека и распространение существенного морального уважения на широкий круг животных) считается нравственным прогрессом. Пессимистическая мысль, которая может возникнуть на основе эволюционных соображений, состоит в том, что человеческая моральная психология могла развиваться таким образом, чтобы наложить неизбежные ограничения на реальность достижения устойчивого прогресса в этих отношениях. Хотя наша моральная

психология и содержит определенные альтруистические элементы, они должны быть ограниченными, учитывая природу давления отбора, с которым сталкивались наши плейстоценовые предки. Эти элементы могут быть сосредоточены на родственниках и членах группы, а другие, противостоящие им настроения, создают серьезные препятствия для распространения альтруизма за пределы группы. Даже если у нас есть способность к автономному моральному мышлению, позволяющая по крайней мере некоторым усвоить более всеобъемлющие моральные истины, в человеческую моральную психологию могут быть встроены такие мощные эволюционные ограничения, что подобные идеи вряд ли закрепятся и вряд ли будут мотивировать устойчивые прогрессивные изменения. Могут существовать эволюционные пределы базового морального прогресса, которого мы, эволюционирующие приматы, можем реально надеяться достичь, будучи наделены преимущественно эксклюзивистской, племенной моралью.

Бьюкенен и Пауэлл называют эту пессимистическую идею эвоконсерватизмом и отмечают, что она использовалась для аргументации против социальных и политических движений, организованных вокруг такой «космополитической моральной реформы». По другую сторону той же эволюционной медали эволибералы разделяют ту же предпосылку об ограничениях, налагаемых эволюцией на моральную психологию человека, но, по их убеждению, поскольку прогрессивные цели желательны, мы должны искать биотехнологические способы для изменения человеческой природы, чтобы сделать их реально достижимыми. Бьюкенен и Пауэлл отвергают обе точки зрения, поскольку они отвергают саму картину эволюции моральной природы человека. Они утверждают, что поразительные сдвиги, уже произошедшие в направлении более инклюзивной морали, а также рассмотрение условий, при которых происходил моральный регресс, поддерживают совершенно иную, более тонкую картину развитой человеческой моральной психологии, не имеющую таких пессимистических подтекстов для глубокого и прочного нравственного прогресса [Buchanan, Powell, 2019].

По «биокультурной» теории авторов, дело не только в том, что человеческая рациональность предоставляет значительные возможности для автономных моральных рассуждений, обуславливающих большую инклюзивность, но и в том, что наша возникшая в

результате эволюции моральная психология сама содержит семена для такого сдвига (хотя она в равной степени содержит семена и для потенциального морального регресса). Согласно их ключевой идее, наша моральная природа не является по своей сути ни племенной, ни инклюзивной, но содержит обе предрасположенности, выражение которых зависит от конкретных социально-институциональных условий окружающей среды.

В частности, наши племенные тенденции активируются, когда воспринимаются «сигналы внешней угрозы» (реальные или нет), включая конкуренцию, болезнь, подрыв внутригрупповой сплоченности и так далее; напротив, наши инклюзивные тенденции активируются, когда такие сигналы становятся менее многочисленными или заметными [Ibid]. Это делает людей опасно восприимчивыми к моральному регрессу из-за манипуляции их восприятием посредством демагогического популизма (например, опасность, которая усугубляется, когда дезинформация быстро распространяется через социальные сети). Модель, предложенная Бьюкененом и Пауэллом, оставляет открытыми реальные надежды на продолжающийся и углубляющийся моральный прогресс: человечество может добиться его и поддерживать, не будучи скованным какими-либо встроенными эволюционными ограничениями, но это потребует постоянных усилий по развитию и поддержанию социально-институциональных условий, которые позволяют лучшим аспектам человеческой моральной природы преобладать [Buchanan, 2020].

Сторонники биологической концепции морали утверждают, что моральные решения принимаются почти исключительно на основе интуиций, запрограммированных в человеке эволюцией, а разум используется лишь для апостериорной рационализации принятых решений. В качестве доказательства приводятся наблюдения за социальными животными (преимущественно приматами), у которых обнаружены поведенческие проявления альтруизма и чувства справедливости, а также результаты моделирования эволюционных процессов в теории игр. Даже если принять эти косвенные доказательства и согласиться с тем, что некоторые, имеющие отношение к принятию моральных решений эмоции и мотивации имеют эволюционные корни, их наличие в психике человека само по себе не объясняет всего разнообразия поведения разных людей и конкретного человека в разных социальных ситуациях.

В психике человека имеются, например, исключительно эгоистические мотивации с хорошо понятными эволюционными корнями. Вполне вероятно, что определенные просоциальные мотивации, по смыслу конфликтующие с эгоистическими, также имеют эволюционные корни. Как этот конфликт решается в каждом конкретном случае – в конечном счете зависит от свободного выбора человека. Он может, например, поддаться импульсу убить своего соперника, или поделиться куском хлеба с голодной собакой, а может подавить эти импульсы и выбрать какой-то другой вариант поведения. Поскольку, как это убедительно показано Кантом, даже чистая рациональность может в принципе служить обоснованием морального поведения, принятие моральных решений человеком вряд ли правомерно объяснять лишь эволюционно запрограммированными импульсами. Но даже если не брать во внимание моральные суждения и ограничиться рассмотрением моральных интуиций, трудно объяснить их наличие исключительно действием естественного отбора.

Дарвиновская (и неодарвинистская) теория эволюции основана исключительно на конкуренции, в которой всегда побеждает сильнейший. Движущей силой этой конкуренции является сугубый эгоизм, причем, как утверждает Докинз, эгоизм даже не на уровне отдельной особи, а на уровне отдельного гена [Dawkins, 2016]. Кооперация, взаимопомощь, забота о других поддерживаются и сохраняются лишь в той степени, в какой они способствуют выживанию данной особи (или данной конгломерации генов). В основе такой кооперации всегда лежит собственный интерес. Предполагаю, что практически любой, читающий эти строки, сразу почувствует «аморальность» подобного поведения, основанного исключительно на собственном интересе, поскольку такое чувство (или интуиция) лежит в основе человеческой морали.

Как бы в реальности ни вели себя разные люди в разных культурах и разных обстоятельствах, и какими бы соображениями ни объясняли свое поведение, практически у каждого человека, за исключением патологических случаев, есть ощущение, что преследование сугубо эгоистических интересов в ущерб другим людям (независимо от того, являются ли эти люди членами твоей семьи или группы), или использование других людей для достижения этих целей – плохо. Это чувство распространяется не только на людей, но и на животных, и, как мы отмечали раньше, проявления альтруизма и заботы по отношению к другим особям, не являющимся не-

посредственными родичами, а иногда и принадлежащим к другим биологическим видам, наблюдаются и у некоторых животных. По мнению эволюционных биологов, это побочные следствия инстинктов, направленных на выживание конкретной конгломерации генов. Такое объяснение кажется маловероятным.

Казалось бы, эволюционный механизм должен был привести к появлению и закреплению моральной интуиции, согласно которой хорошо не только помогать родственникам и членам своей группы, но и убивать посторонних. Такая мораль явно способствовала бы победе в конкуренции генов, ее программирующих. Однако, несмотря на то, что недружелюбие к «чужакам» – действительно частое явление во многих культурах и политических системах, на уровне интуиции каждый чувствует, что убивать или притеснять другого человека, независимо от его принадлежности к той или иной группе, – плохо. Ярким примером может быть сюжет Бхагавадгиты, созданной более трех тысяч лет назад, где герой, профессиональный воин, испытывает отвращение к необходимости убивать врагов (членов конкурирующего клана).

Как мы видели в гл. 1, посвященной эволюции, теория, по которой конкуренция и естественный отбор победителей являются единственными двигателями эволюционного процесса, сталкивается с трудностями в объяснении некоторых явлений. В последние годы появилось немало попыток модифицировать эту теорию, добавив что-то, что в противовес конкуренции объединяло бы и направляло процесс эволюции. В сфере социального поведения основанная на эгоизме, разобщающая людей конкуренция дополняется противостоящей ей моральной интуицией, объединяющей людей независимо от их принадлежности к той или иной группе.

8.3.3. Эмпатия

Из обзора теоретических представлений о происхождении и сущности морали очевидно отсутствие консенсуса в отношении ключевых процессов, лежащих в ее основе. В традиции, идущей от философии Канта, краеугольным камнем морали признается рациональное суждение. Противоположная традиция, идущая от Дэвида Юма, ведущую роль в морали отводит чувствам. В качестве такого, лежащего в основе морали чувства чаще всего рассматривают сочувствие (*empathy*). Многие религиозные и философские традиции считали сочувствие и сострадание ключом к нравственному мышлению, поведению или мотивации. Дэвид Юм

сравнивал ум человека с зеркалом, способным отразить ум других людей; по его убеждению, боли и удовольствия других волнуют нас только потому, что мы способны их чувствовать [Hume, 1978].

В последние десятилетия точка зрения о центральности сочувствия в морали многими критиковалась на основе того, что это слишком узкое явление, поскольку оно не может объяснить все аспекты морального поведения. Тем ни менее важность сочувствия как побудительной силы моральных суждений трудно отрицать [Maibom, 2014]. Понятие сочувствия (эмпатии) используется для обозначения широкого спектра психологических способностей, которые считаются центральными для людей как социальных существ и позволяют им судить о мыслях и чувствах других людей, разделять их, эмоционально взаимодействовать с ними и заботиться об их благополучии [Stueber, 2019].

Немецкие философы XIX в. рассматривали эмпатию также в качестве основы эстетического восприятия красоты объектов внешнего мира. Мы ценим объект как красивый, потому что сочувствие дает нам возможность видеть его как аналогию с другим человеческим телом [Lipps, 1905]. Сочувствие в этом контексте понимается как феномен «внутреннего подражания», когда разум отражает умственную деятельность и переживания другого человека, основываясь на наблюдении за его телесными действиями и на врожденной предрасположенности к моторной мимикрии, — факт, хорошо известный в психологической литературе и отмеченный уже Адамом Смитом [Smith, 1853]. По мнению Липпса, хотя такая предрасположенность не всегда проявляется внешне, она всегда присутствует как внутренняя тенденция, вызывающая у наблюдателя кинестетические ощущения, аналогичные тем, которые испытывает наблюдаемый объект. Видя гневное лицо другого человека, например, мы инстинктивно склонны копировать его и таким образом «имитировать» его гнев. Поскольку мы не осознаем таких тенденций, мы лишь понимаем, что видим гнев [Lipps, 1907].

Идея о том, что эмпатия, понимаемая как внутреннее подражание, является основным эпистемологическим средством для понимания других, была возрождена в 1980-х гг. в контексте междисциплинарных дебатов, посвященных вопросу: как лучше всего объяснить наши способности интерпретировать и предсказывать поведение других людей [Davies, Stone, 1995]. Открытие так называемых зеркальных нейронов, которые, как предполагается,

играют важную роль в распознавании эмоциональных состояний другого человека и в понимании целенаправленности его поведения, было истолковано как эмпирическое доказательство идеи Липпса об эмпатии как внутренней имитации. С помощью термина «зеркальный нейрон» ученые ссылаются на тот факт, что существует значительное перекрытие между областями активации в мозге при наблюдении за действиями другого человека и областями, которые активируются, когда мы выполняем то же действие. Подобное же перекрытие было установлено для распознавания эмоций другого человека на основе выражения его лица [Goldman, 2009; Keysers, 2011].

Нейробиолог Галлезе считает зеркальные нейроны основой интересубъективности [Gallese, 2016]. Стюбер [Stueber, 2006], вдохновленный концепцией Липпса об эмпатии как внутренней имитации, называет зеркальные нейроны механизмом базовой эмпатии, который позволяет нам непосредственно воспринимать эмоции другого человека и понимать движения его тела как целенаправленные действия. Данные о зеркальных нейронах и использование при восприятии других людей совсем иных нейробиологических механизмов, чем при восприятии физических объектов, означают, что уже в первичном перцептивном контакте с миром мы не просто воспринимаем физические объекты, а уже отличаем простые физические объекты от объектов, похожих на нас [Meltzoff, Brooks, 2001].

Но активность зеркальных нейронов не является в прямом смысле слова «отражением» активности мозга человека, за которым ведется наблюдение. Джин Десети с соавторами показали, например, что восприятие боли, испытываемой другим человеком, приводит к активации части нейронной сети, участвующей в обработке личного опыта боли. Однако, по мнению авторов, полное совпадение паттернов активации могло бы вызвать страдание и стремление избежать его, а значит, вело бы к снижению эмпатии и вероятности просоциального поведения [Decety, Lamm, 2006]. Соответственно, предполагается, что активированная матрица боли не чувствительна в этом случае к феноменальному качеству боли, а используется лишь как индикатор состояния, испытываемого другим человеком [Decety, Cowell, 2015]. «Резонанс» зеркальных нейронов также модулируется различными контекстными факторами – например, насколько наблюдаемый субъект нам близок, считаем ли мы боль морально оправданной (как, например, в слу-

чае наказания) или рассматриваем ли мы ее как неизбежную и необходимую, скажем, при медицинской процедуре [Gallese, 2016; Rizzolatti, Sinigaglia, 2016].

«Чтение мыслей» в сфере базовой эмпатии следует отличать от «чтения мыслей» более высокого уровня [Goldman, 2006]. Повседневное «чтение мыслей» не ограничивается сферой элементарного сочувствия. Обычно мы не только осознаем, что другие люди чего-то боятся или выполняют определенные движения, но и понимаем их поведение с точки зрения причин их действий в сложных социальных контекстах, используя весь спектр психологических концепций, включая концепции убеждений и желаний. Как показывают данные нейробиологии, выполнение задач такого рода, связанных с созданием модели психического состояния другого человека (theory of mind), сопровождается активацией очень разных областей мозга, таких как медиальная префронтальная кора, височно-теменная кора и поясная извилина, в значительной степени перекрывающихся с топографией дефолтной сети [Frith U., Frith C.D., 2003; Zaki, Ochsner, 2012]. Как именно следует понимать высокоуровневые способности создания модели психического (связаны ли они в первую очередь со стратегиями моделирования на низком уровне или с суждениями и концепциями высокого уровня) – является предметом дебатов.

Следуя традиции, берущей начало от Дэвида Юма и Адама Смита, часто утверждается, что эмпатия – это единственная основа моральных суждений [Slote, 2010], «уникальный эволюционный блок» морали [De Waal, 2006], сила, культивирование которой обладает уникальными революционными способностями трансформировать мир в периоды кризиса [Rifkin, 2009], и «универсальный растворитель» многих социальных проблем [Baron-Cohen, 2011]. Скептики, однако, указывают на темные стороны этого явления [Prinz, 2011; Bloom, 2016]. Согласно эгоистической интерпретации связанных с эмпатией явлений, сочувствие нуждающемуся человеку, с одной стороны, сопровождается негативными эмоциями, такими как чувство вины и стыда, или осознанием негативных последствий отказа от помощи, например, социальных санкций; с другой стороны, может привести к осознанию положительных последствий помощи, таких как социальное вознаграждение или добрые чувства. Сочувствие, согласно этой интерпретации, побуждает нас помогать посредством чисто эгоистических побуждений.

В рамках рассматриваемой концептуальной схемы предполагается, что эгоизм может обеспечить единое и относительно простое описание мотивационной структуры человеческой деятельности. С.Д. Бэтсон доказал несостоятельность этой простой модели и в серии экспериментов убедительно показал, что помимо эгоистических мотиваций нами также движут подлинно альтруистические мотивы [Batson, 1997]. Однако эти исследования не доказывают, что эмпатия является основой морали или что она составляет единственный источник моральной мотивации.

Некоторые из собственных исследований Бэтсона ставят под сомнение утверждение о том, что эмпатия является основой морали, поскольку альтруизм, вызванный эмпатией, может привести к поведению, которое противоречит нашим принципам справедливости. Кто-то может, например, дать лучшую работу или приоритет в получении медицинской помощи лицам, которым он сочувствует, в нарушение моральных принципов. Именно по этой причине Бэтсон проводит различие между альтруистической мотивацией, связанной с благополучием другого человека, и моральной мотивацией, основанной на принципах справедливости. К сожалению, мы не всегда осознаем данный факт, когда абстрактно противопоставляем моральную мотивацию в широком смысле эгоистической мотивации. А значит, нужно быть более осторожными при обсуждении отношений между альтруистическими и моральными мотивами. Наконец, вышеприведенные исследования не дают решения вопроса о том, требуются ли высокоуровневые способности создания модели психического для полноценной моральной свободы воли, поскольку Бэтсон понимает эмпатию прежде всего как эмоциональное состояние [Batson, 2014].

В психологической литературе одно из наиболее полных описаний эмпатии и ее связи с нравственным развитием человека содержится в работе Мартина Хоффмана [Hoffman, 2000]. Автор рассматривает сочувствие как имеющую биологическое основание склонность к альтруистическому поведению. Он считает, что в основе эмпатии лежат различные виды возбуждения, позволяющие нам эмпатически реагировать на сигналы бедствия от другого человека. Хоффман упоминает мимирию, классическое обусловливание и прямую ассоциацию – когда один сопереживает, потому что ситуация другого напоминает одно из его собственных болезненных переживаний. В качестве разновидностей эмпатии, требу-

ющих мышления, он называет опосредованную ассоциацию, в которой сигналы для эмпатической реакции предоставляются в лингвистической среде, или принятие роли.

Хоффман различает шесть стадий развития эмпатических реакций, начиная от реактивного плача новорожденного, эгоцентрического эмпатического дистресса, квазиэгоцентрического эмпатического дистресса до альтруистического сочувствия, сочувствия к другому человеку за пределами непосредственной ситуации и сочувствия к целым группам людей. Соответственно, в процессе развития эмпатические реакции составляют континуум, который варьируется от эмоционального заражения (*emotional contagion*), как в случае реактивного плача новорожденного, до различных форм истинного сочувствия, достигаемых на четвертой стадии, когда ребенок способен эмоционально реагировать на страдания другого более изощренным образом из-за появления способности различать себя и других и осознания того, что у других есть психические состояния, независимые от его собственного. Только на четвертой стадии эмпатия также трансформируется или ассоциируется с симпатией, что приводит к появлению просоциального поведения.

Взгляды Хоффмана на индивидуальное развитие согласуются с описанием эмпатии как феномена, наблюдаемого у разных видов животных на разных уровнях сложности, связанных с разными степенями когнитивного развития [Preston, de Waal, 2002]. Хоффман сочетает свою интерпретацию развития эмпатии с анализом ее роли в развитии моральной свободы воли. Он осознает ограничения нашей естественной способности сопереживать или сочувствовать другим, особенно то, что он называет эффектом «здесь и сейчас», т. е. тот факт, что мы склонны больше сочувствовать людям, которых в каком-то смысле воспринимаем как более близких – например, представителям собственной расы. Как и Бэтсон, Хоффман не считает, что сфера морали ограничена исключительно нашей способностью сопереживать другим людям. Помимо эмпатических способностей, моральная свобода действий требует также знания абстрактных моральных принципов, таких как принципы заботы и справедливости. Тем не менее Хоффман считает эмпатию основой морали [Hoffman, 2014].

Точку зрения Хоффмана критикуют Принц [Prinz, 2011] и Блум [Bloom, 2016]. Принц указывает на эффекты привлекательности, заметности и близости – тот факт, что мы склонны больше

сопереживать людям привлекательным или людям, которые находятся в непосредственной близости, и только если их страдания особенно заметны. Исследования показали, что испытуемые, как правило, ощущают больше эмпатии к членам своей группы, чем к чужакам, и больше контрэмпатии (чувство удовольствия от несчастья другого или чувство неудовольствия от удачи другого) к членам чужой группы [Cikara et al., 2014]. Это особенно верно, если другая группа рассматривается как конкурирующая со своей группой. Сочувствие к чужим можно еще больше уменьшить с помощью различных стратегий дегуманизации, которые, безусловно, использовались в контексте геноцида XX в. и системы расового рабства в Соединенных Штатах [Fuchs, 2019].

Повышенное сочувствие к обидам, нанесенным членам своей группы, также может привести к насильственному и аморальному поведению [Bloom, 2016]. Кроме того, эмпатия имеет тенденцию сосредотачиваться на одном (особенно если он или она идентифицируема), а не на многих. Сочувствие может ввести нас в заблуждение, особенно когда нам необходимо принимать во внимание статистически значимую информацию при решении моральной или социальной проблемы, например, при обсуждении преимуществ вакцинации, когда более уместно думать о большом количестве спасенных детей, а не о пагубных последствиях для одного конкретного ребенка.

Представляется, таким образом, очевидным, что одной эмпатии недостаточно, чтобы надежно удерживать нас в рамках морали. По этой причине Принц в качестве основы морали предпочитает рассматривать не эмпатию, а «моральные эмоции», такие как гнев, вина и стыд. Можно, однако, задаться вопросом, почему мы должны ожидать, что эти эмоции менее подвержены предвзятости и влиянию универсальной человеческой тенденции отдавать предпочтение членам своей группы. Блум считает сочувствие, основанное на разуме, более правдоподобной основой морали, чем просто сочувствие. Это во многом соответствует взглядам Дэвида Юма и Адама Смита, которые еще в XVIII в. предположили, что нам необходимо регулировать эмпатию с помощью определенных корректирующих механизмов, таких как «некоторые устойчивые и общие точки зрения», или перспектива «беспристрастного зрителя», чтобы компенсировать ограниченные возможности эмпатии.

Таким образом, роль эмпатии в морали до сих пор является предметом дебатов. Эмпатия позволяет «почувствовать» состояние другого человека и побудить нас действовать для его блага, зачастую пренебрегая собственными интересами. То есть эмпатия противостоит сугубо эгоистическим мотивам и в этом смысле похожа на мораль. Однако эмпатические мотивы могут не совпадать с моральными, так как подвержены перекосу в сторону членов собственной группы или просто приятных людей. Этот факт и то, что базовая эмпатия наблюдается и у некоторых животных, делает правдоподобной гипотезу, по которой она могла появиться в эволюции в процессе группового отбора [Preston, de Waal, 2002]. Ее связь с моралью, однако, не кажется очевидной.

Представление Юма и современных его последователей о том, что мораль – это та же эмпатия, немного подправленная разумом, кажется маловероятным, если говорить о морали не в широком смысле, как о системе норм поведения, зависящих к тому же от культуры, а как об элементарной интуиции. Заметим, интуиция, которая (в разной степени выраженности) есть у каждого человека, независимо от пола, расы и культурной принадлежности, позволяет практически мгновенно отделить «правильные» поступки от «неправильных», идет ли речь о собственном поведении или поведении другого человека. Например, практически каждый «чувствует», что говорить заведомую ложь другому человеку плохо. Каждый чувствует также, что если он проголосовал на каком-то собрании в пользу властной фигуры или постановления, которое он не одобряет, лишь потому, что хочет избежать нежелательных последствий для своей карьеры, то это плохо.

Подобные интуиции никак не связаны с эмпатией, которая подразумевает сочувствие конкретному человеку в ситуации дистресса. Они также не являются следствием рассуждений, так как появляются практически одновременно с мыслью о том, как можно поступить в данной ситуации. Скорее, рассуждение в этом случае может быть направлено на оправдание неблагоприятного поступка. На мой взгляд, несмотря на наличие большого количества возможных объяснений, начиная с отрицания универсальности моральных интуиций до привлечения эмоций и «быстрых когний» в духе Канемана, правдоподобного объяснения существования этого феномена пока не найдено.

8.3.4. Мораль и мозг

Для чего нужна мораль, в чем ее «функция»? Вопросы подобного рода произвольно возникают у биологов при обсуждении процессов в организме или поведения животных. Такой подход (в философии его называют телеологическим) предполагает наличие какой-то цели, на достижение которой рассчитана система. Тогда функцию любого компонента системы можно понять исходя из того, какой вклад этот компонент вносит в достижение цели. Легко понять, в чем функция карбюратора в автомобиле, поскольку мы знаем, что автомобиль создан для обеспечения передвижения. В физиологии при обсуждении функций органов опираются на то, что целью их активности является сохранение гомеостаза, необходимого для поддержания жизни организма. Например, Гарвей так определял функцию сердца: основная функция сердца – это передача и перекачка крови по артериям к конечностям тела [Harvey, 1928]. Как эта цель (поддержание жизни) вписывается в общую картину мира – зависит от философской позиции.

У Платона целью всех сущностей, включая живые организмы, является стремление к идеалу, установленному Демиургом [Kraut, 2017]. У Аристотеля телеология, управляющая поведением живых существ, имманентна и направлена на достижение «конечных целей» [Shields, 2020]. Современные философы считают, что функциональное объяснение однозначно применимо к биологии и ссылаются на теорию Дарвина, чтобы обосновать практику приписывания функций. Различают два способа использования естественного отбора для обоснования биологической телеологии. Косвенные подходы рассматривают адаптивную, самоорганизующуюся природу живых организмов как естественную основу телеологических свойств их черт, но отдают должное способности естественного отбора создавать такую самоорганизующуюся сложность. Прямые подходы в явном виде ссылаются на естественный отбор при объяснении функций либо в этиологическом смысле, основанном на истории отбора, либо в диспозиционном смысле, основанном на приспособленности организмов, обладающих данными признаками [Allen, Neal, 2020].

Мораль может иметь функции в контексте социальной группы как средство поддержания кооперации и сохранения единства этой группы. Например, знаменитое «золотое правило» («относись к другим так, как ты хочешь, чтобы относились к тебе»), в

той или иной форме присутствующее во всех религиозных и нерелигиозных системах морали [Spooner, 1914], может выполнять функцию сплочения людей в обществе. Но мораль может также иметь функции в контексте регуляции поведения и мыслей отдельного человека.

Далеко не всегда моральный поступок поддерживает единство группы, к которой принадлежит человек, совершивший этот поступок. Можно, например, представить ситуацию, когда кто-то открыто выступает против доминирующего в обществе режима, навлекая тем самым на себя и, возможно, близких людей, негативные последствия. У кого-то поступок вызовет восхищение, у кого-то раздражение и явно не будет способствовать сохранению единства группы. Для данного человека, однако, он представляется необходимым, так как соответствует его внутреннему моральному стандарту и позволяет сохранить «верность себе». Эту функцию морали мы рассмотрим позже при обсуждении темы самости. Заметим, что даже в таком контексте функции морали связаны с регуляцией социального поведения. Принимая в расчет эту явную социальную направленность морали, а также то, что, по наиболее популярной в настоящее время теории, движущей силой эволюции мозга у гоминидов были именно социальные факторы [Dunbar, 1998], можно было бы ожидать, что мораль должна иметь явное и значительное представительство в мозге.

Действительно, способность интуитивно различать морально приемлемые и неприемлемые поступки представляется более важной для поддержания гармонии в социальной группе, чем, например, способность к абстрактному мышлению. Соответственно, по убеждениям эволюционных психологов, моральные суждения должны возникать в результате активности появившихся в процессе эволюции «предметно-ориентированных психологических модулей» (*domain-specific psychological modules*), т. е. специализированных нейронных «вычислительных систем». Согласно этой идее, в мозге должны существовать области, специализирующиеся на обеспечении моральных суждений. По альтернативному представлению, моральные суждения зависят от активности модулей общего назначения (*domain-general psychological modules*) [Cosmides, Tooby, 2008].

Существующие эмпирические данные определенно показывают, что в мозге нет «морального центра» и «моральный мозг» можно разбить на несколько модулей, чьи изначальные функции

не имеют ничего общего с моралью [Verplaetse et al., 2009]. Например, немалое количество исследований связывает «моральные чувства», такие как отвращение к инцесту [Schaich Borg et al., 2008], неприятие несправедливых предложений [Sanfey et al., 2003], эмпатическое страдание от наблюдения за болью, которую испытывают близкие родственники [Singer et al., 2004], и возмущение исключением из группы [Eisenberger et al., 2003], с активностью передней части островка (anterior insular cortex). Однако хорошо известно, что чувство базового отвращения также активирует эту область мозга [Calder et al., 2001], а электрическая стимуляция вызывает неприятные и трудно переносимые ощущения в полости рта [Penfield, Faulk, 1955; Krolak-Salmon et al., 2003].

Нейровизуализационные исследования подтверждают важную роль островка в возникновении неприятных ощущений запаха и вкуса [Small et al., 2003; Heining et al., 2003]. Наконец, исследования последних лет связывают эту часть мозга с регуляцией процессов внимания [Seeley et al., 2007]. Некоторые моральные эмоции имеют много общего с переживанием чувства отвращения [Haidt, 2003], что может объяснять активацию островка в соответствующих ситуациях. Эти данные подтверждают общий принцип, лежащий в основе большинства исследований мозга, – области, изначально «предназначенные» для выполнения задач «нижнего уровня», используются для реализации «высших» умственных задач [Verplaetse et al., 2009].

Чтобы «произвести» эмоции, суждения или поведение, называемые нами моральными, мозг активирует эти базовые модули и сети. Сочувствие боли, испытываемой другим человеком, активирует переднюю часть поясной извилины – область, которая связана (помимо всего прочего) с реакцией на негативные ощущения [Singer et al., 2004]. Злорадство вызывает повышенную активность в прилежащем ядре, играющем важную роль в обеспечении чувства удовлетворенности [Singer et al., 2006]. Решение сотрудничать с незнакомым человеком во время игры «дилемма заключенного» сопровождается активацией хвостатого ядра – области, участвующей в системе вознаграждения [Rilling et al., 2002]. Решение наказать несправедливого партнера стимулирует нейронную активность в правой дорсолатеральной префронтальной коре, связанной с интеграцией эмоций и когниций [Sanfey et al., 2003]. Нет ни одной области мозга, которая была бы связана исключительно с моралью. Все часто цитируемые в связи с процессами

морали области являются частью различных мозговых цепей, выполняющих моральные задачи в дополнение к ряду других задач [Verplaetse et al., 2009].

Нейровизуализационные исследования мозгового обеспечения процессов, связанных с принятием моральных решений, имеют те же недостатки, что и исследования мозговых коррелятов многих других ментальных процессов. Прежде всего не нужно забывать, что все данные в этих исследованиях получают в процессе экспериментов, во время которых испытуемый лежит в сканере и общается с экспериментатором или партнером посредством микрофона и наушников или визуальных стимулов на экране. Ситуации морального выбора моделируют в эксперименте либо с помощью показа на экране изображений соответствующих ситуаций, либо с помощью игр типа «дилемма заключенного» с экспериментатором или неизвестным партнером. Испытуемыми, как правило, являются студенты психологических факультетов университетов. В какой степени получаемые в этих экспериментах данные отражают активность мозга в реальных жизненных ситуациях морального выбора – остается большим вопросом.

Вторым, хорошо известным недостатком подобных исследований является то, что они плохо воспроизводятся (т. е. повторный эксперимент, поставленный той же или другой исследовательской группой, часто не воспроизводит полученный ранее эффект). Эта проблема остро стоит практически для всех сфер нейробиологии «высших» ментальных процессов. Плохая воспроизводимость по сути означает отсутствие надежных связей между активностью мозга и соответствующими ментальными процессами. Выявляемые в отдельных работах эффекты зависят от случайных, плохо контролируемых факторов. Например, недавние исследования ставят под сомнение решающую роль передней части островка в социоморальном отвращении. Вопреки ожиданиям, морально отталкивающие проступки, не связанные с сексом, такие как насильственные нападения на невинных людей, не увеличивают активность в этой области мозга [Schaich Borg et al., 2008]. Вторая проблема – произвольность интерпретации выявленных эффектов.

Как я уже отмечал ранее, стандартный подход в нейровизуализационных исследованиях – это интерпретация эффектов активации определенных областей мозга в свете предшествующих исследований функциональных коррелятов этих областей, что по сути является циркулярным объяснением. Часто исследователи

формулируют задачу эксперимента исходя из теоретической концепции, которая не делает четких и конкретных предсказаний в отношении активности тех или иных областей мозга. Поскольку данные о функциональных коррелятах разных областей мозга тоже достаточно неопределенны, то практически любой выявленный эффект можно интерпретировать как подтверждающий теорию. При просмотре литературы о функциональных коррелятах какой-то области мозга всегда можно подобрать коррелят, который можно интерпретировать в свете тестируемой теории. Противоречивые интерпретации умалчиваются.

Двойные толкования, часто имеющие место в ходе обсуждения результатов, вызывают особое беспокойство, поскольку они всегда соответствуют данным. Предположим, в каком-то исследовании психопаты демонстрируют повышенную активность в ДЛПК при выполнении экспериментальной задачи. К сожалению, согласно теории, мы должны ожидать меньшей активации ДЛПК у психопатов в этом эксперименте. Используя здравые рассуждения, следует заключить, что данные опровергают теорию. Двойные интерпретации позволяют избежать такого вывода. Можно утверждать, что психопаты прилагают больше усилий, чтобы достичь такого же уровня выполнения задачи, т. е. по сравнению со здоровыми участниками им требуется большая активность ДЛПК. Большинство читателей не заметят этого неявного сдвига в интерпретации от «нормального возбуждения» к «дополнительному усилию». Более того, учитывая текущее положение дел в сфере нейровизуализационных исследований, мы не можем опровергнуть это надуманное объяснение, так как сканеры не различают естественную активацию от компенсирующих усилий. Такая уловка помогает избежать неутешительных результатов [Verplaatse et al., 2009].

Помимо нейровизуализации, арсенал методов изучения функций мозга пополнился в последние десятилетия методами транскраниальной магнитной стимуляции (ТМС), позволяющими активировать или подавлять активность поверхностно расположенных областей мозга. Было опубликовано несколько работ, в которых ТМС использовалась при изучении мозговых коррелятов морали. В исследовании Дарьи Кнох с соавторами, опубликованном в 2006 г. в журнале *Science*, испытуемые участвовали в игре «ультиматум» [Knoch et al., 2006]. В этой простой игре первый игрок предлагает поделить определенную сумму денег (скажем, 10 руб-

лей) с другим испытуемым. Например, он решает взять 6 рублей себе, чтобы другой получил 4 рубля. Ответчик имеет право наложить вето на это решение. Если он воспользуется своим правом, сделка срывается и оба игрока ничего не получают. При несправедливых предложениях испытуемые склонны блокировать сделку, хотя с рациональной точки зрения это неразумно – даже если вторая сторона оставляет 9 рублей себе и дает напарнику 1 рубль, выгодней принять данное предложение, чем заблокировать сделку и ничего не получить.

В предшествующем фМРТ-исследовании было показано, что у испытуемых, получающих несправедливые предложения, наблюдается активация в островке, передней поясной коре и ДЛПК [Sanfey et al., 2003]. В исследовании Кнох с соавторами использовали ТМС для торможения правой ДЛПК и показали, что в этом случае треть испытуемых соглашалась со всеми предложениями, даже максимально несправедливыми. Так, на несправедливое предложение поделить сумму в пропорции 8/2 в экспериментальной группе соглашалось 45 % респондентов (из 52), а в контрольной – не более 10 %. При этом ТМС не влияла на моральную оценку предложений. По результатам анкетирования все участники, независимо от того, получали они ТМС или нет, одинаково считали неравные предложения несправедливыми.

Предполагается, что торможение ДЛПК препятствовало тому, что моральное негодование реализовалось в соответствующий поведенческий ответ. Из этой работы, как и из других исследований с использованием ТМС, очевидна бóльшая вариабельность ответов испытуемых. У 55 % испытуемых, получавших ТМС, поведение не отличалось от поведения в контрольной группе. Это согласуется с данными нейровизуализационных исследований и означает отсутствие надежной связи между активностью какой-то части мозга и соответствующими ментальными процессами или поведением.

Альтернативный ТМС подход к манипулированию активностью мозга – использование фармакологических субстанций. В сфере изучения морали и социального поведения наиболее часто используют окситоцин – нейропептид, который играет ключевую роль в регуляции социальной привязанности у млекопитающих. В исследовании Майкла Косфельда с соавторами, опубликованном в журнале Nature [Kosfeld et al., 2005], испытуемые получили интраназальную дозу окситоцина, а контрольная группа

получала плацебо. После этого они участвовали в игре доверия (trust game), в которой каждый испытуемый играет роль либо инвестора, либо попечителя. У инвестора есть опция выбрать рискованный вариант передачи денег попечителю. При этом сумма автоматически утроится, но попечитель в конечном итоге выбирает, как разделить инвестиции. Он может обмануть доверие инвестора, оставив вложенную сумму себе, поэтому рациональные инвесторы склонны не переводить ничего (или только небольшие суммы денег) ввиду того, что эгоцентрические попечители могут обмануть доверие инвесторов.

Косфельд с соавторами обнаружили, что введение окситоцина влияло на готовность инвесторов идти на риск – по сравнению с контрольной группой они переводили попечителям в среднем на 17 % больше денег. Так же, как и в случае использования ТМС, окситоцин не влиял на всех испытуемых одинаково и значительная часть из них, получавших окситоцин, демонстрировала поведение, не отличимое от поведения в контрольной группе. Кроме того, окситоцин не влиял на решения, принимаемые попечителями. Почему окситоцин влияет на одних людей больше, чем на других, и почему он не повлиял на поведение попечителей – остается неясным.

В целом, результаты экспериментов по изучению мозговых коррелятов морали дают достаточно пеструю и неопределенную картину. Хотя термины типа «моральный мозг» или «социальный мозг» очень часто можно встретить в соответствующих работах, наполнение этих терминов в плане набора конкретных структур мозга и их предполагаемых функций в соответствующих когнитивных процессах остается очень туманным. Значительная часть дебатов в области изучения мозговых коррелятов морали крутится вокруг вопроса о преобладании эмоций или разума в принятии моральных решений.

Вопрос этот, как мы видели, берет начало в философских концепциях Юма и Канта, и предполагалось, что исследование мозга позволит окончательно поставить точку в споре философов. Этого, однако, не произошло, и исследователи чаще всего интерпретируют эмпирические данные о мозговых коррелятах морали в соответствии со своей исходной позицией о преобладании тех или иных процессов. Так, в «моральный мозг», помимо островка и передней части поясной извилины, как правило, включают орбитофронтальный кортекс и вентральный стриатум, которые обычно

относят к мозговой системе вознаграждения [Schulz, 1998], а также медиальные и вентральные части префронтальной коры, угловую извилину (преимущественно правую), заднюю часть поясной извилины и амигдалу [Glenn, Raine, 2009].

Приверженцы сентиментализма, т. е. позиции, согласно которой принятие моральных решений в основном связано с эмоциями, считают, что эти области участвуют в обеспечении эмоциональной составляющей, и нужно строить модели, объясняющие, как они это делают [Prinz, 2016]. Другие утверждают, что именно рассуждение играет решающую роль в принятии моральных решений, и, соответственно, все эти области мозга участвуют в рассудочной активности [Kennett, Gerrans, 2016]. Сторонники умеренной позиции доказывают важность как суждений, так и эмоций, и области мозга, обычно определяемые как вовлеченные в эмоциональную обработку, каузально участвуют в построении моральных суждений путем «вычисления» значений, связанных с положительным и отрицательным подкреплением. В частности, вентромедиальная префронтальная и орбитофронтальная кора интегрируют сигналы от мозговой системы вознаграждения и от различных когнитивных систем, обеспечивая таким образом мотивационную составляющую моральных суждений [Woodward, 2016].

Термины типа «вычисление значения» и «интеграция информации» повсеместно используются в подобных рассуждениях и их использование мотивируется аллюзиями к ВТР и аналогиям с компьютером. Однако, как мы видели при рассмотрении ВТР (см. раздел 4.3.2), теории эти остаются чисто умозрительными, и практически ничего не известно по поводу того, как подобные «вычисления» могут производиться в мозге. Напомню еще раз, что все эти области были выявлены в фМРТ-исследованиях с использованием модельных ситуаций принятия «моральных решений» и в результате статистического анализа, выявляющего «достоверный» в данной выборке испытуемых эффект.

В качестве эффекта обычно рассматривается разница величины фМРТ-сигнала между представляющими интерес (например, ситуация морального выбора) и контрольными (когнитивно похожая задача, но без морального выбора – например, выбор покупки в магазине) трейлами. Результаты такого анализа сильно зависят от содержания контрольных трейлов и должны интерпретироваться в контексте этого содержания, что не всегда достаточно учитывается при описании данных. В любом случае экспериментальные

данные лишь показывают, что перечисленные области мозга имеют тенденцию потреблять больше кислорода в модельных ситуациях морального выбора, чем в покое, или модельных ситуациях другого рода выборов.

Поскольку все эти области крайне полифункциональны (их активация выявлена в большом количестве разнородных ситуаций) и поскольку отсутствует даже спекулятивное представление о том, как конкретно они объединяются в «моральный мозг» и обеспечивают принятие моральных решений, полученные данные очень далеки от того, чтобы представить механистическую картину принятия моральных решений в мозге. Поэтому, по факту, накопленные нейровизуализационные данные практически ничего не дают для решения философских вопросов в сфере морали, которое продолжает опираться в основном на данные, полученные с помощью традиционных психологических методов [Gonzalez-Sabrera, 2017]. Как и в случае эмоций или когнитивных состояний и процессов, можно лишь утверждать, что моральные интуиции и решения являются по отношению к активности мозга эмерджентной сущностью. Их содержание невозможно «свести» или объяснить на основе регистрируемой активности мозга.

8.4. Красота

Говоря о ценностных ориентациях, невозможно обойти тему красоты, которая традиционно считалась одной из высших ценностей, вместе с добром, правдой и справедливостью. Природа красоты – одна из самых противоречивых тем в западной философии. Вероятно, наиболее известный вопрос теории красоты состоит в том, является ли она субъективной или объективной характеристикой. До XVIII в. большинство философов рассматривали красоту как объективное качество. Хотя Платон и Аристотель расходятся во мнениях относительно того, что такое красота, оба считают ее объективной. Классическая концепция рассматривает красоту как вопрос установления определенных пропорций или отношений между частями, иногда выражаемых в математических соотношениях, например, золотое сечение [Plotinus, 1952].

Августин прямо спрашивает, красивы ли вещи потому, что они доставляют удовольствие, или они доставляют удовольствие, потому что они красивы, и решительно выбирает второй вариант [Augustine, 1953]. Уже в XVIII в., однако, Юм писал, что красота

не является качеством самих вещей – она существует только в уме, созерцающем их, и каждый ум воспринимает различную красоту. Один человек может даже воспринимать уродство там, где другой видит красоту, и каждый человек должен следовать за своими собственными чувствами, не претендуя на то, чтобы регулировать чувства других [Hume, 1894].

Похожую точку зрения высказывал Кант. Суждение вкуса, по его представлениям, не есть суждение познания и, следовательно, является не логическим, а эстетическим, т. е. оно может быть не чем иным, как субъективным [Kant, 1951]. Однако если красота исключительно субъективна, то это слово фактически теряет смысл, и мы ничего не сообщаем, когда называем что-то красивым, кроме одобрительного личного отношения. Вдобавок, хотя разные люди, конечно, могут различаться в конкретных суждениях, очевидно, что наши суждения в значительной степени совпадают. Было бы странно отрицать великолепие идеальной розы или грандиозной сцены заката. Можно не соглашаться и спорить о том, красиво ли что-то, или попытаться показать кому-то, что что-то красиво, или узнать от кого-то, почему это так. Это бы не имело смысла, если бы красота считалась сугубо субъективной, как предпочтении сладкого или соленого [Sartwell, 2017].

И Юм, и Кант понимали потерю чего-то важного, когда к красоте относятся просто как к субъективному состоянию. По их убеждению, если красота приравнивается к индивидуальным ощущениям, она перестает быть первостепенной ценностью или даже узнаваемой как ценность для всех людей или обществ. С другой стороны, кажется бессмысленным утверждение о красоте, не связанной с субъективным восприятием, или о ее полной объективности. Это может означать, что мир без наблюдателей может быть красивым или уродливым, или что красоту можно обнаружить с помощью научных инструментов. Поэтому и Юм, и Кант трактуют суждения о красоте не как чисто субъективные и не как объективные, а как интерсубъективные, или имеющие социальный и культурный аспект, и концептуально подразумевающие интерсубъективные претензии на достоверность [Ibid]. По мнению Юма, суждения критиков, обладающих развитым вкусом, как правило, совпадают и в конечном итоге приближаются к согласию, что объясняет, например, непреходящее почитание поэм Гомера. Таким образом, испытание временем, оцененное вердиктами лучших критиков, действует как нечто подобное объективному стандарту.

В последующие годы из трактовок красоты Юмом и Кантом была преимущественно воспринята ее субъективность. Соответственно, в XX в. философы в значительной степени перестали уделять внимание теме красоты. Если красота – это лишь субъективное удовольствие, то она не имеет более высокого статуса, чем все, что просто развлекает или отвлекает, она несопоставима по важности с истиной или справедливостью. Двадцатый век отказался также от красоты как доминирующей цели искусства, возможно, отчасти потому, что ее принижение в теории заставило художников поверить в необходимость стремления к более реальным и серьезным целям [Danto, 2003]. Кульминацией этого направления является антиэстетика постмодернизма [Foster, 1998].

Однако в последние десятилетия наблюдается возрождение интереса к красоте как в искусстве, так и в философии. Американский философ-аналитик Гай Сирчелло предложил свою теорию красоты как попытку подтвердить ее статус в качестве важной философской концепции. Он отвергал субъективизм Канта и стремился определить свойства, присущие объекту, которые делают его красивым. Он называл такие качества, как яркость, смелость и тонкость, «свойствами качественной степени» (PQD) и заявлял, что они делают объект красивым, если в нем отсутствует видимость недостатка или дефекта [Sircello, 1989].

Некоторые современные попытки развития теории красоты перекликаются с позицией Мура, согласно которой сказать, что вещь прекрасна, значит на самом деле сказать не то, что она хороша сама по себе, но что это необходимый элемент в чем-то. Доказать, что вещь действительно прекрасна, – значит доказать, что это целое, к которому она имеет особое отношение как часть, действительно хорошо [Moore, 1903]. Одно из толкований этого высказывания заключается в следующем: фундаментально ценной является ситуация, в которую и объект, и субъект встроены; ценность красоты включает в себя как черты прекрасного объекта, так и удовольствия человека, переживающего его.

Криспин Сартвелл в своей книге «Шесть имен красоты» [Sartwell, 2004] приписывает красоту не только субъекту или объекту, но и отношениям между ними, и даже в более широком смысле также ситуации или среде, в которую они встроены. Он указывает, что когда мы, например, описываем красоту ночного неба, мы не просто сообщаем о состоянии удовольствия, а обращены вовне к реальному миру. С другой стороны, если бы не было

людей, способных воспринимать такие вещи, не было бы красоты. Красота возникает в ситуациях, когда субъект и объект связаны. У красоты также есть социальный аспект. Красота – это то, чем мы делимся или хотим поделиться с другими, и совместный опыт красоты – это особенно интенсивные формы общения. Таким образом, переживание красоты происходит в первую очередь не внутри черепа переживающего, а связывает наблюдателей и объекты, такие как произведения искусства и литературы, в сообщества ценителей. Эстетическое суждение не требует всеобщего согласия, и ни красивый объект, ни произведение искусства не создают религиозное сообщество, но они создают меньшие общества, не менее важные и серьезные, и, с точки зрения их членов, каждое из них ортодоксально, однако, все остальные при этом не считаются еретиками [Nehamas, 2007].

Особенности восприятия красоты психологами и нейробиологами в сфере экспериментальной эстетики и нейроэстетики будут изложены ниже. Рассмотрим наиболее важные концепции красоты.

Классическая концепция красоты заключается в том, что она представляет собой объединение составных частей в единое целое в соответствии с пропорциями, гармонией, симметрией и другими подобными понятиями [Wölfflin, 1932]. Это исконная западная концепция красоты, воплощенная в классической и неоклассической архитектуре, скульптуре, литературе и музыке, где бы они ни появлялись. По мнению Аристотеля, чтобы быть красивым, живое существо и любое другое целое, состоящее из частей, должно представлять определенный порядок в своем расположении частей. Главные формы красоты – это порядок, симметрия и определенность, которые математические науки демонстрируют в особой степени [Aristotle, 1984].

Этот взгляд, как предполагает Аристотель, иногда сводится к математической формуле, такой как золотое сечение, но нет необходимости рассматривать его в таких строгих терминах. Примером данной концепции являются прежде всего такие тексты, как «Элементы» Евклида и такие архитектурные произведения, как Парфенон, а также Канон скульптора Поликлета (конец V – начало IV в. до н.э.). (Замечу, канон был не только статуей, изображающей идеальные пропорции, но и ныне утерянным трактатом о красоте).

Важно отметить, что понятие «симметрия» в классических текстах отличается от его нынешнего использования для обозначения двустороннего зеркального отражения. Оно относится к

видам гармоничных и измеримых пропорций среди частей, характерных для предметов, прекрасных в классическом смысле, и имеет также моральный вес. Например, в «Софисте» Платон описывает добродетельные души как симметричные. Фома Аквинский в типично аристотелевской плюралистической формулировке отмечает, что для красоты есть три требования. Во-первых, целостность или совершенство – если что-то повреждено, это некрасиво. Затем должная пропорция или созвучие, а также ясность [Aquinas, 1981].

По определению Фрэнсиса Хатчесона, то, что мы называем прекрасным в объектах, говоря математическим языком, находится в сложном соотношении единообразия и разнообразия [Hutcheson, 2004]. Представитель романтизма Эдмунд Берк не согласен с классическим определением красоты и приводит серию опровержений и контрпримеров идее о том, что красота может быть вопросом любых пропорций между частями: художник, соблюдая любые пропорции каждой части человеческого тела, тем не менее, может, если ему угодно, изобразить очень некрасивую фигуру [Burke, 1990].

По идеалистической концепции, представленной у Платона, красота понимается как совершенное единство, или даже как сам принцип единства. Начиная с индивидуальной красоты, в поисках универсальной красоты человек должен подниматься по небесной лестнице, шагая от ступеньки к ступеньке, т. е. от одного к двум и от двух к каждому прекрасному телу, и от телесной красоты к красоте души, от души к обучению и от обучения в целом к специальным знаниям, которые не имеют отношения ни к чему, кроме самой красоты, – пока, наконец, он не узнает, что такое красота [Plato, 1961].

Неоплатонисты отождествляли красоту с идеальной формой, которая является источником единства разрозненных вещей, и сама по себе – совершенное единство [Plotinus, 1952]. Плотин критикует классическую концепцию красоты, по которой только сложные, состоящие из компонентов объекты могут быть красивыми. Он указывает, что красота совокупности требует красоты деталей. Симметричное и «гармоничное» расположение компонентов, каждый из которых сам по себе уродлив, не может быть красивым. Красивыми могут быть не только сложные, но и простые вещи, такие как вспышка молнии в ночи или свет звезды. В музыкальной композиции каждый отдельный тон может быть прекрасен.

Наиболее красивым физическим явлением для Плотина является пламя – всегда устремленное вверх, тончайшее и наиболее живое из всех тел, наиболее близкое к бестелесному [Plotinus, 1952]. Для Плотина, как и для Платона, вся множественность в конечном итоге должна быть слита в единство, и все пути познания ведут к Добру/Прекрасному/Истинному/Божественному. Это породило мистическое в своей основе видение красоты Бога, которое, как утверждал Умберто Эко, сохранялось на протяжении всего Средневековья: наслаждение изобилием, которое в конечном итоге сливается в единое духовное единство [Eco, 1959].

Чувственные/эстетические наслаждения рассматриваются как выражение восхищения безмерным прекрасным изобилием Бога [Pseudo-Dionysius, 1897]. Эта концепция получила множество выражений и в более позднюю эпоху, в том числе в таких фигурах, как Шефтсбери, Шиллер и Гегель, согласно которым, эстетика или переживание искусства и красоты является мостом между материальным и духовным. Выражение Шиллером аналогичной серии мыслей оказало фундаментальное влияние на концепции красоты, разработанные в рамках немецкого идеализма. Для Шиллера красота, игра или искусство (он использует эти слова почти взаимозаменяемо) осуществляют процесс интеграции естественного и духовного, или чувственного и рационального [Schiller, 1795]. Это очень похоже на «лестницу» Платона: красота как способ подняться к абстрактному или духовному. Но Шиллера больше интересует интеграция царств природы и духа, которую выполняет красота и искусство, чем выход за пределы физической реальности.

Очень похожую концепцию красоты можно найти и у Гегеля. Философское понятие прекрасного, по мнению Гегеля, должно содержать согласованные внутри себя обе крайности – идеальное и эмпирическое, поскольку оно объединяет метафизическую универсальность с реальной частностью. Красота, или, по крайней мере, художественная красота, – это путь от чувственного и особенного к Абсолюту и к свободе, от конечного к бесконечности [Hegel, 1975]. И Гегель, и Шефтсбери, которые связывают красоту и искусство с разумом и духом, считают, что красота искусства выше красоты природы, потому что мир природы рожден от Бога, но красота искусства снова преобразует этот материал духом художника.

Чувство красоты у идеалистов связано с любовью, понимаемой как неутолимая тоска и стремление к идеальному. Любовь всегда находится в состоянии недостатка и, следовательно, в жела-

нии – желании обладать прекрасным [Plato, 1961]. Эта идея возрождалась многократно – например, романтиками. Она подпитывала культ идеализированной, или куртуазной, любви в средние века, когда возлюбленные становились символом бесконечности. Недавние работы по теории красоты возродили эту идею, отказавшись от удовольствия и обратившись к любви или тоске (которые не обязательно являются полностью приятными переживаниями) как эмпирическим коррелятам красоты. Сартвелл определяет красоту как «объект стремления» и характеризует тоску как сильное и неудовлетворенное желание. Он называет это фундаментальным состоянием конечного во времени существа, когда мы всегда находимся в процессе потери всего, что имеем, и поэтому неизбежно пребываем в состоянии тоски [Sartwell, 2004]. Нехамас считает красоту символом того, чего нам не хватает, – красивые вещи оживляют чувство жизни, придавая ему новую форму и направление [Nehamas, 2007].

В противоположность классикам и идеалистам, мыслители XVIII в., многие из которых были ориентированы на эмпиризм, отождествляли красоту с удовольствием. Под прекрасным они понимали все, что восхищает, радует и вызывает приятные ощущения (Ludovico Antonio Muratori, цит. по: [Carritt, 1931]). Из изложения Хатчесона неясно, должны ли мы понимать красоту в первую очередь в терминах классических формальных элементов или в терминах реакции зрителя. По его утверждению, объекты, которые представляют «сложное соотношение единообразия и разнообразия», способны доставлять удовольствие [Hutcheson, 1725].

Юм считал, что красота – это такой порядок и построение частей, которые, по базовому строению нашей природы, по обычаю или по прихоти, способны доставить нам удовольствие и удовлетворение. Удовольствие и боль, таким образом, – не только необходимые спутники красоты и уродства, но и составляют их самую суть [Hume, 1978]. Вопрос о том, почему восприятие красивых объектов сопровождается чувством удовольствия, остается за скобками.

В течение двух столетий после «Третьей критики» Канта идея о прямой связи между красотой и удовольствием стала считаться настолько очевидной, что мыслители часто определяли красоту как определенный вид удовольствия [Santayana, 1896]. Одним из результатов подобного подхода к красоте – или, возможно, крайнего выражения этой ориентации – является утверждение позитив-

вистов о том, что такие слова, как «красота», бессмысленны или лишены познавательного содержания, или являются просто выражением субъективного одобрения. Такие слова, как «красивый» и «отвратительный», используются не для констатации фактов, а просто для выражения определенных чувств и вызова определенной реакции. Соответственно, нет смысла приписывать объективную значимость эстетическим суждениям и спорить по вопросам ценности в эстетике [Ayer, 1952]. Вопрос красоты – это не стоящий вопрос, и мы можем спокойно оставить его в покое. Большинство философов XX в. именно так и поступали [Sartwell, 2017].

Другое направление в объяснении «функции» красоты – ее предполагаемая связь с полезностью. Кант, противник такого подхода, считал, что суждение о чем-то красивом должно быть беспристрастным, т. е. не должно зависеть от обычных человеческих желаний, например, экономических или сексуальных. Если кто-то, гуляя по музею, восхищается картинами, потому что они были бы чрезвычайно дорогими, если их выставить на аукционе, или задаваясь вопросом, нельзя ли их украсть и продать, то он не испытывает чувства красоты этих картин [Kant, 1951]. Философы кантианской традиции отождествляют переживание красоты с бескорыстным удовольствием, психической дистанцией и т. п. и противопоставляют эстетическое практическому. Вкус – это, по их мнению, способность судить об объекте или способе его представления по совершенно бескорыстному удовлетворению или неудовлетворенности. Объект такого удовлетворения называется прекрасным [Ibid]. В наше время Роджер Скратон в своей книге «Красота» возвращается к модифицированному кантианству в отношении как красоты, так и возвышенности. Он отмечает, что мы называем что-то прекрасным, когда получаем удовольствие от созерцания этого как отдельного объекта, ради него самого [Scruton, 2009].

Многие философы пошли в противоположном направлении и отождествили красоту с удобством использования (например, Беркли, см.: [Carritt 1931]). Сугубо утилитарное понимание красоты, однако, противоречит здравому смыслу. Полезность лопаты, например, ни у кого не вызывает сомнения, но вряд ли кто-то увидит в ней красоту и сделает ее объектом любви и преклонения.

Кантом предложена своя интерпретация темы полезности, или видимой «целенаправленности» красоты: объект считается красивым, когда кажется, что он демонстрирует «целесообраз-

ность»; т. е. когда его форма воспринимается как имеющая характер вещи, созданной в соответствии с некоторым принципом и приспособленной для определенной цели. Он отличал «свободную красоту» от «зависимой красоты», объясняя, что «первая не предполагает никакого представления о том, каким должен быть объект; вторая действительно предполагает такое понятие и совершенство объекта в соответствии с ним». Согласно этому определению, свободную красоту можно видеть в морских ракушках и музыке, а зависимую в зданиях и человеческом теле [Kant, 1951].

Таким образом, идеалистические и тяготеющие к ним интерпретации рассматривают чувство красоты как восприятие латентного идеала. По этой интерпретации за красотой отдельного объекта и всего мира в целом стоит гармония высшего порядка. Способность «увидеть» эту гармонию за завесой частных реального мира различна у разных людей. Сила и характер соответствующего чувства варьируют как у разных людей, так и у одного человека в зависимости от контекста ситуации и особенностей объекта, но то общее, что можно выделить во всех случаях, – и есть ощущение красоты, которое по сути является смесью восхищения и тоски от понимания недостижимости этой гармонии для смертного существа.

В противоположность этому, материалистические и тяготеющие к ним интерпретации рассматривают чувство красоты как восприятие полезности или «приятности» тех или иных воздействий, объектов или явлений. По этой интерпретации чувство красоты принципиально не отличается от чувства удовольствия, получаемого, например, от поедания вкусной пищи или занятий сексом. Однако, если в последнем случае чувство удовольствия легко объяснить как субъективное проявление присущей всем живым существам базовой тенденции к сохранению гомеостаза и размножению (и, соответственно, обосновать существование этого чувства естественным отбором), то чувство красоты трудно связать с базовыми биологическими процессами или даже процессами культурной эволюции.

Очевидно, что восприятие красоты и его отражение в произведениях искусства существует у представителей всех культур, и ископаемые остатки музыкальных инструментов и наскальной живописи показывают, что оно существовало на самых ранних этапах появления современного человека и даже у неандертальцев. Очевидно также, что характер чувств, сопровождающих вос-

приятие красоты, имеет мало общего с характером чувств, сопровождающих поедание вкусной пищи или занятие сексом. Далеко не всегда их можно отнести к разряду удовольствий. Они могут сопровождаться чувством печали («слезы наворачиваются на глаза»), или благоговейного ужаса. Это особенно относится к восприятию «возвышенного» (sublime) как в явлениях природы, так и в произведениях искусства. Хотя некоторые философы выделяли восприятие возвышенного в отдельную категорию и даже противопоставляли его восприятию красоты [Burke, 1990].

Несмотря на то, что потенциально можно выделить нечто общее в восприятии красоты разных объектов, невозможно отрицать различия, связанные с восприятием в них красоты. Прежде всего это касается восприятия красоты в явлениях и объектах природы с одной стороны, и в произведениях искусства или в интеллектуальной деятельности – с другой. Мы не будем останавливаться на поистине неисчерпаемой теме красоты в искусстве и лишь вспомним еще раз мысль Гегеля о том, что красота искусства выше красоты природы в том смысле, что художник извлекает и дистиллирует красоту при работе над произведением [Hegel, 1835].

Красоту видят не только в произведениях искусства, но и в других продуктах интеллектуальной деятельности. Часто говорят, например, о красоте идеи или красоте математической формулы. Очевидно, что восприятие красоты в объектах природы и в плодах интеллектуальной деятельности имеет как что-то общее, так и элементы различия. Когда мы видим гениальное произведение искусства или «красивое» физическое уравнение, мы восторгаемся не только красотой природного явления, послужившего прототипом или побудительной причиной для творца, но и, прежде всего, интеллектуальным свершением автора. Ощущение красоты в этом случае является для нас подтверждением истинности того, что художник или ученый увидел. Поэтому попытки «объективизации» красоты, т. е. попытки алгоритмизировать свойства объекта или явления, вызывающие ощущение красоты, могут дать разные результаты при оценке красоты природного явления и продуктов интеллектуальной деятельности.

Если говорить только о природных объектах, то можно выделить в особую категорию восприятие красоты человеческого лица и тела. Этот объект настолько важен для человека, что черты человеческого лица или жесты часто переносят на описание других

объектов – как живых, так и неживых. Характеристика человека как «красивого» часто основана на некоторой комбинации внутренней красоты, которая включает психологические факторы, такие как личность, интеллект, изящество, вежливость, харизма, целостность, конгруэнтность и элегантность, и внешнюю красоту (т. е. физическую привлекательность), которая ценится с эстетической точки зрения.

Стандарты красоты менялись с течением времени в связи с изменением культурных ценностей. Картины и скульптуры демонстрируют широкий спектр различных стандартов красоты, однако относительно молодые люди с гладкой кожей, пропорционально сложенные и с правильными чертами лица традиционно считались более красивыми на протяжении всей истории человечества. Ярким показателем физической красоты является «среднестатистичность». Когда изображения человеческих лиц усредняются, чтобы сформировать составное изображение, они постепенно приближаются к «идеальному» изображению и воспринимаются как более привлекательные [Valentine et al., 2004]. Предполагается, что эволюционно выгодно, чтобы сексуальные существа чувствовали приязнь к партнерам, обладающим преимущественно средними чертами, поскольку это предусматривает отсутствие генетических или приобретенных дефектов [Highfield, 2008]. По альтернативной интерпретации усреднение устраняет случайные отклонения от латентного идеала (как писал Александр Блок: «Сотри случайные черты – И ты увидишь: мир прекрасен»). Есть данные, что предпочтение красивых лиц проявляется в раннем младенчестве и, вероятно, является врожденным [Kramer et al., 2019], и что критерии привлекательности одинаковы для разных полов и культур [Rhodes, 2006].

8.4.1. Эволюционные теории чувства красоты

В рамках эволюционной психологии эволюционная эстетика исходит из предположения, что основные эстетические предпочтения *Homo sapiens* эволюционировали с целью повышения выживаемости и репродуктивного успеха. Основываясь на этой теории, предпочтение цвета, пропорций тела партнера, формы, эмоциональные связи с объектами и многие другие аспекты эстетического опыта объясняют со ссылкой на эволюцию человека. Утверждается, например, что люди имеют сильные эстетические

предпочтения к ландшафтам, которые были хорошими средами обитания для предков. Когда маленьких детей из разных стран просят выбрать, какой пейзаж им больше нравится из набора стандартных пейзажных фотографий, предпочтение отдается саваннам с деревьями. Восточноафриканская саванна – это среда, в которой, как утверждается, происходила бóльшая часть человеческой эволюции. Также отдается предпочтение пейзажам с водой, как с открытыми, так и с лесными участками, с деревьями с ветвями, расположенными на подходящей высоте для лазания и сбора плодов, с играющими животными и с некоторым количеством облаков [Dutton, 2003].

Такого рода рассуждения вряд ли можно всерьез рассматривать в качестве доказательства эволюционного происхождения чувства красоты. Как, например, эта теория может объяснить красоту горных пейзажей? Снежная вершина или дикий утес, не пригодный для жизни и совершенно не похожий на восточноафриканскую саванну, может восприниматься как красивый, так же как величественные и опасные для жизни явления природы вроде водопада или грозы. А как насчет красоты звездного неба? Вспомним стихотворение Лермонтова «На севере диком» и одноименную картину Шишкина. Почему эта сосна, засыпанная снегом, кажется более прекрасной, чем увешанная бананами пальма? С точки зрения эволюционной теории опасные хищники вроде льва и тигра уж явно должны вызывать лишь страх и отвращение, тем не менее они считаются эталоном красоты в царстве животных (приведем в пример хотя бы знаменитое стихотворение «Тигр» Уильяма Блейка).

Эволюционное музыковедение стремится объяснить возникновение музыки эволюционными процессами. Уже Дарвин в своей книге «Происхождение человека» связывал появление музыки с процессами полового отбора. Первобытный человек или, скорее, какой-то ранний прародитель человека, вероятно, впервые использовал свой голос для создания истинных музыкальных ритмов, т. е. в пении, как это делают некоторые обезьяны-гibbonы в наши дни. Мы можем по аналогии сделать вывод, что это умение, которое особенно проявлялось во время ухаживания, способствовало выражению различных эмоций, таких как любовь, ревность, триумф, и служило для победы над соперником [Darwin, 1871].

Мы не будем рассматривать многочисленные более поздние теории, которые связывают появление музыки с половым отбором, появлением языка, двуногостью или устрашением соперников. Все эти теории остаются преимущественно спекулятивными, потому что, по очевидным причинам, они основываются не на прямых эмпирических данных, а на более или менее правдоподобных рассуждениях и фантазиях. Вторая общая черта всех этих теорий в том, что они могут объяснять, какую функцию могла выполнять музыка в первобытном обществе, но ничего не говорят о том, как эта функция выполнялась. Легко себе представить, по аналогии со сценами из кургузых романов, как два дикаря, пытаясь покорить сердце некой дикарки, соревнуются друг с другом на вокальном поприще. Трудно понять, однако, почему дикарка предпочитает пение одного пению другого. Этот вопрос, кстати, возникает и по поводу механизмов полового отбора у животных.

Классический пример полового отбора – хвост павлина, который явно не помогает ему в повседневной жизни, но обеспечивает привлекательность для противоположного пола. Для объяснения появления в эволюции подобных признаков Рональд Фишер предложил теорию безудержного отбора (*runaway selection*), согласно которой отбираемый признак (например, плюмаж павлина) изначально коррелировал с лучшей приспособленностью и самки отдавали предпочтение обладающим этим признаком самцам в силу их большей «перспективности» в плане выживания. В дальнейшем, когда в результате полового отбора этот признак так гипертрофировался, что уже перестал коррелировать с приспособленностью, самки «вопреки здравому смыслу» упорно продолжали его выбирать, что и привело к появлению изумительных по красоте павлиньих хвостов [Fisher, 1930].

В этом сценарии, хотя и математически безупречном, остаются неясности. Во-первых, непонятно, почему орнаментированный хвост мог изначально коррелировать с лучшей приспособленностью. Во-вторых, непонятно, почему самки, изначально столь чувствительные к потенциальной выживаемости партнеров, потом «потеряли голову» и стали преследовать цели, противоречащие первоначальному идеалу. В-третьих, трудно объяснить, почему возникшее в результате этого «слепого» процесса явление (хвост павлина) вызывает восхищение у человека. Проще уж вернуться к объяснению самого Дарвина, который предполагал наличие

эстетического вкуса у высших животных [Gayon, 2010]. Возвращаясь к эволюционному музыковедению, невозможно не заметить, что симфонии Бетховена и Моцарта так же отличаются от брачных криков обезьян-гibbonов, как современные математические теории от счета на пальцах. Даже если процессы естественного отбора участвовали в возникновении каких-то предпочтений, они не могут объяснить появление и развитие музыки и других видов искусства.

8.4.2. Современные эмпирические данные и теории

Экспериментальная эстетика – одно из самых первых направлений в экспериментальной психологии, берущее начало в работах немецкого психолога XIX в. Густава Теодора Фехнера. Согласно Фехнеру, эстетика – это восприятие свойств объекта, которое эмпирически постижимо в свете характеристик как субъекта, переживающего опыт, так и объекта [Fechner, 1860]. Эмпирические исследования восприятия красоты включают работы с использованием других (кроме эксперимента) методов сбора психологических данных (опросники, наблюдения за поведением), а также методов нейронаук (нейроэстетика). Поскольку, как мы видели, не существует объективных методов определения красоты, во всех этих исследованиях в качестве критерия красоты принимается субъективное суждение о том, в какой степени нравится тот или иной объект.

Экспериментальные данные показывают, что произведения искусства, смысл которых легче понять, в среднем чаще нравятся и оцениваются как красивые. В частности, фигуративное искусство, связь которого с реальностью более очевидна, нравится чаще, чем произведения абстрактной живописи. Авторы теории регуляции страха (TPC, terror management theory) объясняют это необходимостью найти смысл в окружающем мире, чтобы снизить страх неизбежной и непредсказуемой смерти [Landau et al., 2006].

Согласно TPC, эволюция наградила человека, как и всех остальных животных, инстинктом самосохранения, однако человек – единственное существо, понимающее неизбежность смерти [Solomon et al., 2015]. Этот психологический конфликт порождает чувство страха, с которым люди пытаются справиться либо посредством бегства от реальности, либо с помощью убеждений, противопоставляющих биологической реальности неизбежной

смерти более значимые ценности. Самыми очевидными примерами культурных ценностей, снижающими страх смерти, являются те, которые прямо постулируют бессмертие (например, вера в загробную жизнь, религия) [Jonas, Fischer, 2006].

Однако, согласно ТРС, многие другие культурные ценности, в том числе и, казалось бы, не связанные со смертью, также обеспечивают символическое бессмертие. Например, национальная идентичность, потомство, человеческое превосходство над животными. В такой интерпретации символическое бессмертие обретается либо потому, что человек представляется частью чего-то большего, что в конечном итоге его переживет (например, страна, вид), или потому, что духовные ценности выше, чем биологические. Поскольку эти ценности определяют значимое, они также становятся основой самооценки, которая описывается как личная, субъективная мера того, насколько хорошо человек соответствует культурным идеалам [Solomon et al., 2015].

Теория регуляции страха является развитием идей, изложенных Эрнестом Беккером в его книге «Отрицание смерти», в которой автор утверждает, что большинство человеческих действий предпринимается для того, чтобы игнорировать или подавить мысль о неизбежности смерти. Ужас абсолютного исчезновения, вызывающий у людей такое глубокое – хотя и подсознательное – беспокойство, побуждает их всю жизнь пытаться избавиться от него. В масштабах общества создаются символы – законы, религиозные системы, культуры и системы убеждений, чтобы объяснить значение жизни и определить, что делает некоторые характеристики, навыки и таланты выдающимися. Общество награждает тех, кто служит примером определенных качеств, и наказывает или убивает тех, кто им не соответствует. На индивидуальном уровне соблюдение этих символов помогает снять стресс, связанный с реальностью смерти, а самооценка служит буфером против тревоги [Becker, 1973].

Предсказания ТРС проверяют в эксперименте, используя парадигмы «напоминания о смертности» (mortality salience). Например, участников просят написать о своей собственной смерти, или проводят эксперименты возле погребального дома или кладбища, либо демонстрируют графические изображения смерти [Burke et al., 2010]. Исследования с помощью подобных парадигм показывают, что напоминания о смертности усиливают предпочтение фи-

гуративного искусства и неприятие абстрактного [Landau et al., 2006]. По заключению авторов, современное искусство часто недолюбляют, поскольку в нем отсутствует осязаемый смысл и, следовательно, оно несовместимо с основным мотивом, постулируемым ТРС и направленным на поддержание осмысленного представления о реальности.

В другой работе авторы использовали шкалу личной потребности в структуре, которая позволяет выявить временное увеличение потребности людей в осмысленности. Показано увеличение оценок по этой шкале после просмотра образцов абсурдистского искусства, по сравнению с образцами реалистического искусства [Proulx et al., 2010]. Согласно полученным данным, красота служит мерилем смысла при восприятии окружающего мира. На это же указывают и данные экспериментов по изучению влияния «сложности» произведений абстрактного искусства на оценку их красоты.

До определенного предела увеличение сложности (оцененной по количеству компонентов и связей между ними) связано с увеличением оценок красоты, но, когда объект становится слишком сложным, эти оценки падают [Nicki, Moss, 1975]. Как можно предположить, сложность в данном случае воспринимается как мера похожести на объекты реального мира, которые достаточно сложны, но, если они воспринимаются как красивые, доступны для понимания и осмысления. Изучение влияния симметрии и уравновешенности композиции на восприятие красоты показывает, что объекты более симметричные и сбалансированные воспринимаются как более красивые [Lindell, Mueller, 2011].

В 1970-х гг. появляются теории, связывающие восприятие красоты с теорией информации [Moles, 1973]. В 1990-х гг. Юрген Шмидхубер предлагает алгоритмическую теорию красоты, которая принимает во внимание субъективность наблюдателя и постулирует, что среди нескольких наблюдений, классифицированных наблюдателем как сопоставимые, наиболее эстетически приятным является то, которое имеет кратчайшее описание, учитывая предыдущие знания наблюдателя и его конкретный метод кодирования данных [Schmidhuber, 1997]. Это тесно сопряжено с принципами алгоритмической теории информации и минимальной длины описания. Один из его примеров – математическая красота: математикам нравятся простые доказательства с кратким описанием на языке математики. Другой пример – эстетически приятное челове-

ческое лицо, пропорции которого исследовали в XV в. Леонардо да Винчи и Альбрехт Дюрер. Его можно описать с помощью очень небольшого количества информации [Schmidhuber, 2007].

Теория Шмидхубера делает различие между красотой и интересностью, утверждая, что интересность соответствует первой производной субъективно воспринимаемой красоты. Здесь предпосылка состоит в том, что любой наблюдатель постоянно пытается улучшить предсказуемость и лаконичность описания, обнаруживая такие закономерности, как повторения, симметрии и фрактальное самоподобие. Каждый раз, когда процесс обучения (который может быть смоделирован искусственной нейронной сетью) приводит к улучшению сжатия данных (когда последовательность наблюдений может быть описана меньшим количеством бит, чем раньше), возникающее чувство интересности данных, сопровождаемое чувством удовольствия, соответствует количеству сэкономленных бит информации [Schmidhuber, 2006].

Похожие подходы к эстетике и красоте развиваются в компьютерных науках, где методы информатики используются для прогнозирования, передачи и вызова эмоциональной реакции на произведение искусства. В этой области считается, что эстетика не зависит от вкуса, а является вопросом познания и, следовательно, обучения [Jahanian, 2016]. В 1928 г. математик Джордж Дэвид Биркгоф предложил определять эстетическую ценность отношением порядка к сложности $M = O/C$ [Akiba, 2013]. Примерно с 2005 г. компьютерные ученые пытаются разработать автоматизированные методы определения эстетического качества изображений. Как правило, используются алгоритмы машинного обучения, когда большое количество фотографий, оцененных экспертами вручную, используется для «обучения» компьютера тому, какие визуальные свойства имеют отношение к эстетическому качеству. В области музыки была использована математическая формулировка эстетической ценности в терминах «избыточности» и «сложности» и понятие скорости передачи информации для предсказания музыкального восприятия [Dubnov, 2010].

Связь чувства красоты с осмысленностью, симметричностью и сбалансированностью объекта, а также попытки алгоритмизировать красоту соотношением сложности объекта и лаконичности его описания согласуются с представлением о том, что красота – это неосознанное восприятие гармонии и простоты, «просвечивающей» сквозь избыточность и вариативность реальных объектов.

Восприятие красоты позволяет не замечать несущественное и множественное и видеть подлежащую суть и единство. Это особенно видно в абстрактной красоте математических формул. Красота в данном случае пропорциональна краткости математического описания и широте его применимости к большому спектру часто, казалось бы, совершенно не связанных друг с другом явлений. Отсюда стремление физиков создать «теорию всего», которая описывала бы (в идеале одним уравнением) все мироздание.

Но, может быть, это лишь иллюзия, основанная на том, что экономное представление информации нам удобно (например, экономит ресурсы памяти) и сжатие информации субъективно воспринимается как что-то полезное и приятное? Эта приятность и есть чувство красоты [Schmidhuber, 2006]. Например, была предложена теория, согласно которой красота объекта тождественна удовольствию, связанному с легкостью его восприятия (processing fluency theory of aesthetic pleasure). Особенности, способствующие быстрой обработке стимулов, включают их объективные характеристики, такие как совершенство формы, симметрия, контрастность фигуры и фона, а также опыт взаимодействия со стимулом, например, многократное воздействие или прототипичность. Беглость обработки сама по себе гедонически значима (т. е. она обладает неотъемлемым аффективным качеством), а высокая беглость субъективно воспринимается как положительный фактор. Беглость обработки влияет на суждения эстетической оценки, потому что люди опираются на свой субъективный опыт при вынесении оценочных суждений [Reber et al., 2004]. Однако не все объекты, которые легко «обрабатывать», воспринимаются как красивые. Нам легче воспринимать хорошо известное, чем что-то совершенно новое, но далеко не всегда известное воспринимается как красивое.

Альтернативная интерпретация связывает чувство красоты с интуитивным восприятием истины. Утверждается, что красота и истина почти синонимичны [Stewart, 2007], это отражено в стихотворении Джона Китса: «Красота есть истина, истина – красота». Как известно, и в математике, и в физике наиболее «красивые» уравнения оказываются, как правило, и наиболее верными, и люди используют красоту как показатель истины в математических задачах [Reber et al., 2008]. Многие физики и математики настаивают на том, что наиболее верные и фундаментальные законы не изобретаются, а открываются как некая неизвестная, но независи-

мо существующая территория, и это открытие сопровождается эстетическим чувством красоты [Penrose, 1989]. Все вышеизложенное хорошо укладывается в концепцию идеалистов, согласно которой красота – это восприятие латентного идеала за видимостью реального объекта.

8.4.3. *Нейроэстетика*

Термин «нейроэстетика», введенный Семиром Зеки в 1999 г. [Zeki, 1999], получил свое формальное определение в 2002 г. как научное исследование нейронных основ созерцания и создания произведений искусства [Nalbantian, 2008]. В нейроэстетике можно выделить описательный и экспериментальный подходы. В рамках описательного подхода гипотезы формулируются, исходя из эволюционных представлений о возникновении человека и сущности когнитивных процессов, предположительно участвующих в восприятии красоты. В экспериментальных исследованиях преимущественно используют нейровизуализацию для выявления активности мозга, сопровождающей восприятие красоты.

Предполагается, что эстетические переживания – это эмерджентные состояния, возникающие в результате взаимодействий между триадой нейронных систем, которые включают в себя сенсорно-моторные, эмоционально-оценочные и когнитивные системы [Chatterjee, 2013]. Например, рассматривание картин Ван Гога вызывает субъективное ощущение движения и активирует области зрительного анализатора (V5/MT), связанные с восприятием движения [Thakral et al., 2012]. Портреты активируют лицевую область в веретенообразной извилине, а пейзажи – область в парагиппокампальной извилине, связанную с восприятием мест [Yue et al., 2007]. Считается, что сенсорные области могут не только классифицировать визуальные элементы, но и участвовать в их оценке [Chatterjee et al., 2009], однако вопрос о том, в какой степени и какого рода оценка имеет место в сенсорной коре, является предметом дебатов.

Картины, изображающие действия, могут активировать моторные области коры, особенно расширенную систему зеркальных нейронов, что интерпретируется как свидетельство эмпатического ответа на произведения искусства [Freedberg, Gallese, 2007]. Удовольствие, которое люди получают от созерцания красивых объектов, автоматически активирует систему вознаграждения [Kühn, Gallinat, 2012]. Например, привлекательные лица активиру-

ют лицевую область в веретенообразной извилине и области вентрального полосатого тела [Kim et al., 2007], даже когда люди в явном виде не думают о привлекательности этих лиц.

Орбито- и медиофронтальная кора, вентральное полосатое тело, передняя поясная извилина и островок реагируют на красивые визуальные образы [Kawabata, Zeki, 2004; Jacobs et al., 2012], а медиальная орбитофронтальная кора и прилегающая к ней поясная кора реагируют на различные источники удовольствия, включая музыку [Ishizu, Zeki, 2011] и даже архитектурные пространства [Vartanian et al., 2013]. Кирк с соавторами [Kirk et al., 2009] исследовали влияние ожиданий на нейронные реакции. Люди оценивали абстрактные «художественные» изображения как более привлекательные, если было указано, что они доставлены из музея, а не созданы с помощью компьютера. Это сопровождалось большей активностью в медиальной орбитофронтальной и вентромедиальной префронтальной коре, а также в энторинальной коре, связанной с памятью. Предполагается, что ожидания людей основываются на воспоминаниях, которые усиливают (или уменьшают) визуальное удовольствие.

Лейси с коллегами [Lacey et al., 2011] обнаружили, что вентральная часть полосатого тела и области орбитофронтальной коры более чувствительны к «художественному статусу», чем к фактическому содержанию зрительных образов. Хуанг и его коллеги [Huang et al., 2011] наблюдали разные нейронные реакции у людей, которым говорили, что они смотрят на подлинный или скопированный портрет Рембранта. Подлинные портреты вызывали реакции в орбитофронтальной коре, а копии в лобно-полярной коре и правом предклинье. Смысл этих исследований состоит в том, что контекст и знания, выходящие за рамки сенсорных качеств визуальных образов, явно влияют на нейронную активность людей в эстетических переживаниях.

По мнению Семира Зеки, одного из пионеров нейроэстетики, в основе эстетического восприятия лежат свойства мозга, сформировавшиеся в процессе эволюции. Среди таких свойств выделяется способность «видеть» прототипичный объект за вариацией конкретного сенсорного восприятия. Например, цвет банана всегда распознается как желтый, несмотря на вариацию освещения и прочих факторов. Это свойство, которое в эволюции имело адаптивное значение, лежит в основе способности художников к дистилляции объекта, чтобы представить его таким, какой он есть на

самом деле, а не таким, как его видят глаза, – в конечном счете это объясняет и идеальные платонические формы, и концепты Гегеля [Zeki, 1999]. Можно думать, что способность «видеть» прототипичный объект за вариацией формы есть и у животных. Например, собаки с легкостью отличают собак от кошек, несмотря на разнообразие форм, однако это не означает, что они имеют представление о платонических формах.

Другое свойство, важное для эстетического восприятия по теории Зеки, – способность к абстракции, которая могла появиться в эволюции как необходимость из-за ограничений памяти и благодаря которой общее представление может быть применено ко многим частностям [Zeki, 2001]. Оба этих постулата теории Зеки очень схематичны. Даже если способность к абстракции появилась в силу необходимости из-за ограничений памяти, это ничего не говорит о том, как данная способность осуществляется в мозге. Как мозг распознает прототипичный объект за вариацией конкретного сенсорного восприятия и как он извлекает общие черты по видимости разнородных объектов – остается тайной.

Вилайанур Рамачандран с коллегами разработали в высшей степени спекулятивную теорию человеческого художественного опыта и опосредующих его нервных механизмов [Ramachandran, Hirstein, 1999]. Они описывают ряд особенностей сенсорного (в основном зрительного) восприятия, которые возникли, по их предположению, в эволюции для более эффективного обнаружения хищников и других адаптивных задач, и которые могут опосредовать определенные особенности художественного восприятия.

Сравнение активации мозга при восприятии произведений живописи, которые кажутся красивыми, и теми, которые не воспринимаются как красивые, выявляет большую активацию в медиальной орбитофронтальной коре в первом случае [Kawabata, Zeki, 2004]. По результатам других исследований, эта же область активна при восприятии красоты разных источников [Ishizu, Zeki, 2011], включая музыку [Blood, Zatorre, 2001], нравственную [Tsukiura, Cabeza, 2011] и даже математическую [Zeki et al., 2014] красоту. Интересно, что переживание возвышенного, в отличие от прекрасного, приводит к другому паттерну мозговой активности [Ishizu, Zeki, 2014].

Согласно существующим данным, связывающим ОФК с воспринимаемой гедонистической значимостью во вкусовых, обонятельных и визуальных модальностях, он является общим центром

для оценки гедонистической составляющей стимула независимо от его происхождения [Kirk et al., 2009]. Его активация при восприятии красоты лишь показывает, что оно связано с активацией системы вознаграждения. Кроме ОФК, восприятие красоты связано с активацией дорсолатеральной префронтальной коры, которая участвует в широком спектре когнитивных процессов, таких как внимание, рабочая память и нисходящий контроль [Cela-Conde et al., 2004]. Эксперименты, разработанные специально для того, чтобы заставить испытуемых рассматривать произведение искусства субъективно (оценивая его эстетическую привлекательность), выявили более высокую активацию эмоциональных областей мозга, в частности, островка [Cupchik et al., 2009]. Однако природа связи между степенью активации островка и положительными или отрицательными эмоциями в этом контексте неизвестна.

В целом критики отмечают рассогласование между узким подходом исследователей к моделированию восприятия красоты и теми грандиозными заявлениями, которые они выдвигают в своих теориях [Freeman, 1999]. Очевидно также рассогласование результатов разных работ. Утверждается, что, по крайней мере в сфере изобразительного искусства, эти, кажущиеся разными, результаты совместимы с общей моделью, связывающей эстетическое восприятие с активностью зрительного анализатора и аффективной системы, а также систем вознаграждения и принятия решений. Привлекательные визуальные свойства картин повышают активность вентральной зрительной коры. Эстетические суждения активируют части дорсолатеральной и медиальной префронтальной коры. Эмоциональные реакции на эти стимулы активируют ОФК и переднюю поясную кору [Chatterjee, 2011]. Это описание не дает возможности понять, в чем же специфика активности мозга при восприятии красоты.

Практически неотличимый набор активаций описан во многих работах при предъявлении эмоционально значимых стимулов. Важно также отметить, что все результаты получены в экспериментах, в которых испытуемым предъявляли ряд изображений и просили рассортировать их на красивые и некрасивые (либо поставить оценку красоты по пятибалльной шкале). В данном случае интроспективное оценивание своих чувств и принятие решений отличаются от простого восприятия красоты. Неизвестно поэтому, в какой степени области мозга, связанные с вниманием и принятием решений, активация которых выявлена в подобных

исследованиях, действительно участвуют в восприятии красоты. В итоге преимущественно остается активация ОФК. Как мы отмечали, она отнюдь не специфична для восприятия красоты и в еще большей степени наблюдается, например, при восприятии приятных запахов, вкусной пищи и сексуальных стимулов.

8.5. Религия

Философия религии – это третий раздел аксиологии, изучающей природу ценностных ориентаций, и мы не можем обойти данную тему, поскольку в настоящее время, когда для практически любого раздела наук о человеке можно найти термин с приставкой «нейро», философия религии не является исключением [Burton, 2008]. На сегодня нет единого мнения о точном определении необходимых и достаточных условий того, что считается религией. Словарь философии религии дает следующее определение: «Религия включает в себя коллективный, передаваемый свод учений и предписанных практик о высшей, священной реальности, вызывающей благоговение и трепет, которые ведут своих приверженцев к тому, что описывается как спасительные, просветляющие или освобождающие отношения с этой реальностью через преобразующие жизнь молитвы, ритуальную медитацию и/или нравственные практики, такие как покаяние и личное возрождение» [Taliaferro, 2019].

Если считать, что те традиции, которые сегодня широко признаны как религии, действительно таковыми являются, то они включают (как минимум) индуизм, буддизм, даосизм, конфуцианство, иудаизм, христианство и ислам. Греко-римские взгляды на богов, ритуалы, загробную жизнь, душу в широком смысле также являются «религиозными» или «религиозно значимыми». Это определение дает основания считать, что сайентология является проторелигией, поскольку она не имеет устойчивого коллективного, передаваемого свода учений, но отвечает другим условиям, чтобы считаться религией [Taliaferro, 2019].

Можно было бы думать, что в наше время широкого распространения атеизма религия не имеет такого значения, какое ей придавали, скажем, несколько сотен лет назад, однако, согласно данным социологических исследований, и сегодня большая часть населения мира исповедует религию или находится под ее влиянием. Здесь, безусловно, не место обсуждать разные аспекты фи-

лософии религии, которая имеет такую же длинную историю, как и сама философия. Можно лишь отметить, что некогда популярная точка зрения позитивистов, считающих, что вопросы религии не имеют смысла, поскольку они не имеют эмпирического наполнения, с середины XX в. была практически вытеснена позицией, основанной на предложении Витгенштейна. Суть его позиции заключается в следующем. Спрашивать, существует ли Бог – не значит задавать теоретический вопрос. Если это вообще что-то значит, то вопрос здесь о поклонении и молитве; вопрос о том, есть ли что-нибудь во всем этом [Phillips, 1976].

Значительное число философов (середины XX в. и далее) пришли к выводу, что все традиционные аргументы и контраргументы по поводу метафизических вопросов религии неопределенны. Если это так, то новая философия религии, вдохновленная Витгенштейном, имеет свое преимущество – она смещает почву в более определенную область, касающуюся реальной практики самих религиозных людей. Унаследовав акцент Витгенштейна на практике, некоторые философы, работающие над этой темой сегодня, уделяют больше внимания значению религии в жизни, а не рассматривают религиозные убеждения на предмет их верности или неверности [Cottingham, 2014].

Для чего нужна религия, какую функцию она выполняет в жизни общества и отдельного человека? Уже в начале XVIII в. натуральные философы пытались объяснить существование религии невежеством: люди часто используют сверхъестественные объяснения, когда им не хватает понимания естественных причин, лежащих в основе экстраординарных событий, – чем более человек невежественен или имеет меньше опыта, тем больше он видит чудес [de Fontenelle, 1824]. «Естественная история религии» Юма является наиболее известным философским примером естественно-исторического объяснения религиозной веры. Он выводит истоки политеизма, который, по его мнению, был самой ранней формой религиозной веры, из незнания естественных причин в сочетании со страхом и опасениями по поводу окружающей среды. По Юму, обожествляя аспекты окружающей среды, первые люди пытались убедить или подкупить богов, тем самым обретая чувство контроля [Hume, 1757/2007].

В XIX – начале XX в. представители вновь возникших научных дисциплин, таких как антропология, социология и психология, исследовали предполагаемые натуралистические корни рели-

гиозных убеждений. По одной из теорий все общества в своих попытках осмыслить мир проходят три стадии развития: теологическую – самую раннюю, где преобладают религиозные объяснения; метафизическую (не вмещающий Бог); и, наконец, позитивную, или научную стадию, характеризующуюся научными объяснениями и эмпирическими наблюдениями [Comte, 1841]. Следуя этим представлениям, Дюркгейм и Фрейд, а также социальные теоретики Карл Маркс и Макс Вебер предложили разные версии тезиса секуляризации, согласно которому религия будет исчезать перед лицом современных технологий, науки и культуры.

Однако исследования, проведенные в Соединенных Штатах в конце 1930-х – 1960-е гг., показали, что, вопреки тезису секуляризации, религия не только не исчезала, а, похоже, претерпевала существенное возрождение [Stark, 1999]. Как показывают социологические исследования недавних лет, почти девять из десяти взрослых в США верят в Бога или универсальный дух. Среди молодых людей процент теистов составляет около 80 % [Masci, Smith, 2016]. Атеизм и агностицизм гораздо шире распространены среди ученых, особенно среди тех, кто работает в элитных учреждениях США. По результатам опроса членов Национальной академии наук США, большинство не верили в существование Бога (72.2 %), 20.8 % были агностиками и только 7 % теистами [Larson, Witham 1998]. Экклунд и Шейтле [2007] проанализировали ответы ученых, работающих в области социальных и естественных наук, из 21 элитного университета США. Около 31.2 % их участников идентифицировали себя как атеисты, а еще 31 % – как агностики. Остальные верили в высшие силы (7 %), верили в Бога с некоторыми сомнениями (20.9 %), или верили в Бога без всяких сомнений (9.7 %). В отличие от населения в целом, ученые старшего возраста в этой выборке не проявляли более высокой религиозности – на самом деле, они с большей вероятностью говорили, что не верят в Бога [Ecklund, Scheitle, 2007; Ecklund, 2010].

Гросс и Симмонс [Gross, Simmons, 2009] исследовали более разнородную выборку ученых из американских колледжей, включая общественные колледжи, элитные институты, присуждающие докторскую степень, не элитные четырехлетние государственные школы и небольшие гуманитарные колледжи. Они обнаружили, что большинство университетских профессоров имели теистические убеждения, либо веря в Бога (34.9 %), либо веря в Бога с некоторыми сомнениями (20.9 %), либо веря в высшие силы (19.2 %).

На веру в Бога влиял как тип учреждения (более низкая вера в более престижных университетах), так и дисциплина (более низкая вера в физических и биологических науках по сравнению с социальными и гуманитарными).

Результаты показывают, что ученые более разнообразны в религиозном отношении, чем это обычно предполагалось, и что большинство из них не против религии. Тем не менее в США процент атеистов и агностиков в академических кругах выше, чем среди населения в целом. Одной из причин может быть предубеждение против верующих в академических кругах. Когда социологи спрашивали работодателей престижных университетов, наняли бы они кого-нибудь, если бы знали о принадлежности кандидата к христианству, 39.1 % заявили, что они с меньшей вероятностью наняли бы этого кандидата – аналогичные результаты были получены с другими религиозными группами, такими как мормоны или мусульмане [Yancey, 2012]. Другая причина может заключаться в том, что верующие усваивают преобладающие в обществе негативные социальные стереотипы и в результате теряют интерес к научной карьере [Rios et al., 2015]. Согласно полученным данным, существование религии и, в более широком смысле, веры в сверхъестественное, вряд ли можно объяснить лишь невежеством.

Начиная с 1920-х гг. научное изучение религии сосредоточилось на конкретных религиозных традициях и верованиях. Антропологи, не полагаясь исключительно на отчеты из вторых рук (обычно плохого качества и из искаженных источников), занимались серьезными полевыми исследованиями и пришли к заключению: религиозные убеждения не являются исключительно результатом незнания естественных механизмов [Malinowski, 1992]. Как было совсем недавно обнаружено, люди в разных культурах с легкостью комбинируют сверхъестественные и естественные объяснения [Legare et al., 2012]. Психологи и социологи религии также начали сомневаться в том, что религиозные убеждения уходят корнями в иррациональность, психопатологию и другие атипичные психологические состояния, как предполагали Джеймс [James, 1902] и другие ранние психологи.

Недавним развитием в научном изучении религии считается появление когнитивной науки о религии. В этой мультидисциплинарной области работают психологи развития, антропологи, философы и когнитивные психологи. В отличие от других научных подходов к религии, когнитивная наука исходит из предположе-

ния, что религия не является чисто культурным феноменом, а зависит от универсальных когнитивных процессов человека [Barrett, 2004]. Например, согласно Полу Блуму [Bloom, 2007], религия возникает как побочный продукт нашего интуитивного различения между разумом и телом: мы можем думать о сознании как о продолжающемся даже после смерти тела (например, приписывая желания умершему члену семьи). Это делает веру в загробную жизнь и в бестелесных духов естественной и спонтанной.

Дебора Келемен с сотрудниками провела серию экспериментальных работ с детьми и взрослыми с целью выявить спонтанные интуиции в отношении восприятия целеполагания в неживой природе и возможности существования ментальности вне материального тела. По результатам экспериментов, дети дошкольного возраста имеют склонность видеть вещи в мире природы как целенаправленно созданные. Например, на вопрос, согласны ли они с тем, что дождь просто идет из облака, а не «создан для чего-то», дошкольники, в отличие от взрослых, придерживаются мнения, что дождь (и другие физические явления) «созданы для чего-то», и именно поэтому они существуют [Kelemen, 1999]. Эти телеологические интуиции сохраняются и в начальной школе, особенно в отношении свойств объекта. Например, когда детей спрашивают, являются ли скальные породы острыми из-за физического процесса или потому, что они выполняют определенные функции, они, в отличие от взрослых, предпочитают телеологические объяснения (например, это нужно, чтобы животные не могли на них сесть и сломать, или чтобы животные могли о них почесаться, когда у них появился зуд).

Склонность к телеологическому объяснению сохраняется даже когда детям говорят, что взрослые применяют физические объяснения к неживым физическим объектам [Kelemen, 2003]. Она начинает снижаться в 9–10 лет, что позволяет отвергнуть интерпретации, согласно которым телеологические интуиции – это отражение выраженной культурной религиозности. В последующих работах Келемен изучала вопрос, сохраняется ли телеологическая интуиция во взрослой жизни.

Кэслер и Келемен [Casler, Kelemen, 2007] провели исследование с участием взрослых цыган из Америки и Румынии и американских школьников. Как оказалось, необразованные взрослые цыгане (из Румынии) более чем в два раза чаще выбирали телеологические объяснения, чем образованные взрослые цыгане (из

Америки), и в этом отношении больше походили на американских школьников, чем на образованных взрослых. Эти результаты показывают, что тенденция к распространению телеологического рассуждения на неживые объекты может повторяться в разных культурах и что она не просто исчезает с возрастом, а должна быть вытеснена образованием. Далее Келемен и Россет [Kelemen, Rosset, 2009] предоставили экспериментальные доказательства того, что в условиях высокой когнитивной нагрузки даже взрослые с университетским образованием проявляют признаки «научно неподобающего» телеологического объяснения.

Еще в одной работе ученые-физики из ведущих американских университетов оценивали телеологические объяснения как истинные или ложные, либо быстро, либо без ограничения по времени. Подобно студентам и неученым участникам того же возраста, ученые продемонстрировали склонность принимать «ненаучные» телеологические объяснения в условиях скорости, хотя отвергали их, когда время ответа не было ограничено [Kelemen et al., 2013]. Согласно этим данным, одна из возможных причин широко распространенных в разных культурах верований в божества, которые либо упорядочивают, либо создают мир природы, заключается в переключке таких идей с рано проявляющимся и стойким интуитивным убеждением в том, что мир природы преднамеренно создан.

Аналогичные результаты получены в отношении другой «ненаучной» идеи – веры в возможность существования ментальных состояний после смерти или до рождения. Так, Беринг и Бьорк-лунд [Bering, Bjorklund, 2004] обнаружили, что, когда детям от 3 до 12 лет рассказывали о мертвой мыши, а затем спрашивали, сохранились ли у нее ментальные (мысли, эмоции, намерения) и телесные (например, голод, жажда) состояния, дети всех возрастных групп склонны были приписывать мертвой мыши больше ментальных, чем телесных способностей, но у старших детей эта тенденция снижалась, что опять свидетельствует о ее врожденном, а не выученном характере. Как показывают доступные данные, с раннего возраста люди интуитивно склонны считать, что, несмотря на драматические биологические изменения, связанные со смертью, ментальные (но не телесные) аспекты продолжают существовать после смерти, когда физическое тело разрушено. Более того, аналогичные интуиции показаны у детей и в отношении существования ментальности до рождения [Emmons, Kelemen,

2014; Kelemen et al., 2021]. В целом можно сделать такое заключение: возникновение религии и веры в сверхъестественное не является следствием невежества, а имеет глубокие корни в природе человека.

Было бы странным, если бы для объяснения существования этих корней не были бы привлечены эволюционные идеи. Религиозное поведение часто сопряжено со значительными издержками, включая экономические расходы, опасные ритуалы или затраты времени, которое можно было бы использовать иным образом. Это предполагает, что естественный отбор должен действовать против религиозного поведения, если только оно не дает значительные преимущества для выживания. В эволюционной психологии религии есть два основных направления. Согласно одному из них, сама религия возникла в результате естественного отбора и является адаптацией, т. е. она давала какое-то эволюционное преимущество. Другой подход заключается в том, что религиозные убеждения и поведение возникли как побочные продукты других адаптивных черт, т. е. они не были изначально объектом естественного отбора из-за их собственных преимуществ [Dávid-Barrett, Carney, 2015].

Одна из групп гипотез в рамках первого направления рассматривает религию как биологическую или культурную адаптацию, которая способствует сотрудничеству и сплоченности внутри групп и помогает людям решать совместные проблемы, увеличивая тем самым шансы на выживание и воспроизводство [Sosis, Alcorta, 2003; Bering, 2011; Norenzayan, 2013]. Другая группа гипотез предполагает, что религия может оказывать непосредственное положительное влияние на здоровье и выживание, так как делает жизнь более осмысленной. Эмпирические данные действительно показывают прямую положительную связь между религиозной практикой – с одной стороны, и здоровьем и долголетием – с другой; в современном обществе [Koenig, Cohen, 2001].

В рамках второго направления Стивен Гулд приводит религию как пример экзаптации, но не указывает определенную черту, на которую, по его мнению, действительно повлиял естественный отбор. Он цитирует предположение Фрейда о том, что наш большой мозг, получивший развитие по другим причинам, привел к появлению сознания, которое заставило людей столкнуться с концепцией личной смертности, послужившей стимулом для возникновения религии [Gould, 1991]. Другие предложили более опреде-

ленные психологические процессы – те, что отбирались в эволюции и могли способствовать появлению религии. Они могут включать способность делать выводы о присутствии вредных организмов, способность придумывать причинно-следственные связи для природных явлений и способность распознавать, что у других людей есть собственные мысли, желания и намерения (теория разума). Эти три адаптации (среди прочего) позволяют людям вообразить целенаправленных агентов за многими наблюдениями, которые нельзя было бы легко объяснить иначе – например, гром, молния, движение планет и т. д. [Atran, Norenzayan, 2004].

Джастин Барретт считает, что вера в Бога естественна, поскольку зависит от умственных способностей людей. По его предположению, одним из фундаментальных ментальных модулей мозга является «устройство обнаружения гиперактивного агента» (УОГА), еще одна потенциальная система для определения опасности. Система может способствовать выживанию, даже если она сверхчувствительна: лучше избегать воображаемого хищника, чем быть убитым настоящим. Это может поощрять веру в призраков и духов [Barrett, 2004]. Согласно ТРС, религия выполняла адаптивную функцию, так как обещание загробной жизни или продолжения существования помогало преодолеть страх неизбежной смерти [Landau et al., 2007]. По теории мема, предложенной Ричардом Докинзом, религии возникают в процессе эволюции культуры и распространяются наподобие «психических вирусов» [Dawkins, 2006].

Естественнонаучный подход к ментальным явлениям, связывающий их прежде всего с активностью мозга, не обошел стороной и религию. Олдос Хаксли впервые использовал термин «нейротеология» в утопическом романе «Остров». Сейчас использование этого термина в научных работах уже широко распространено. Дисциплина изучает когнитивную нейробиологию религиозного опыта и духовности. Доминирующий подход – запись активности мозга (преимущественно с помощью фМРТ), сопровождающей измененные состояния сознания во время молитвы, медитации и других видов духовной практики. Эндрю Ньюберг и соавторы обнаружили, что сфокусированное духовное созерцание вызывает изменение активности мозга, приводящее к восприятию трансцендентного религиозного опыта как осязаемой реальности [Newberg et al., 2002].

В одной из работ фМРТ записывали у монахинь-кармелиток, которых просили вспомнить и пережить заново (с закрытыми глазами) самый сильный мистический опыт, который они когда-либо испытывали в своей жизни в качестве члена Ордена кармелитов. Активация была выявлена в множестве областей мозга, включая правую медиальную орбитофронтальную и среднюю височную кору, нижнюю и верхнюю теменные доли, хвостатые ядра, левую медиальную префронтальную и переднюю поясную кору, левую нижнюю теменную дольку, левую островковую область и ствол мозга [Beauregard, Paquette, 2006]. Авторы заключают, что в мозге нет одного «центра Бога».

Фергюсон с соавторами, используя данные фМРТ, полученные от 19 набожных мормонов, показали, что узнаваемое чувство, центральное в их религиозной практике, связано с активацией в прилежащем ядре, вентромедиальной префронтальной коре и лобных областях, связанных с регуляцией внимания. Активация прилежащего ядра предшествовала пику духовных переживаний на 1–3 секунды и была воспроизведена в четырех отдельных задачах. По предположению авторов, активация системы вознаграждения мозга (прилежащее ядро и вентромедиальная префронтальная кора) свидетельствует о том, что доктринальные концепции могут быть вознаграждением и способны мотивировать поведение религиозных людей [Ferguson et al., 2017]. Та же группа авторов недавно опубликовала исследование, в котором анализировались эффекты повреждений мозга на спиритуальный опыт. Было обнаружено, что повреждения в области околосреднего серого вещества, которое ранее связывали с формированием страха, модуляцией боли и альтруистическим поведением, связано с самооценкой религиозности [Ferguson et al., 2021].

Исследование психологических аспектов молитвы с использованием факторного анализа выявило три фактора. Первым было осознание себя, вторым и третьим – стремление вверх (к Богу) и движение вовне (к другим). Данное исследование поддерживает современную модель молитвы как коммуникации (будь то с собой, высшим существом или другими) [Ladd, Spilka, 2002]. В соответствии с этим результаты фМРТ-исследований показывают совпадение между молитвой и разговором с любимым человеком в областях мозга, связанных с теорией разума, предполагая, что на уровне мозга и то и другое представлено как межличностные от-

ношения. Эти области мозга также связаны с дефолтной сетью, участвующей в процессах самосознания [Neubauer, 2014]. Показано также снижение активности фронто-париетальной системы внимания к внешним стимулам во время молитвы, одновременно со снижением чувствительности к боли [Elmholdt et al., 2017].

Большое количество исследований посвящено выявлению активности мозга во время медитации. Хотя техники медитации изначально возникли в рамках религиозных учений, в наше время они широко используются в качестве психотерапии в отрыве от религии, поэтому мы не будем подробно останавливаться на этих работах. Метаанализ фМРТ-исследований медитации показывает, что она сопровождается активацией областей мозга, участвующих в саморегуляции, целенаправленном решении проблем, адаптивном поведении и interoцепции. Практика медитации вызывает функциональные и структурные изменения мозга у опытных медитаторов, особенно в областях, связанных с процессами самореференции и самосознания (дефолтная сеть) и саморегуляции [Vossia et al., 2015]. Паттерны активации мозга различны для различных техник медитации, но некоторые области, включая островок, дополнительную моторную кору, дорсальную переднюю поясную кору и лобный полюс, активируются во всех случаях [Fox et al., 2016]. Также было высказано предположение, что стимуляция височной доли психоактивными ингредиентами «магических грибов» имитирует религиозные переживания. Эта гипотеза нашла подтверждение в отношении псилоцибина [Griffiths et al., 2008].

Подводя итог, можно заключить, что, по современным представлениям, религия и вера в сверхъестественное не являются результатом невежества, а возникают спонтанно из, вероятно, врожденных интуиций, согласно которым мир природы преднамеренно создан, а ментальные состояния могут существовать после смерти или до рождения. Эти интуиции могут быть подавлены образованием, но остаются на подсознательном уровне даже у людей, считающих себя атеистами. Сформулированные гипотезы об эволюционных корнях этих интуиций являются, естественно, чисто умозрительными.

Исследования мозговых коррелятов молитвы, медитации и других видов спиритуального опыта дает разнородную картину, в которой непросто выделить элементы единства. Одним таким элементом, так же, как и при изучении мозговых коррелятов эстети-

ческого восприятия, является активация центров системы вознаграждения – прилежащее ядро и ОФК. Выявлена также активация эмоциональных центров (островков) и центров так называемого социального мозга. В отношении областей мозга, связанных с процессами самореференции и самосознания (дефолтная сеть), а также функционально противоположных им центров вовне ориентированного внимания (фронтально-париетальная сеть, ФПС) данные неоднозначны и зависят от конкретного вида спиритуальной практики. Христианская молитва связана с доминированием ДСМ над ФПС, но некоторые виды медитации связаны с противоположным трендом. Например, медитативные практики, берущие начало в дзен-буддизме и нацеленные на снижение внимания к себе и восприятие сиюминутной («здесь и сейчас») действительности, снижают активность ДСМ и увеличивают активность ФПС [Brewer et al., 2011; Kemmer et al., 2015; Taren et al., 2017].

8.6. Итоги по ценностным ориентациям

В целом анализ систем ценностных ориентаций человека позволяет сделать некоторые выводы. Базовая, присущая любому живому существу тенденция классифицировать все явления и события на хорошие, нейтральные и плохие и направлять свое поведение на достижение хорошего и избегание плохого есть и у человека, но наполнение ее совершенно другое. Сугубо эгоистические драйвы, направленные на обеспечение собственного существования и продолжение рода, естественны и приемлемы до тех пор, пока они не вступают в конфликт с благополучием других людей или ценностями высшего порядка, но практически во всех культурах они не расцениваются как хорошие. Системы ценностных ориентаций, основанные на морали, восприятию красоты и религии, различны в разных культурах, но в основе их лежат одни и те же интуиции, происхождение которых неизвестно, несмотря на немалое количество спекуляций об их эволюционных корнях. Мораль направлена на подавление эгоистических драйвов и приоритизацию широко понимаемой социальной гармонии. Эстетика и религия отражают присущее человеку стремление к пониманию смысла и гармонии окружающего мира.

Интересно, что всем этим сугубо человеческим и по сути различным видам ментальности нет адекватного объяснения на уровне активности мозга. «Моральный мозг», «эстетический мозг»,

«теологический мозг» трудно отличить друг от друга и от разных других вариантов вроде «социального» или «культурного» мозга. Везде участвует набор одних и тех же областей, которые к тому же задействованы и в большом количестве других процессов. Например, во всех процессах, связанных с ценностными ориентациями, так или иначе участвуют центры системы вознаграждения. Они активируются при восприятии как сугубо биологических стимулов, так и эстетических или религиозных ценностей. Моральную, эстетическую или религиозную суть состояния невозможно «расшифровать» из набора сопровождающих его активаций мозга, что опять же указывает на эмерджентный характер этих состояний.

ЗАКЛЮЧЕНИЕ

Отождествление ментальности с активностью мозга стало стандартом как в научных, так и в далеких от науки сферах. В философии это направление представлено концепциями материалистического монизма, или физикализма, в рамках которого три интерпретации отношения ментальности к активности мозга включают элиминативизм, редукционизм и эмерджентизм (с вариантами слабой и сильной эмерджентности). В науке в рамках материалистического монизма развиваются теории, описывающие ментальные процессы в терминах обработки и кодирования информации в мозге и производимых мозгом вычислений. В популярной литературе и обыденной философии стандартом стало объяснять психологические явления нейрофизиологическими процессами и эволюционной теорией. Популярные статьи пестрят сообщениями о том, что ученые раскрыли, например, природу родительской любви. Оказывается, все дело в механизме продукции окситоцина в гипоталамусе, который появился в эволюции млекопитающих для запуска поведения, необходимого для охраны потомства и обеспечения репродуктивного успеха в ряду поколений.

«Твердо установленные», или «самоочевидные» истины далеко не всегда на поверку оказываются такими бесспорными. В этой книге была предпринята попытка собрать существующие научные данные о связи психических процессов с активностью мозга и на их основе оценить правдоподобность физикалистской интерпретации психических явлений. Мы прошли через разные тематические зоны, начиная с эволюции человеческого мозга и кончая данными о мозговых коррелятах религиозности. Подводя общий итог, можно с уверенностью сказать, что простая физикалистская картина, по которой человеческий мозг появился как продукт естественного отбора наиболее эффективных поведенческих паттернов и выступает самодостаточной причиной поведения, а психические явления не имеют самостоятельной каузальной силы и могут быть сведены к активности мозга, не имеет однозначного подтверждения в существующих эмпирических данных.

Палеонтологические данные не дают основания утверждать, что человеческий мозг появился в результате естественного отбора наиболее эффективных поведенческих паттернов. Появление в эво-

люции человеческого мозга до сих пор остается загадкой и может быть объяснено только чередой счастливых случайностей. Данные о характере активности мозга, сопровождающей когнитивные процессы, показывают принципиальное различие между природой нервной активности и природой ментальности. Связь активности мозга с феноменологией ментальной активности носит статистический характер, и каждая из этих двух сфер подчиняется своим законам. Исследование систем ценностных ориентаций человека показывает, что моральную, эстетическую или религиозную суть ментального состояния невозможно «расшифровать» из паттерна сопровождающих его активаций мозга. Во всех процессах участвует набор одних и тех же областей мозга, которые к тому же задействованы и в большом количестве других процессов сугубо биологического характера. Можно заключить поэтому, что на основе имеющихся в настоящее время данных редукционизм вряд ли можно рассматривать как жизнеспособную опцию.

Философы и ученые, понимающие ограниченность редукционизма, все чаще прибегают к разным вариантам эмерджентизма при описании ментальных состояний. По единодушному мнению философов, слабый эмерджентизм совместим с материализмом и конкретно он чаще всего и имеется в виду. Однако пригодность концепции слабой эмерджентности для описания ментальных состояний требует выполнения некоторых условий. Хотя в случае эмерджентизма мы не должны ожидать, что ментальное содержание будет каким-то образом закодировано в активности мозга, у нас все же есть основания ожидать некой согласованности между ними [O'Connor, 2020, p. 1]. В приложении к содержанию сознания это означает, что разные содержания сознания должны соответствовать разным паттернам активности мозга, а одно и то же содержание сознания должно сопровождаться одной и той же активностью мозга при повторных его появлениях как у одного и того же субъекта, так и у разных субъектов. Систематическая проверка выполнимости этих условий не проводилась, однако ряд уже имеющихся данных позволяет сомневаться в их выполнимости по крайней мере для некоторых видов семантики сознания. Если слабый эмерджентизм также окажется нежизнеспособным, можно будет рассматривать сильный, обладающий нисходящей каузальной силой вариант эмерджентизма [Chalmers, 2006], однако совместимость этого варианта с философией материализма вызывает сомнения [Bedau, 1997].

Надо сказать, что в рамках физикализма сама концепция эмерджентной сущности достаточно туманна. Беду [Ibid] выделяет два отличительных признака эмерджентных явлений: (1) они каким-то образом порождаются лежащими в их основе процессами и (2) каким-то образом автономны от лежащих в их основе процессов. Это двойное «каким-то образом» отражает статус эмерджентной сущности, как «вечной философской загадки», которая в лучшем случае порождает «призрак незаконного возникновения чего-либо из ничего» [Ibid]. В своей основе концепция эмерджентной сущности – просто признание необъяснимости возникновения некоего явления. Присваивание данному явлению лейбла «эмерджентная сущность» создает лишь видимость его объяснения, особенно относительно концепции сильной эмерджентности. Поэтому некоторые философы предлагают полностью от нее отказаться и при объяснении феноменологии ментального отдать предпочтение философским системам, альтернативным физикализму, таким как панпсихизм [Nagel, 1979a; Chalmers, 2013] или нейтральный монизм [Atmanspacher, 2014], которые постулируют протоментальный характер базовой структуры мира. Это, однако, уже другая история, находящаяся за пределами темы настоящей книги.

СПИСОК ЛИТЕРАТУРЫ

- Кунин Е.В.** Логика случая. О природе и происхождении биологической эволюции. М.: Центрполиграф, 2012.
- Маяковский В.В.** Полное собрание сочинений: в 13 т. М.: АН СССР; Ин-т мировой лит. им. А.М. Горького, 1955–1961. Т. 10.
- Свааб Д.** Мы – это наш мозг. СПб.: Изд-во Ивана Лимбаха, 2019.
- Abnar S., Ahmed R., Mijnheer M., Zuidema W.** Experiential, distributional and dependency-based word embeddings have complementary roles in decoding brain activity // arXiv. 2017. 1711.09285.
- Acker F.** New findings on unconscious versus conscious thought in decision making: Additional empirical data and meta-analysis // *Judgm. Decis. Mak.* 2008. V. 3. P. 292–303.
- Akiba F.** Preface: Natural computing and computational aesthetics. Natural computing and beyond // *Int. Conf. Commun. Inf. Technol.* 2013. V. 6. P. 117–118.
- Allen C., Neal J.** Teleological notions in biology / Zalta E.M. (ed.) *The Stanford encyclopedia of philosophy*. 2020. URL: <https://plato.stanford.edu/archives/spr2020/entries/teleology-biology/>.
- Allport A.** Attention and control: Have we been asking the wrong questions? A critical review of twenty five years. *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* / Meyer D.E., Kornblum S. (eds.). Cambridge MA: MIT Press, 1993. P. 183–218.
- Alvarez J.A., Emory E.** Executive function and the frontal lobes: A meta-analytic review // *Neuropsychol. Rev.* 2006. V. 16, N 1. P. 17–42.
- Andersen R.A., Andersen K.N., Hwang E.J., Hauschild M.** Optic ataxia: From Balint's syndrome to the parietal reach region // *Neuron*. 2014. V. 81. P. 967–983.
- Anderson M.L., Penner-Wilger M.** Neural reuse in the evolution and development of the brain: Evidence for developmental homology? // *Dev. Psychobiol.* 2013. V. 55. P. 42–51.
- Aquinas T.** *Summa theologiae*. Fathers of the English Dominican province. trans. L.: Christian Classics, 1981 (13th century text).
- Aristotle.** *The complete works of Aristotle: 2 v.* / Barnes J. (ed.). Princeton: Princeton Univ. Press, 1984 (4th century BCE text).
- Armstrong D.** What is consciousness? // *The Nature of Mind*. Ithaca, N.Y.: Cornell Univ. Press, 1981.
- Armstrong K.M., Moore T.** Rapid enhancement of visual cortical response discriminability by microstimulation of the frontal eye field // *PNAS*. 2007. V. 104, N 22. P. 9499–9504.
- Arnold M.B.** *Emotion and personality*. N.Y.: Columbia Univ. Press, 1960.

- Atkinson R.C., Shiffrin R.M.** Human memory: A proposed system and its control processes // Spence K.W., Spence J.T. (eds.) *The psychology of learning and motivation*. N.Y.: Academic Press, 1968. P. 89–195.
- Atmanspacher H.** 20th century variants of dual-aspect thinking // *Mind & Matter*. 2014. V. 12, N 2. P. 245–288.
- Atmanspacher H.** Quantum approaches to consciousness / Zalta E.N. (ed.) *The Stanford encyclopedia of philosophy*. 2020. URL: <https://plato.stanford.edu/archives/sum2020/entries/qt-consciousness/>.
- Atran S., Norenzayan A.** Religion's evolutionary landscape: Counterintuition, commitment, compassion, communion // *Behav. Brain Sci.* 2004. V. 27, N 6. P. 713–730.
- Attenborough D.** *Micro monsters 3D*. N.Y.: Colossus Productions, 2013.
- Augustine.** *Earlier writings* / Burleigh J.H. (ed.). N.Y.: WJK Publishing, 1953.
- Augusto L.M.** Unconscious knowledge: a survey // *Adv. Cogn. Psychol.* 2010. V. 6. P. 116–141.
- Axelrod R.** *The evolution of cooperation*. N.Y.: Basic Books, 1984.
- Ayer A.J.** *Language, truth, and logic*. N.Y.: Dover, 1952.
- Baars B.J.** *A cognitive theory of consciousness*. Cambridge, MA: Cambridge Univ. Press, 1988.
- Baars B.J.** *In the theater of consciousness*. N.Y., NY: Oxford Univ. Press, 1997.
- Baars B.J.** The conscious access hypothesis: Origins and recent evidence // *Trends Cogn. Sci.* 2002. V. 6. P. 47–52.
- Baddeley A.** Exploring the central executive // *Q. J. Exp. Psychol. Section A*. 1996. V. 49. P. 5–28.
- Baddeley A.** Working memory: theories, models, and controversies // *Ann. Rev. Psychol.* 2011. V. 63. P. 1–29.
- Baddeley A., Eysenck M.W., Anderson M.C.** *Memory*. N.Y.: Psychology Press, 2009.
- Baddeley A.D., Hitch G.J.** Working memory / Bower G.A. (ed.) *The psychology of learning and motivation*. N.Y.: Academic Press, 1974. P. 47–89.
- Baker D.H., Graf E.W.** Natural images dominate in binocular rivalry // *PNAS*. 2009. V. 106, N 13. P. 5436–5441.
- Baker L.** *Saving belief*. Princeton: Princeton Univ. Press, 1987.
- Bambini V.** *Neurolinguistics* // Östman J.O., Verschueren J. (eds.) *Handbook of pragmatics*. Amsterdam: John Benjamins Publishing Company, 2012.
- Barbey A.K., Koenigs M., Grafman J.** Dorsolateral prefrontal contributions to human working memory // *Cortex*. 2013. V. 49, N 5. P. 1195–1205.
- Barlow H.B.** *The neuron doctrine in perception* // Gazzaniga M. (ed.) *The cognitive neurosciences*. Boston: MIT Press, 1994.
- Barnard S., Matthews L., Messori S., Podaliri-Vulpiani M., Ferri N.** Laterality as an indicator of emotional stress in ewes and lambs during a separation test // *Anim. Cogn.* 2015. V. 19. P. 1–8.

- Baron-Cohen S.** The science of evil: On empathy and the origins of cruelty. N.Y.: Basic Books, 2011.
- Baroni M., Murphy B., Barbu E., Poesio M.** A corpus-based semantic model based on properties and types // *Cogn. Sci.* 2010. V. 34. P. 222–254.
- Barrett J.L.** Why would anyone believe in God? // Lanham, MD: Altamira Press, 2004.
- Barrett L.F.** Feeling is perceiving: core affect and conceptualization in the experience of emotion // Barrett L.F., Niedenthal P.M., Winkielman P. (eds.) *Emotion and consciousness*. N.Y.: Guilford Press, 2005. P. 255–284.
- Barrett L.F.** Are emotions natural kinds? // *Perspect. Psychol. Sci.* 2006. V. 1. P. 28–58.
- Barrett L.F., Russell J.A. (eds.)** The psychological construction of emotion. N.Y.: Guilford Press, 2015.
- Bartlett F.C.** *Remembering: An experimental and social study*. Cambridge, UK: Cambridge Univ. Press, 1932.
- Basile L.F., Sato J.R., Pasquini H.A., Velasques B., Ribeiro P., Anghinah R.** Individual versus task differences in slow potential generators // *Neurol. Sci.* 2021. V. 42, N 9. P. 3781–3789.
- Bastos A.M., Usrey W.M., Adams R.A., Mangun G.R., Fries P., Friston K.J.** Canonical microcircuits for predictive coding // *Neuron*. 2012. V. 76. P. 695–711.
- Bates E., Dick F.** Beyond phrenology: Brain and language in the next millennium // *Brain Lang.* 2000. V. 71. P. 18–21.
- Batson C.D.** Self-other merging and the empathy-altruism hypothesis: reply to Neuberg et al. // *J. Pers. Soc. Psychol.* 1997. V. 73. P. 517–522.
- Batson C.D.** Empathy-induced altruism and morality: No necessary connection // Maibom H. (ed.) *Empathy and morality*. Oxford: Oxford Univ. Press, 2014. P. 41–58.
- Bayley P.J., Frascino J.C., Squire L.R.** Robust habit learning in the absence of awareness and independent of the medial temporal lobe // *Nature*. 2005. V. 436. P. 550–553.
- Bayne T.** Free will and the phenomenology of agency // Timpe K., Griffith M., Levy N. (eds.) *The Routledge companion to free will*. Abingdon: Routledge, 2017. P. 633–644.
- Beauregard M., Paquette V.** Neural correlates of a mystical experience in Carmelite nuns // *Neurosci. Lett.* 2006. V. 405. P. 186–190.
- Bechara A., Damásio A.R.** The Somatic Marker Hypothesis: A neural theory of economic decision // *GEB*. 2005. V. 52. P. 336–372.
- Beck F., Eccles J.** Quantum aspects of brain activity and the role of consciousness // *PNAS*. 1992. V. 89. P. 11357–11361.
- Becker E.** *The denial of death* (1st ed.). N.Y., NY: The Free Press, 1973.
- Bedau M.** *Weak emergence. Philosophical perspectives, mind, causation, and world*. Oxford: Blackwell, 1997. P. 375–399.

- Bednarik R.G.** The domestication of humans // *Anthropologie*. 2008. V. 46. P. 1–17.
- Beer R.** Dynamical approaches to cognitive science // *Trends Cogn. Sci.* 2000. V. 4. P. 91–99.
- Bering J.M.** The God instinct. The psychology of souls, destiny and the meaning of life. L.: Nicholas Brealy, 2011.
- Bering J.M., Bjorklund D.F.** The natural emergence of reasoning about the afterlife as a developmental regularity // *Dev. Psychol.* 2004. V. 40. P. 217–233.
- Berker S.** The normative insignificance of neuroscience // *Philosophy & Public Affairs*. 2009. V. 37, N 4. P. 293–329.
- Berry D.C., Broadbent D.E.** Interactive tasks and the implicit-explicit distinction // *Brit. J. Psychol.* 1988. V. 79. P. 251–272.
- Birkhoff G., von Neumann J.** The logic of quantum mechanics // *Ann. Math.* 1936. V. 37. P. 823–843.
- Biswal B., Yetkin F.Z., Haughton V.M., Hyde J.S.** Functional connectivity in the motor cortex of resting human brain using echo-planar MRI // *Magn. Reson. Med.* 1995. V. 34. P. 537–541.
- Blackburn S.** *Essays in quasi-realism*. Oxford: Oxford Univ. Press, 1993.
- Blake R., Brascamp J., Heeger D.J.** Can binocular rivalry reveal neural correlates of consciousness? // *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 2014. V. 369, N 1641. P. 20130211.
- Blasi A.** The moral functioning of mature adults and the possibility of fair moral reasoning // Narvaez D., Lapsley D.K. (eds.) *Personality, identity, and character*. Cambridge: Cambridge Univ. Press, 2009. P. 396–440.
- Block N.** Psychologism and behaviorism // *Phil. Rev.* 1981. V. 90. P. 5–43.
- Block N.** Can the mind change the world? // Boolos G. (ed.) *Meaning and method: essays in honor of Hilary Putnam*. Cambridge: Cambridge Univ. Press, 1990.
- Block N.** What is Dennett's theory a theory of? // *Philos. Top.* 1994. V. 22, N 1/2. P. 23–40.
- Block N.** On a confusion about a function of consciousness // *Behav. Brain Sci.* 1995. V. 18. P. 227–247.
- Block N.** Consciousness, accessibility and the mesh between psychology and neuroscience // *Behav. Brain Sci.* 2007. V. 30. P. 481–548.
- Block N.** If perception is probabilistic, why does it not seem probabilistic? // *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 2018. V. 373, N 1755. P. 20170341.
- Block N., Fodor J.** What psychological states are not // *Philos. Rev.* 1972. V. 81. P. 159–181.
- Blood A.J., Zatorre R.J.** Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion // *PNAS*. 2001. V. 98, N 20. P. 11818–11823.

- Bloom P.** Religion is natural // *Dev. Sci.* 2007. V. 10. P. 147–151.
- Bloom P.** Against empathy: The case for rational compassion. N.Y.: Ecco, 2016.
- Boccia M., Piccardi L., Guariglia P.** The meditative mind: A comprehensive meta-analysis of MRI Studies // *BioMed Res. Int.* 2015. V. 2015. P. 419808.
- Bode S., He A.H., Soon C.S., Trampel R., Turner R. et al.** Tracking the unconscious generation of free decisions using ultra-high field fMRI // *PLoS ONE.* 2011. V. 6, N 6. P. e21612.
- Boghossian P.** The status of content // *Philos. Rev.* 1990. V. 99. P. 157–184.
- Boghossian P.** The status of content revisited // *Pac. Philos. Q.* 1991. V. 71. P. 264–278.
- Bohr N.** The quantum postulate and the recent development of atomic theory // *Nature.* 1928. V. 121. P. 580–590.
- Bohr N.** Atomic physics and human knowledge. N.Y.: Wiley, 1958.
- Boly M., Massimini M., Tsuchiya N., Postle B.R., Koch C., Tononi G.** Are the neural correlates of consciousness in the front or in the back of the cerebral cortex? Clinical and neuroimaging evidence // *J. Neurosci.* 2017. V. 37, N 40. P. 9603–9613.
- Born M.** Zur Quantenmechanik der Stossvorgänge. 1926. V. 38. P. 803–827.
- Bowden E.M., Jung-Beeman M.** Aha! Insight experience correlates with solution activation in the right hemisphere // *Psychon. Bull. Rev.* 2003. V. 10. P. 730–737.
- Boyd R.** Materialism without reductionism: What physicalism does not entail // Block N. (ed.) *Readings in the Philosophy of Psychology.* Cambridge, MA: Harvard Univ. Press, 1980. V. 1.
- Bradley M.** The causal efficacy of qualia // *J. Conscious. Stud.* 2011. V. 18, N 11/12. P. 32–44.
- Brady M.S.** Emotional insight: the epistemic role of emotional experience. N.Y.: Oxford Univ. Press, 2013.
- Brewer J.A., Worhunsky P.D., Gray J.R., Tang Y.Y., Weber J., Kober H.** Meditation experience is associated with differences in default mode network activity and connectivity // *PNAS.* 2011. V. 108. P. 20254–20259.
- Brigandt I., Love A.** Reductionism in biology // *The Stanford encyclopedia of philosophy.* Zalta E.N. (ed.). 2017. URL: <https://plato.stanford.edu/archives/spr2017/entries/reduction-biology/>.
- Broadbent D.E.** Perception and communication. Oxford: Pergamon Press, 1958.
- Broadbent D.E., FitzGerald P., Broadbent M.H.P.** Implicit and explicit knowledge in the control of complex systems // *Brit. J. Psychol.* 1986. V. 77. P. 33–50.
- Brosnan R., de Waal F.** Evolution of responses to (un)fairness // *Science.* 2014. V. 346, N 6207. P. 1251776.

- Brown K.S., Marean C.W., Herries A.I.R., Jacobs Z., Tribolo C., Braun D., Roberts D.L., Meyer M.C., Bernatchez J.** Fire as an engineering tool of early modern humans // *Science*. 2009. V. 325, N 5942. P. 859–862.
- Bucciarelli M., Khemlani S., Johnson-Laird P.N.** The psychology of moral reasoning // *Judgm.* 2008. V. 3. P. 121–139.
- Buchanan T.W.** Retrieval of emotional memories // *Psychol. Bull.* 2007. V. 133, N 5. P. 761–779.
- Buchanan A., Powell R.** Precis of the evolution of moral progress: A biocultural theory // *Analyse & Kritik*. 2019. V. 41, N 2. P. 183–193.
- Buchanan A.** Our moral fate: Evolution and the escape from tribalism. Cambridge: MIT Press, 2020.
- Bundesen C., Habekost T.** Principles of visual attention: Linking mind and brain. Oxford: Oxford Univ. Press, 2008.
- Bunge M.** The mind-body problem. Oxford: Pergamon, 1980.
- Burge T.** Origins of objectivity. Oxford, UK: Oxford Univ. Press, 2010.
- Burgess P., Stuss D.T.** Fifty years of prefrontal cortex research: Impact on assessment // *J. Int. Neuropsychol. Soc.* 2017. V. 23. P. 755–767.
- Burke B.L., Martens A., Faucher E.H.** Two decades of terror management theory: A meta-analysis of mortality salience research // *Pers. Soc. Psychol. Rev.* 2010. V. 14, N 2. P. 155–195.
- Burke E.** A philosophical enquiry into the origin of our ideas of the sublime and beautiful. Oxford: Oxford Univ. Press, 1990.
- Burton R.A.** Neurotheology. On being certain. Believing you are right even when you're not. N.Y. City: Macmillan Publishers/St. Martin's Press, 2008.
- Byrne J.H.** Synaptic transmission in the central nervous system. 2020. URL: <https://nba.uth.tmc.edu/neuroscience/m/s1/chapter06.html>.
- Calder A.J., Lawrence A.D., Young A.W.** Neuropsychology of fear and loathing // *Nat. Rev. Neurosci.* 2001. V. 2, N 5. P. 352–363.
- Callaway E.** Oldest *Homo sapiens* fossil claim rewrites our species' history // *Nature*. 2017. V. 546. P. 289–293.
- Campbell J.** Control variables and mental causation // *Proc. Aristot. Soc.* 2010. V. 110. P. 15–30.
- Cannon W.B.** Bodily changes in pain, hunger, fear and rage, second edition. N.Y.: Appleton, 1929.
- Caramazza A., Zurif E.B.** Dissociation of algorithmic and heuristic processes in language comprehension: Evidence from aphasia // *Brain Lang.* 1976. V. 3. P. 572–582.
- Carritt E.F.** Philosophies of beauty. L.: Oxford Univ. Press, 1931.
- Carruthers P.** The centred mind: What the science of working memory shows us about the nature of human thought. Oxford: Oxford Univ. Press, 2015.

- Carruthers P.** Phenomenal consciousness. Cambridge: Cambridge Univ. Press, 2000.
- Casasanto D.** Embodiment of abstract concepts: Good and bad in right- and left-handers // *J. Exp. Psychol. Gen.* 2009. V. 138. P. 351–367.
- Casasanto D., Chrysikou E.G.** When left is «right»: Motor fluency shapes abstract concepts // *Psychol Sci.* 2011. V. 22. P. 419–422.
- Casler K., Kelemen D.** Reasoning about artifacts at 24 months: The developing teleo-functional stance // *Cognition.* 2007. V. 103, N 1. P. 120–130.
- Cela-Conde C.J., Marty G., Maestú F., Ortiz T., Munar E., Fernández A. et al.** Activation of the prefrontal cortex in the human visual aesthetic perception // *PNAS.* 2004. V. 101, N 16. P. 6321–6325.
- Cerf M., Thiruvengadam N., Mormann F., Kraskov A., Quian Quiroga R., Koch C., Fried I.** Online, voluntary control of human temporal lobe neurons // *Nature.* 2010. V. 467. P. 1104–1108.
- Chalmers D.J.** Facing up to the problem of consciousness // *J. Conscious. Stud.* 1995. V. 2, N 3. P. 200–219.
- Chalmers D.J.** The conscious mind. Oxford: Oxford Univ. Press, 1996.
- Chalmers D.J.** Strong and weak emergence // Clayton P., Davies P. (eds.) *The re-emergence of emergence.* Oxford: Oxford Univ. Press, 2006.
- Chalmers D.J.** Panpsychism and panprotopsychism. Amherst lecture in philosophy. 2013. URL: <http://consc.net/papers/panpsychism.pdf>
- Chang L., Tsao D.Y.** The code for facial identity in the primate brain // *Cell.* 2017. V. 169. P. 1013–1028.
- Chatterjee A.** Neuroaesthetics: A coming of age story // *J. Cogn. Neurosci.* 2011. V. 23. P. 53–62.
- Chatterjee A.** The aesthetic brain: How we evolved to desire beauty and enjoy art. N.Y., NY: Oxford Univ. Press, 2013.
- Chatterjee A., Thomas A., Smith S.E., Aguirre G.K.** The neural response to facial attractiveness // *Neuropsychol.* 2009. V. 23, N 2. P. 135–143.
- Chao L.L., Martin A.** Representation of manipulable man-made objects in the dorsal stream // *NeuroImage.* 2000. V. 12. P. 478–484.
- Chemero A.** Radical embodied cognitive science. Cambridge, MA: MIT Press, 2009.
- Chittka L., Niven J.** Are bigger brains better? // *Curr. Biol.* 2009. V. 19, N 21. P. R995–R1008.
- Church A.** An unsolvable problem of elementary number theory // *Am. J. Math.* 1936. V. 58. P. 345–363.
- Churchland P.M.** Eliminative materialism and the propositional attitudes // *J. Philos.* 1981. V. 78. P. 67–90.
- Churchland P.S.** Consciousness: The transmutation of a concept // *Pac. Philos. Q.* 1983. V. 64. P. 80–95.
- Churchland P.S.** Neurophilosophy: Toward a unified science of the mind-brain. Cambridge, Massachusetts: The MIT Press, 1986.

- Cikara M., Bruneau E., Van Bavel J.J., Saxe R.** Their pain gives us pleasure: How intergroup dynamics shape empathic failures and counter-empathic responses // *J. Exp. Soc. Psychol.* 2014. V. 55. P. 110–115.
- Clark A.** Whatever next? Predictive brains, situated agents, and the future of cognitive science // *Behav. Brain Sci.* 2013. V. 36. P. 181–204.
- Clark A.** Predictions, precision, and agentic attention // *Conscious. Cogn.* 2017. V. 56. P. 115–119.
- Clark-Polner E., Johnson T.D., Barrett L.F.** Multivoxel pattern analysis does not provide evidence to support the existence of basic emotions // *Cereb. Cortex.* 2016. V. 27. P. 1944–1948.
- Clarke D.D., Sokoloff L.** Circulation and energy metabolism of the brain // Agranoff B.W., Siegel G.J. (eds.) *Basic neurochemistry. Molecular, cellular and medical aspects.* Philadelphia: Lippincott-Raven, 1999. P. 637–670.
- Clarke R.** *Libertarian accounts of free will.* Oxford: Oxford Univ. Press, 2003.
- Cloninger C.R.** A unified biosocial theory of personality and its role in the development of anxiety states: A reply to commentaries // *Psychiatry Develop.* 1988. V. 2. P. 83–120.
- Cohen M.A., Cavanagh P., Chun M.M., Nakayama K.** The attentional requirements of consciousness // *Trends Cogn. Sci.* 2012. V. 16. P. 411–417.
- Collette F., van der Linden M.** Brain imaging of the central executive component of working memory // *Neurosci. Biobehav. Rev.* 2002. V. 26. P. 105–125.
- Colombetti G.** *The feeling body: Affective science meets the enactive mind.* Cambridge, MA: MIT Press, 2014.
- Colombo M., Seriès P.** Bayes in the brain-on bayesian modelling in neuroscience // *Br. J. Philos. Sci.* 2012. V. 63. P. 697–723.
- Comte A.** *Cours de philosophie positive: la partie historique de la philosophie sociale en tout ce qui concerne l'état théologique et l'état métaphysique.* Paris: Bachelier, 1841. V. 5.
- Conway M.A., Pleydell-Pearce C.W.** The construction of autobiographical memories in the self-memory system // *Psychol. Rev.* 2000. V. 107. P. 261–288.
- Corr P.J.** Reinforcement sensitivity theory and personality // *Neurosci. Biobehav. Rev.* 2004. V. 28. P. 317–332.
- Cosmides L., Barrett H.C., Tooby J.** Colloquium paper: Adaptive specializations, social exchange, and the evolution of human intelligence // *PNAS.* 2010. V. 107. Suppl. 2. P. 9007–9014.
- Cosmides L., Tooby J.** Can a general deontic logic capture the facts of human moral reasoning? // Sinnott-Armstrong W. (ed.) *Moral psychology.* Boston: Boston Review, 2008. V. 1: The evolution of morality: Adaptations and innateness. P. 53–119.

- Costa-Mattioli M., Sonenberg N.** Translational control of gene expression: A molecular switch for memory storage // *Progr. Brain Res.* 2008. V. 169. P. 81–95.
- Cottingham J.** *Philosophy of religion: Towards a more humane approach.* N.Y.: Cambridge Univ. Press, 2014.
- Cover T., Thomas J.** *Elements of information theory.* Hoboken: Wiley, 2006.
- Crick F., Koch C.** Toward a neurobiological theory of consciousness // *Semin. Neurosci.* 1990. V. 2. P. 263–275.
- Cupchik G.C., Vartanian O., Crawley A., Mikulis D.J.** Viewing artworks: contributions of cognitive control and perceptual facilitation to aesthetic experience // *Brain Cogn.* 2009. V. 70. P. 84–91.
- Curry J.R.** *Children of God: children of earth.* Bloomington, IN: Author House, 2008.
- Cushen P.J., Wiley J.** Cues to solution, restructuring patterns, and reports of insight in creative problem solving // *Conscious. Cogn.* 2012. V. 21. P. 1166–1175.
- Cushman F., Young L., Greene J.D.** *Multi-system moral psychology. The moral psychology handbook.* Oxford: Oxford Univ. Press, 2010. P. 47–71.
- Damásio A.R.** The somatic marker hypothesis and the possible functions of the prefrontal cortex // *Philos. Trans. R. Soc. B.* 1996. V. 351. P. 1413–1420.
- Damasio A.** *The feeling of what happens: Body and emotion in the making of consciousness.* N.Y.: Harcourt, 1999.
- Damasio A.R.** *Looking for Spinoza: Joy, sorrow, and the feeling brain.* Orlando, FL: Harcourt, 2003.
- Danko S., Starchenko M., Bechtereva N.** EEG local and spatial synchronization during a test on the insight strategy of solving creative verbal tasks // *Hum. Physiol.* 2003. V. 29. P. 129–132.
- Danto A.** *The abuse of beauty.* Chicago: Open Court, 2003.
- Darwin C.** *The descent of man, and selection in relation to sex* (1st ed.). L.: John Murray, 1871.
- Darwin C.** *The expression of the emotions in man and animals.* 3rd ed. (ed. Ekman P.). L.: Harper Collins; N.Y.: Oxford Univ. Press, 1998.
- Dávid-Barrett T., Carney J.** The deification of historical figures and the emergence of priesthoods as a solution to a network coordination problem // *Religion Brain Behav.* 2015. V. 6. P. 307–317.
- Davidson J.E., Sternberg R.J. (eds.).** *The psychology of problem solving.* Cambridge: Cambridge Univ. Press, 2003.
- Davies M., Stone T.** *Folk psychology.* Oxford: Blackwell Publishers, 1995.
- Davis P.** *Information and the nature of reality.* Cambridge: Cambridge Univ. Press, 2010.
- Dawkins R.** *The god delusion.* Boston, NY: Houghton Mifflin Co., 2006.
- Dawkins R.** *The selfish gene.* 4-th ed. Oxford: Oxford Univ. Press, 2016.

- DeCasien A.R., Williams S.A., Higham J.P.** Primate brain size is predicted by diet but not sociality // *Nat. Ecol. Evol.* 2017. V. 1. P. 112.
- Decety J., Cowell J.M.** Empathy, justice, and moral behavior // *AJOB Neurosci.* 2015. V. 6. P. 3–14.
- Decety J., Lamm C.** Human empathy through the lens of social neuroscience // *Sci. World J.* 2006. V. 6. P. 1146–1163.
- Dediu D., Levinson S.C.** Neanderthal language revisited: Not only us // *Curr. Opin. Behav. Sci.* 2018. V. 21. P. 49–55.
- Deecke L.** There are conscious and unconscious agendas in the brain and both are important-our will can be conscious as well as unconscious // *Brain Sci.* 2012. V. 2. P. 405–420.
- De Fontenelle B.** De l'origine des fables // *Oeuvres de Fontenelle.* Paris: J. Pinard, 1824. P. 294–310.
- Dehaene S., Changeux J.P., Nacchache L., Sackut J., Sergent C.** Conscious, preconscious, and subliminal processing: a testable taxonomy // *Trends Cogn. Sci.* 2006. V. 10. P. 204–211.
- Dehaene S., Naccache L.** Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework // *Cogn.* 2000. V. 79. P. 1–37.
- Dehaene S., Piazza M., Pinel P., Cohen L.** Three parietal circuits for number processing // *Cogn. Neuropsychol.* 2003. V. 20. P. 487–506.
- De Neys W.** Conflict detection, dual processes and logical intuitions: Some clarifications // *Think. Reason.* 2014. V. 20. P. 169–187.
- De Neys W., Bialek M.** Dual processes and moral conflict: Evidence for deontological reasoners' intuitive utilitarian sensitivity // *Judgm. Decis. Mak.* 2017. V. 12. P. 148–167.
- Dennett D.** Why you can't make a computer that feels pain // *Brainstorms.* Cambridge, MA: MIT Press, 1978. P. 190–229.
- Dennett D.** Quining qualia // Marcel A., Bisiach E. (eds.) *Consciousness in contemporary science.* N.Y.: Oxford Univ. Press, 1988. P. 42–77.
- Dennett D.** *Consciousness explained.* N.Y.: The Penguin Press, 1991.
- Deonna J.A., Teroni F.** Emotions as attitudes // *Dialectica.* V. 69. P. 293–311.
- D'Errico F., Henshilwood C., Vanhaeren M., van Niekerk K.** Nassarius kraussianus shell beads from Blombos Cave: Evidence for symbolic behaviour in the middle stone age // *J. Hum. Evol.* 2005. V. 48. P. 3–24.
- Descartes R.** *The Principles of philosophy.* Translated by E. Haldane and G. Ross. Cambridge: Cambridge Univ. Press, 1911.
- Deubel H., Schneider W.X.** Attentional selection in sequential manual movements, movements around an obstacle and in grasping // Humphries G.W., Riddoch M.J. (eds.) *Attention in action.* Hove: Psychology Press, 2004. P. 61–91.
- Deutsch J.A., Deutsch D.** Attention: Some theoretical considerations // *Psychol. Rev.* 1963. V. 70. P. 80–90.

- De Waal F.B.M.** Peacemaking among primates. Cambridge, MA: Harvard Univ. Press, 1989.
- De Waal F.** Good natured: The origins of right and wrong in humans and other animals. Cambridge, MA: Harvard Univ. Press, 1996.
- De Waal F.** Primates and philosophers: How morality evolved. Princeton, N.J.: Princeton Univ. Press, 2006.
- DeWitt I., Rauschecker J.P.** Phoneme and word recognition in the auditory ventral stream // PNAS. 2012. V. 109. P. E505–E514.
- Dictionary** of philosophy of religion, Taliaferro and Marty. 2010. P. 196–197; 2018. P. 240.
- Dienes Z.** Subjective measures of unconscious knowledge // Progr. Brain Res. 2007. V. 168. P. 49–64.
- Dienes Z., Perner J.** A theory of the implicit nature of implicit learning // French R.M., Cleeremans A. (eds.) Implicit learning and consciousness: An empirical, philosophical, and computational consensus in the making? N.Y.: Psychology Press, 2002. P. 68–92.
- Dienes Z., Perner J.** Unifying consciousness with explicit knowledge // Cleeremans L. (ed.) The unity of consciousness. Oxford: Oxford Univ. Press., 2003. P. 214–232.
- Dietrich A., Kanso R.** A review of EEG, ERP, and neuroimaging studies of creativity and insight // Psychol. Bull. 2010. V. 136. P. 822–848.
- Dijksterhuis A., Nordgren L.F.** A theory of unconscious thought // Perspect. Psychol. Sci. 2006. V. 1. P. 95–109.
- Di Paolo E., Thompson E.** The enactive approach // Shapiro L. (ed.) The Routledge handbook of embodied cognition. N.Y.: Routledge, 2014. P. 68–78.
- Dittmar H., Bond R., Hurst M., Kasser T.** The relationship between materialism and personal well-being: A meta-analysis // J. Pers. Soc. Psychol. 2014. V. 107. P. 879–924.
- Drell J.R.R.** Neanderthals: A history of interpretation // Oxford J. Archaeol. 2000. V. 19. P. 1–24.
- Dretske F.** Knowledge and the flow of information. Oxford: Blackwell, 1981.
- Dretske F.** Differences that make no difference // Philos. Top. 1994. V. 22, N 1/2. P. 41–58.
- Dreyfus H.** What computers still can't do. Cambridge, MA: MIT Press, 1992.
- Dubnov S.** Musical information dynamics as models of auditory anticipation // Weng W. (ed.) Machine audition: Principles, algorithms and systems. N.Y.: IGI Global Publ., 2010.
- Dudai Y.** The neurobiology of consolidations, or, how stable is the engram? // Ann. Rev. Psychol. 2003. V. 55. P. 51–86.
- Duke C.G., Kennedy A.J., Gavin C.F., Day J.J., Sweatt J.D.** Experience-dependent epigenomic reorganization in the hippocampus // Learn. Mem. 2017. V. 24. P. 278–288.

- Dunbar R.I.M.** Neocortex size as a constraint on group size in primates // *J. Hum. Evol.* 1992. V. 22. P. 469–493.
- Dunbar R.I.M.** The social brain hypothesis // *Evol. Anthropol.* 1998. V. 6. P. 178–190.
- Duncan J.** Selective attention and the organization of visual information // *J. Exp. Psychol. Gen.* 1984. V. 113. P. 501–517.
- Dutton D.** Aesthetics and evolutionary psychology // *The Oxford handbook for aesthetics.* Oxford: Oxford Univ. Press, 2003.
- Dworkin R.** Law's empire. Cambridge, MA: Belknap Press, 1986.
- Eagleman D.** Visual illusions and neurobiology // *Nat. Rev. Neurosci.* 2001. V. 2. P. 920–926.
- Ecklund E.H.** Science vs religion: What scientists really think. Oxford: Oxford Univ. Press, 2010.
- Ecklund E.H., Scheitle C.P.** Religion among academic scientists: distinctions, disciplines, and demographics // *Soc. Probl.* 2007. V. 54. P. 289–307.
- Eco U.** Art and beauty in the middle ages. H. Bredin trans. New Haven: Yale Univ. Press, 1986.
- Edelman G.** The remembered present: A biological theory of consciousness. N.Y.: Basic Books, 1989.
- Edmonds D.** Would you kill the fat man? The trolley problem and what your answer tells us about right and wrong. Princeton, N.J.: Princeton Univ. Press, 2014. P. 137–139.
- Egloff B., Schmukle S.C.** Predictive validity of an implicit association test for assessing anxiety // *J. Pers. Soc. Psychol.* 2002. V. 83. P. 1441–1455.
- Eisenberger N., Lieberman M.D., Williams K.** Does rejection hurt? An fMRI study of social exclusion // *Science.* 2003. V. 302. P. 290–292.
- Ekman P.** Emotion in the human face. 2nd ed. Cambridge, UK: Cambridge Univ. Press, 1982.
- Ekman P.** An argument for basic emotions // *Cogn. Emot.* 1992. V. 6. P. 169–200.
- Elfenbein H.A., Ambady N.** On the universality and cultural specificity of emotion recognition: A meta-analysis // *Psychol. Bull.* 2002. V. 128. P. 203–235.
- Elliot A.J.** The hierarchical model of approach-avoidance motivation // *Motiv. Emot.* 2006. V. 30. P. 111–116.
- Elliott M.L. et al.** What is the test-retest reliability of common task-functional MRI measures? New empirical evidence and a meta-analysis // *Psychol. Sci.* 2020. V. 31. P. 792–806.
- Ellis G.F.R.** On the nature of causation in complex systems // *Trans. R. Soc. S. Afr.* 2008. V. 63. P. 69–84.
- Ellis G.** How can physics underlie the mind? Berlin: Springer-Verlag, 2016.
- Elmholdt E.-M., Skewes J., Dietz M., Møller A., Jensen M.S., Roepstorff A., Wiech K., Jensen T.S.** Reduced pain sensation and reduced

- BOLD signal in parietofrontal networks during religious prayer // *Front. Hum. Neurosci.* 2017. V. 11. P. 337.
- Emmons N.A., Kelemen D.** The development of children's prelife reasoning: Evidence from two cultures // *Child Dev.* 2014. V. 85. P. 1617–1633.
- Elman J.** Finding structure in time // *Cogn. Sci.* 1990. V. 14. P. 179–211.
- Ewert J.-P.** Motion perception shapes the visual world of amphibians // Prete F.R. (ed.) *Complex worlds from simpler nervous systems.* Cambridge, MA: MIT Press, 2004. P. 117–160.
- Falk D.** Brain evolution in homo: The «radiator» theory // *Behav. Brain Sci.* 1990. V. 13. P. 333–344.
- Falk E.B., Berkman E.T., Mann T., Harrison B., Lieberman M.D.** Predicting persuasion-induced behavior change from the brain // *J. Neurosci.* 2010. V. 30. P. 8421–8424.
- Farah M.J.** *Visual agnosia: Disorders of object recognition and what they tell us about normal vision.* 2nd ed. Cambridge, MA: MIT Press, 2004.
- Fechner G.T.** *Elemente der psychophysik.* Leipzig: Breitkopf & Härtel, 1860.
- Ferguson M.A., Nielsen J.A., King J.B., Dai L., Giangrosso D.M., Holman R., Korenberg J.R., Anderson J.S.** Reward, salience, and attentional networks are activated by religious experience in devout Mormons // *Soc. Neurosci.* 2017. V. 13. P. 104–116.
- Ferguson M.A., Schaper F., Cohen A., Siddiqi S., Merrill S.M., Nielsen J.A., Grafman J., Urgesi C., Fabbro F., Fox M.D.** A neural circuit for spirituality and religiosity derived from patients with brain lesions // *Biol. Psychiatry.* 2021. V. 91. P. 380–388.
- Figdor C.** Semantic externalism and the mechanics of thought // *Minds Mach.* 2009. V. 19. P. 1–24.
- Fifer W.P., Byrd D.L., Kaku M., Eigsti I.-M., Isler J.R., Grose-Fifer J. et al.** Newborn infants learn during sleep // *PNAS.* 2010. V. 107. P. 10320–10323.
- Firth R.** Ethical absolutism and the ideal observer // *Philos. Phenomenol. Res.* 1952. V. 12. P. 317–345.
- Firth J.R.** A synopsis of linguistic theory 1930–1955 // *Stud. Linguist. Anal.* 1957. V. 1. P. 1–32.
- Fisher R.A.** The evolution of sexual preference // *Eugenics Rev.* 1915. V. 7. P. 184–192.
- Fisher R.A.** *The genetical theory of natural selection.* Oxford, UK: Clarendon Press, 1930.
- FitzPatrick W.** Morality and evolutionary biology // Zalta E.N. (ed.) *The Stanford encyclopedia of philosophy.* 2021. URL: <https://plato.stanford.edu/archives/spr2021/entries/morality-biology/>.
- Flew A.** *Axiology. A dictionary of philosophy* editorial consultant. N.Y.: Macmillan, 1979.

- Flinn M.V., Geary D.C., Ward C.V.** Ecological dominance, social competition, and coalitionary arms races: Why humans evolved extraordinary intelligence // *Evol. Hum. Behav.* 2005. V. 26. P. 10–46.
- Flohr H.** An information processing theory of anesthesia // *Neuropsychologia.* 1995. V. 33/9. P. 1169–1180.
- Fodor J.** The language of thought. N.Y.: Thomas Y. Crowell, 1975.
- Fodor J.** Methodological solipsism considered as a research strategy in cognitive psychology // *Behav. Brain Sci.* 1980. V. 3. P. 63–73.
- Fodor J.** Concepts. Oxford: Clarendon Press, 1998.
- Fodor J.** The mind doesn't work that way. Cambridge, MA: MIT Press, 2000.
- Fodor J.** Reply to Steven Pinker 'So how does the mind work?' // *Mind Lang.* 2005. V. 20. P. 25–32.
- Fodor J.A., Pylyshyn Z.W.** Connectionism and cognitive architecture: A critical analysis // *Cognition.* 1988. V. 28. P. 3–71.
- Foerde K., Poldrack R.A.** Procedural learning in humans. The new encyclopedia of neuroscience. Amsterdam: Elsevier, 2009. P. 1083–1091.
- Foster J.** A defense of dualism // Smythies J., Beloff J. (eds.) The case for dualism. Charlottesville, VA: Univ. of Virginia Press, 1989.
- Foster J.** The immaterial self: A defense of the Cartesian dualist conception of mind. L.: Routledge, 1996.
- Foster H.** The anti-aesthetic: Essays on postmodern culture. L.: New Press, 1998.
- Fox E.** Emotion science: An integration of cognitive and neuroscientific approaches. L.: Palgrave MacMillan, 2008. P. 16–17.
- Fox K.C.R. et al.** Functional neuroanatomy of meditation: A review and meta-analysis of 78 functional neuroimaging investigations // *Neurosci. Biobehav. Rev.* 2016. V. 65. P. 208–228.
- Frankena W.** Thinking about morality. Ann Arbor, MI: Univ. of Michigan Press, 1980.
- Frankish K.** Illusionism as a theory of consciousness // *J. Conscious. Stud.* 2016. V. 23. P. 11–39.
- Frankish K.** Illusionism as a theory of consciousness. Exeter: Imprint Academic Publishing, 2017.
- Freedberg D., Gallese V.** Motion, emotion and empathy in esthetic experience // *Trends Cogn. Sci.* 2007. V. 11. P. 197–203.
- Freeman A.** Signs of the times: Cracking the code of art's allure. Paris: The Unesco Courier, 1999.
- Fridlund A.J.** Human facial expression: An evolutionary view. San Diego, CA: Acad. Press, 1994.
- Friedman W.J.** Time in autobiographical memory // *Soc. Cogn.* 2004. V. 22. P. 591–605.
- Friston K.** Hierarchical models in the brain // *PLoS Comput. Biol.* 2008. V. 4. P. e1000211.

- Friston K., Kilner J., Harrison L.** A free energy principle for the brain // *J. Physiol.* 2006. V. 100. P. 70–87.
- Frith U., Frith C.D.** Development and neurophysiology of mentalizing // *Philos. Trans. R. Soc. B.* 2003. V. 358. P. 459–473.
- Fuchs T.** Empathy, group identity, and the mechanisms of exclusion: An investigation into the limits of empathy // *Topoi.* 2019. V. 38. P. 239–250.
- Gallagher S.** A pattern theory of self // *Front. Hum. Neurosci.* 2013. V. 7. P. 443.
- Gallagher S.** *Enactivist interventions: Rethinking the mind.* Oxford: Oxford Univ. Press, 2017.
- Gallese V.** Finding the body in the brain: From simulation theory to embodied simulation // *McLaughlin B., Kornblith H.K. (eds.) Goldman and his critics.* Oxford: Wiley, 2016. P. 297–314.
- Gallistel C.R.** *The organization of action: A new synthesis.* Hillsdale, N.J.: Erlbaum, 1980.
- Gallistel C.R., King A.** *Memory and the computational brain.* Malden: Wiley-Blackwell, 2009.
- Galván B., Hernández C.M., Mallol C., Mercier N., Sistiaga A., Soler V.** New evidence of early Neanderthal disappearance in the Iberian Peninsula // *J. Hum. Evol.* 2014. V. 75. P. 16–27.
- Garner R.T., Rosen B.** *Moral philosophy: A systematic introduction to normative ethics and meta-ethics.* N.Y.: Macmillan, 1967.
- Gayon J.** Sexual selection: Another darwinian process // *Comptes Rendus Biologies.* 2010. V. 333. P. 134–144.
- Gazzaniga M.** *Mind matters: How mind and brain interact to create our conscious lives.* Boston: Houghton Mifflin, 1988.
- Gazzaniga M.** *Who's in charge? Free will and the science of the brain.* N.Y.: Harper Collins, 2011.
- Gelder B., Tamietto M., van Boxtel G., Goebel R., Sahraie A., van den Stock J., Stienen B.M.C., Weiskrantz L., Pegna A.** Intact navigation skills after bilateral loss of striate cortex // *Curr. Biol.* 2008. V. 18. P. R1128–R1129.
- Gennaro R.** *Consciousness and self-consciousness: A defense of the higher-order thought theory of consciousness.* Amsterdam; Philadelphia: John Benjamins, 1995.
- Gert B.** *Morality: Its nature and justification, revised edition.* N.Y.: Oxford Univ. Press, 2005.
- Gert B., Gert J.** The definition of morality // *Zalta E.N. (ed.) The Stanford encyclopedia of philosophy.* 2020. URL: <https://plato.stanford.edu/archives/fall2020/entries/morality-definition/>.
- Geschwind N.** Disconnexion syndromes in animals and man // *Brain.* 1965. V. 88. P. 237–294.

- Geschwind N., Levitsky W.** Human brain: Left-right asymmetries in temporal speech region // *Science*. 1968. V. 161. P. 186–187.
- Gibbons A.** Bonobos join chimps as closest human relatives. *Science*. 2012. URL: <https://www.sciencemag.org/news/2012/06/bonobos-join-chimps-closest-human-relatives>.
- Gierliński G.D. et al.** Possible hominin footprints from the late Miocene (c. 5.7 Ma) of Crete? // *Proc. Geol. Assoc.* 2017. V. 128. P. 697–710.
- Gleick J.** *The information*. N.Y.: Pantheon, 2011.
- Glenn A.L., Raine A.** The immoral brain // Verplaetse J., De Schrijver J., Vanneste S., Braeckman J. (eds.) *The moral brain*. Heidelberg: Springer Dordrecht, 2009.
- Gödel.K.** Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme, I // *Monatshefte für Mathematik und Physik*. 1931. V. 38. P. 173–198.
- Goel V.** Anatomy of deductive reasoning // *Trends Cogn. Sci.* 2007. V. 11. P. 435–441.
- Goff P., Seager W., Allen-Hermanson S.** Panpsychism // Zalta E.N. (ed.) *The Stanford encyclopedia of philosophy*. 2020. URL: <https://plato.stanford.edu/archives/sum2020/entries/panpsychism/>.
- Goldman A.** *Simulating minds: the philosophy, psychology, and neuroscience of mindreading*. Oxford: Oxford Univ. Press, 2006.
- Goldman A.** *Mirroring, simulating, and mindreading* // *Mind Lang.* 2009. V. 24. P. 235–252.
- Goldman R.I., Stern J.M., Engel J., Cohen M.S.** Simultaneous EEG and fMRI of the alpha rhythm // *NeuroReport*. 2002. V. 13. P. 2487–2492.
- Gonzalez-Cabrera I.** *Moving beyond dichotomies* // Matthew L.S. (ed.) *Moral brains: The neuroscience of morality*. Oxford: Oxford Univ. Press, 2017.
- Goodale M.A., Milner A.D.** *Sight unseen: An exploration of conscious and unconscious vision*. Oxford: Oxford Univ. Press, 2004.
- Gould S.J.** Exaptation: A crucial tool for an evolutionary psychology // *J. Soc. Iss.* 1991. V. 47. P. 43–65.
- Gow D.W.** The cortical organization of lexical knowledge: A dual lexicon model of spoken language processing // *Brain Lang.* 2012. V. 121. P. 273–288.
- Graf P., Squire L.R., Mandler G.** The information that amnesic patients do not forget // *J. Exp. Psychol.* 1984. V. 10. P. 164–178.
- Graves A., Wayne G., Danihelka I.** Neural turing machines // *arXiv*. 2014. 1410.5401.
- Gray J.A.** The psychophysiological basis of introversion-extraversion // *Behav. Res. Ther.* 1970. V. 18. P. 249–266.
- Gray J.** The contents of consciousness: A neuropsychological conjecture // *Behav. Brain Sci.* 1995. V. 18/4. P. 659–722.

- Greenberg G.** Approach/withdrawal theory // Vonk J., Shackelford T.K. (eds.) *Encyclopedia of animal cognition and behavior*. N.Y.: Springer International Publishing, 2017.
- Greenberg D., Verfaellie M.** Interdependence of episodic and semantic memory: Evidence from neuropsychology // *J. Int. Neuropsychol. Soc.* 2010. V. 16. P. 748–753.
- Greene J.** *Moral tribes: Emotion, reason, and the gap between us and them*. N.Y.: Penguin, 2013.
- Greene J.D.** Beyond point-and-shoot morality: Why cognitive (neuro)science matters for ethics // *Ethics*. 2014. V. 124. P. 695–726.
- Greene J.D.** The rat-a-gorical imperative: moral intuition and the limits of affective learning // *Cognition*. 2017. V. 167. P. 66–77.
- Greene J.D., Sommerville R.B., Nystrom L.E., Darley J.M., Cohen J.D.** An fMRI investigation of emotional engagement in moral judgment // *Science*. 2001. V. 293. P. 2105–2108.
- Greenwald A.G.** Unconscious cognition reclaimed // *Am. Psychologist*. 1992. V. 47. P. 766–779.
- Greenwald A.G.** Self-knowledge and self-deception: Further consideration // Myslobodsky M.S. (ed.) *The mythomanias: An inquiry into the nature of deception and self-deception*. Mahwah, N.J.: Erlbaum, 1997. P. 51–71.
- Greenwald A.G., Banaji M.R.** The implicit revolution: Reconceiving the relation between conscious and unconscious // *Am. Psychologist*. 2017. V. 72. P. 861–871.
- Greenwald A.G., McGhee D.E., Schwartz J.L.K.** Measuring individual differences in implicit cognition: The implicit association test // *J. Pers. Soc. Psychol.* 1998. V. 74. P. 1464–1480.
- Greicius M.D., Krasnow B., Reiss A.L., Menon V.** Functional connectivity in the resting brain: A network analysis of the default mode hypothesis // *PNAS*. 2003. V. 100. P. 253–258.
- Griffiths R., Richards W., Johnson M., McCann U., Jesse R.** Mystical-type experiences occasioned by psilocybin mediate the attribution of personal meaning and spiritual significance 14 months later // *J. Psychopharmacol.* 2008. V. 22. P. 621–632.
- Griffiths T.L., Steyvers M., Tenenbaum J.B.** Topics in semantic representation // *Psychol. Rev.* 2007. V. 114. P. 211–244.
- Grill-Spector K., Knouf N., Kanwisher N.** The fusiform face area subserves face perception, not generic within-category identification // *Nat. Neurosci.* 2004. V. 7. P. 555–562.
- Grimaldi M.** Toward a neural theory of language: Old issues and new perspectives // *J. Neurolinguistics*. 2012. V. 25. P. 1–24.
- Grodzinsky Y., Santi A.** The battle for Broca's region // *Trends Cogn. Sci.* 2008. V. 12. P. 474–480.

- Gross C.** Single neuron studies of inferior temporal cortex // *Neuropsychologia*. 2008. V. 46. P. 841–852.
- Gross N., Simmons S.** The religiosity of american college and university professors // *Soc. Relig.* 2009. V. 70. P. 101–129.
- Guillon J.B.** Van inwagen on introspected freedom // *Philos. Stud.* 2014. V. 168. P. 645–663.
- Gulli R. et al.** Context-dependent representations of objects and space in the primate hippocampus during virtual navigation // *Nat. Neurosci.* 2020. V. 23. P. 103–112.
- Hacker P.M.S.** The sad and sorry history of consciousness: being among other things a challenge to the «consciousness studies community» // *Royal Institute of Philosophy*. 2012. V. 70. P. 1–21.
- Haggard P.** Conscious intention and motor cognition // *Trends Cogn. Sci.* 2005. V. 9. P. 290–295.
- Haidt J.** The emotional dog and its rational tail: A social intuitionist approach to moral judgment // *Psychol. Rev.* 2001. V. 108. P. 814–834.
- Haidt J.** The moral emotions // Davidson R.J., Scherer K.R., Goldsmith H.H. (eds.) *Handbook of affective sciences*. Oxford: Oxford Univ. Press, 2003. P. 852–870.
- Haidt J.** *The righteous mind: why good people are divided by politics and religion*. N.Y.: Pantheon, 2011.
- Hameroff S.R., Watt R.C.** Information processing in microtubules // *J. Theor. Biol.* 1982. V. 98. P. 549–561.
- Hamilton W.D.** The genetical evolution of social behavior // *J. Theor. Biol.* 1964. V. 7. P. 1–52.
- Hardcastle V.** *The myth of pain*. Cambridge, MA: MIT Press, 1999.
- Hardy B.L., Moncel M.H., Kerfant C. et al.** Direct evidence of Neanderthal fibre technology and its cognitive and behavioral implications // *Sci. Rep.* 2020. V. 10. P. 4889.
- Hare B., Wobber V., Wrangham R.** The self-domestication hypothesis: Evolution of bonobo psychology is due to selection against aggression // *Anim. Behav.* 2012. V. 83. P. 573–585.
- Hare R.M.** *The language of morals*. Oxford: Clarendon Press, 1952.
- Hart J., Kraut M.A.** Neural hybrid model of semantic object memory // Hart J., Kraut M.A. (eds.) *Neural basis of semantic memory*. N.Y.: Cambridge Univ. Press, 2007. P. 331–359.
- Harvey W.** *Exercitatio anatomica de motu cordis et sanguinis in animalibus with an English translation and annotations by Leake C.D. (trans.)*. Springfield, Illinois and Baltimore, Maryland: Charles C. Thomas, 1928.
- Hasker W.** *The emergent self*. Ithaca, N.Y.: Cornell Univ. Press, 1999.
- Hasson U., Malach R., Heeger D.J.** Reliability of cortical activity during natural stimulation // *Trends Cogn. Sci.* 2010. V. 14. P. 40–48.

- Hauser M., Cushman F., Young L., Kang-Xing Jin R., Mikhail J.** A dissociation between moral judgments and justifications // *Mind Lang.* 2007. V. 22. P. 1–21.
- Hebb D.O.** *The organization of behavior; a neuropsychological theory.* N.Y.: Wiley, 1949.
- Hegel G.W.F.** *Hegel's aesthetics: lectures on fine art: in 2 v.* / Knox T.M. (trans.). Oxford: Clarendon, 1975.
- Heining M., Young A.W., Ioannou G., Andrew C.M., Brammer M.J., Gray J.A. et al.** Disgusting smells activate human anterior insula and ventral striatum // *Ann. N. Y. Acad. Sci.* 2003. V. 1000. P. 380–384.
- Heisenberg W.** *Physics and philosophy.* N.Y.: Harper and Row, 1958.
- Helmholtz H. von.** *Handbuch der Physiologischen Optik.* Leipzig: Voss, 1867.
- Henry A.** Agentialism and the objection from attention capture // Paper presented to Canadian Philosophical Association, Ryerson Univ. Toronto, 2017.
- Henshilwood C., Marean C.** The origin of modern human behavior: Critique of the models and their test implications // *Curr. Anthropol.* 2003. V. 44. P. 627–651.
- Hickok G.** The cortical organization of speech processing: Feedback control and predictive coding the context of a dual-stream model // *J. Commun. Disord.* 2012. V. 45. P. 393–402.
- Hickok G., Poeppel D.** The cortical organization of speech processing // *Nat. Rev. Neurosci.* 2007. V. 8. P. 393–402.
- Highfield R.** *Why beauty is an advert for good genes.* L.: Telegraph Media Group Limited, 2008.
- Hilbert D., Ackermann W.** *Principles of mathematical logic.* Providence, Rhode Island, USA: AMS Chelsea Publishing, 1950.
- Hobbes T.** *Leviathan* / ed. Curly E. Indianapolis: Hackett Publishing Company, 1994.
- Hoefle S., Engel A., Basilio R., Alluri V., Toiviainen P., Cagy M., Moll J.** Identifying musical pieces from fMRI data using encoding and decoding models // *Sci. Rep.* 2018. V. 8. P. 2266.
- Hoffman M.** *Empathy and moral development.* Cambridge: Cambridge Univ. Press, 2000.
- Hoffman M.** Empathy, justice and social change // Maibom H. (ed.) *Empathy and morality.* Oxford: Oxford Univ. Press, 2014. P. 71–96.
- Hohwy J.** Attention and consciousness in the hypothesis testing brain // *Front. Psychol.* 2012. V. 3. P. 96.
- Hohwy J.** *The predictive mind.* Oxford: Oxford Univ. Press, 2013.
- Hood B.** *The self illusion: How the social brain creates identity.* Oxford, UK: Oxford Univ. Press, 2012.
- Horgan T.** Injecting the phenomenology of agency into the free will debate // *Oxford Studies in Agency and Responsibility.* 2015. V. 3. P. 34–61.

- Houdé O.** Inhibition and cognitive development // *Cogn. Dev.* 2000. V. 15. P. 63–73.
- Houdé O., Borst G.** Measuring inhibitory control in children and adults // *Front. Psychol.* 2014. V. 5. P. 616.
- Houdé O., Borst G.** Evidence for an inhibitory-control theory of the reasoning brain // *Front. Hum. Neurosci.* 2015. V. 9. P. 148.
- Hrdy S.B.** Mothers and others: The evolutionary origins of mutual understanding. Cambridge, MA: Harvard Univ. Press, 2011.
- Hsu N.S., Kraemer D.J.M., Oliver R.T., Schlichting M.L., Thompson-Schill S.L.** Color, context, and cognitive style: Variations in color knowledge retrieval as a function of task and subject variables // *J. Cogn. Neurosci.* 2011. V. 23. P. 2554–2557.
- Huang M., Bridge H., Kemp M.J., Parker A.J.** Human cortical activity evoked by the assignment of authenticity when viewing works of art // *Front. Hum. Neurosci.* 2011. V. 5. P. 134.
- Hughes R.I.G.** The structure and interpretation of quantum mechanics. Cambridge, MA: Harvard Univ. Press, 1989.
- Hufendiek R.** Explaining embodied emotions-with and without representations // *Philos. Explor.* 2018. V. 21. P. 319–331.
- Hume D.** A treatise of human nature. Oxford: Oxford Univ. Press, 1978.
- Hume D.** Of the standard of taste. essays moral and political. L.: George Routledge and Sons, 1894.
- Hume D.** The natural history of religion // Beauchamp T.L. (ed.) A dissertation on the passions: The natural history of religion. A critical ed. Oxford: Clarendon Press, 2007. P. 30–87.
- Hutcheson F.** An inquiry into the original of our ideas of beauty and virtue. Indianapolis: Liberty Fund, 2004.
- Huth A.G., de Heer W.A., Griffiths T.L., Theunissen F.E., Gallant J.L.** Natural speech reveals the semantic maps that tile human cerebral cortex // *Nature.* 2016. V. 532, N 7600. P. 453–458.
- Hutto D.D.** Truly enactive emotion // *Emot. Rev.* 2012. V. 4. P. 176–181.
- Hutto D., Myin E.** Radical enactivism: Basic minds without content. Cambridge, MA: MIT Press, 2012.
- Inbar Y., Pizarro D.A., Bloom P.** Conservatives are more easily disgusted than liberals // *Cogn. Emot.* 2009. V. 23. P. 714–725.
- Indriati E., Swisher C.C., Lepre C., Quinn R.L., Suriyanto R.A., Hascaryo A.T. et al.** The age of the 20 meter solo river terrace, Java, Indonesia and the survival of *Homo erectus* in Asia // *PLOS ONE.* 2011. V. 6. P. e21562.
- Irvine W.B.** Ahal!: The moments of insight that shape our world. N.Y., N.Y.: Oxford Univ. Press, 2015.
- Ishizu T., Zeki S.** Toward a brain-based theory of beauty // *PLOS ONE.* 2011. V. 6. P. e21852.

- Ishizu T., Zeki S.** A neurobiological enquiry into the origins of our experience of the sublime and beautiful // *Front. Hum. Neurosci.* 2014. V. 8. P. 891.
- Ison M., Quian Quiroga R., Fried I.** Rapid encoding of new memories by individual neurons in the human brain // *Neuron.* 2015. V. 87. P. 220–230.
- Izard C.E.** Forms and functions of emotions: Matters of emotion-cognition interactions // *Emot. Rev.* 2011. V. 3. P. 371–378.
- Jackson F.** Mind and illusion // Ludlow P., Nagasawa Y., Stoljar D. (eds.) *There's something about Mary: Essays on the knowledge argument.* Cambridge, MA: MIT Press, 2004.
- Jacobs R.H., Renken R., Cornelissen F.W.** Neural correlates of visual aesthetics-beauty as the coalescence of stimulus and internal state // *PLOS ONE.* 2012. V. 7. P. e31248.
- Jahanian A.** Quantifying aesthetics of visual design applied to automatic design. Berlin: Springer, 2016. P. 11–12.
- James W.** Are we automata? // *Mind.* 1879. V. 4. P. 1–22.
- James W.** What is an emotion? // *Mind.* 1884. V. 9. P. 188–205.
- James W.** *The principles of psychology.* N.Y.: Henry-Holt and Company, 1890.
- James W.** *The varieties of religious experience: A study in human nature.* N.Y.: Longmans, Green, 1902.
- James W.** Humanism and truth // *Mind.* 1904. V. 13. P. 457–475.
- Johnson R., Cureton A.** Kant's moral philosophy // Zalta E.N. (ed.) *The Stanford encyclopedia of philosophy*, 2021. URL: <https://plato.stanford.edu/archives/spr2021/entries/kant-moral/>.
- Johnson-Laird P.N.** Mental models and deduction // *Trends Cogn. Sci.* 2001. V. 5. P. 432–442.
- Jonas E., Fischer P.** Terror management and religion: Evidence that intrinsic religiousness mitigates worldview defense following mortality salience // *J. Pers. Soc. Psychol.* 2006. V. 91. P. 553–567.
- Jones M.N., Willits J., Dennis S.** Models of semantic memory // Busemeyer J.R., Townsend J.T. (eds.) *Oxford handbook of mathematical and computational psychology.* Oxford: Oxford Univ. Press, 2014.
- Jordania J.** *Why do people sing? Music in human evolution.* Berlin: Logos Verlag, 2011. P. 186–196.
- Joyce R.** *The evolution of morality.* Cambridge, MA: MIT Press, 2006.
- Jung-Beeman M., Bowden E., Haberman J., Frymiare J., Aramber-Liu S., Greenblatt R., Kounios J.** Neural activity when people solve problems with insight // *PLoS Biol.* 2004. V. 2. P. 500–510.
- Kagan J., Reznick J.S., Snidman N.** The physiology and psychology of behavioral inhibition in children // *Child Dev.* 1987. V. 58. P. 1459–1473.
- Kahneman D.** *Thinking, fast and slow.* N.Y.: Macmillan, 2011.

- Kahneman D., Tversky A.** Prospect theory: An analysis of decision under risk // *Econometrica*. 1979. V. 47. P. 263–291.
- Kaiser D.** History: Shut up and calculate! // *Nature*. 2014. V. 505. P. 153–155.
- Kamitani Y., Tong F.** Decoding the visual and subjective contents of the human brain // *Nat. Neurosci*. 2005. V. 8. P. 679–685.
- Kamm F.** *Intricate ethics: Rights, responsibilities, and permissible harm*. Oxford: Oxford Univ. Press, 2007.
- Kant I.** *Critique of pure reason*. Translated by N. Kemp Smith. N.Y.: MacMillan, 1929.
- Kant I.** *Critique of judgement* / J.H. Bernard, trans. N.Y.: Macmillan, 1951.
- Kanwisher N., McDermott J., Chun M.M.** The fusiform face area: A module in human extrastriate cortex specialized for face perception // *J. Neurosci*. 1997. V. 17. P. 4302–4311.
- Kass L.** *The ethics of human cloning*. Washington: AEI Press, 1998.
- Kasser T.** Materialistic values and goals // *Annu. Rev. Psychol*. 2016. V. 67. P. 489–514.
- Kauffman S.A.** *Autonomous agents* // Barrow J.D., Davies P.C.W., Harper Jr. C.L. (eds.) *Science and ultimate reality: quantum theory, cosmology and complexity*. Philadelphia; L.: Templeton Foundation, 2004.
- Kawabata H., Zeki S.** Neural correlates of beauty // *J. Neurophysiol*. 2004. V. 91. P. 1699–1705.
- Kay K.N., Naselaris T., Prenger R.J., Gallant J.L.** Identifying natural images from human brain activity // *Nature*. 2008. V. 452, N 7185. P. 352–355.
- Kazez J.** Computationalism and the causal role of content // *Philos. Stud*. 1995. V. 75. P. 231–260.
- Kelemen D.** The scope of teleological thinking in preschool children // *Cognition*. 1999. V. 70. P. 241–272.
- Kelemen D.** British and American children's preferences for teleofunctional explanations of the natural world // *Cognition*. 2003. V. 88. P. 201–221.
- Kelemen D., Emmons N., Brown S.A., Gallik C.** Beliefs about origins and eternal life: How easy is formal religious theory development? // *J. Cogn. Dev*. 2021. V. 22. P. 356–378.
- Kelemen D., Rosset E.** The human function compunction: Teleological explanation in adults // *Cognition*. 2009. V. 111. P. 138–143.
- Kelemen D., Rottman J., Seston R.** Professional physical scientists display tenacious teleological tendencies: Purpose-based reasoning as a cognitive default // *J. Exp. Psychol. Gen*. 2013. V. 142. P. 1074–1083.
- Kelly D.** *Yuck! The nature and moral significance of disgust*. Cambridge: MIT Press, 2011.
- Kemmer P.B., Guo Y., Wang Y., Pagnoni G.** Network-based characterization of brain functional connectivity in zen practitioners // *Front. Psychol*. 2015. V. 6. P. 603.

- Kennett J., Gerrans P.** The rationalist delusion? // Liao S.M. (ed.) *Moral brains: the neuroscience of morality*. Oxford: Oxford Scholarship Online, 2016.
- Kent A.** Quanta and qualia // *Found. Phys.* 2018. V. 48. P. 1021–1037.
- Keysers C.** The empathic brain: How the discovery of mirror neurons changes our understanding of human nature. N.Y.: Social Brain Press, 2011.
- Kiefer M., Sim E.-J., Liebich S., Hauk O., Tanaka J.** Experience-dependent plasticity of conceptual representations in human sensory-motor areas // *J. Cogn. Neurosci.* 2007. V. 19. P. 525–542.
- Kim H., Adolphs R., O'Doherty J.P., Shimojo S.** Temporal isolation of neural processes underlying face preference decisions // *PNAS.* 2007. V. 104, N 46. P. 18253–18258.
- Kinsbourne M.** Integrated field theory of consciousness // Marcel A., Bisioch E. (eds.) *Consciousness in contemporary science*. Oxford: Oxford Univ. Press. 1988.
- Kirk U., Skov M., Hulme O., Christensen M.S., Zeki S.** Modulation of aesthetic value by semantic context: an fMRI study // *NeuroImage.* 2009. V. 44. P. 1125–1132.
- Kitcher P.** *The ethical project*. Cambridge, MA: Harvard Univ. Press, 2011.
- Klein S.** *Survival of the nicest*. N.Y.: The Experiment, 2014.
- Klein C.** The dual track theory of moral decision-making: A critique of the neuroimaging evidence // *Neuroethics.* 2010. V. 4. P. 143–162.
- Kluger J.** Inside the minds of animals. *TIME.* 2010. URL: <http://content.time.com/time/magazine/article/0,9171,2008867,00.html>
- Knierim J., Neunuebel J.** Tracking the flow of hippocampal computation: Pattern separation, pattern completion, and attractor dynamics // *Neurobiol. Learn. Mem.* 2016. V. 129. P. 38–49.
- Knoch D., Pascual-Leon A., Meyer K., Treye V., Feh E.** Diminishing reciprocal fairness by disrupting the right prefrontal cortex // *Science.* 2006. V. 314. P. 829–832.
- Knudsen E.I.** Fundamental components of attention // *Ann. Rev. Neurosci.* 2007. V. 30. P. 57–78.
- Knyazev G.G., Slobodskoj-Plusnin J.Y., Bocharov A.V.** Subject's state modulates oscillatory responses to emotional facial expressions // *J. Psychophysiol.* 2012. V. 26. P. 83–91.
- Koch C.** *Biophysics of computation: Information processing in single neurons*. Oxford: Oxford Univ. Press, 1999.
- Koch C.** *Consciousness: Confessions of a romantic reductionist*. Cambridge, MA: MIT Press, 2012.
- Koch C., Tsuchiya N.** Attention and consciousness: Two distinct brain processes // *Trends Cogn. Sci.* 2007. V. 11. P. 16–22.
- Kochiyama T., Ogihara N., Tanabe H.C.** Reconstructing the Neanderthal brain using computational anatomy // *Sci. Rep.* 2018. V. 8. P. 6296.

- Koenig H.G., Cohen H.J.** The link between religion and health: psychoneuroimmunology and the faith factor. Oxford: Oxford Univ. Press, 2001.
- Koenigs M., Barbey A.K., Postle B.R., Grafman J.** Superior parietal cortex is critical for the manipulation of information in working memory // *J. Neurosci.* 2009. V. 29. P. 14980–14986.
- Kogo N., Trengove C.** Is predictive coding theory articulated enough to be testable? // *Front. Comput. Neurosci.* 2015. V. 9. P. 111.
- Koriat A., Goldsmith M., Pansky A.** Toward a psychology of memory accuracy // *Ann. Rev. Psychol.* 2000. V. 51. P. 481–537.
- Korsgaard C.** Morality and the distinctiveness of human action // De Waal F. (ed.) *Primates and philosophers: how morality evolved.* New Jersey: Princeton Univ. Press, 2006. P. 98–119.
- Kosfeld M., Heinrichs M., Zak P.J., Fischbacher U., Fehr E.** Oxytocin increases trust in humans // *Nature.* 2005. V. 435. P. 673–676.
- Kounios J., Jung-Beeman M.** The Aha! Moment the cognitive neuroscience of insight // *Curr. Dir. Psychol. Sci.* 2009. V. 18. P. 210–216.
- Kounios J., Fleck J., Green D., Payne L., Stevenson J., Bowden E., Jung-Beeman M.** The origins of insight in resting-state brain activity // *Neuropsychologia.* 2008. V. 46. P. 281–291.
- Kounios J., Frymiare J., Bowden E., Fleck J., Subramaniam K., Parrish T., Jung-Beeman M.** The prepared mind: Neural activity prior to problem presentation predicts subsequent solution by sudden insight // *Psychol. Sci.* 2006. V. 17. P. 882–891.
- Kramer S., Zebrowitz L., Giovanni J.P.S., Sherak B.** Infants' preferences for attractiveness and babyfacedness. *Studies in perception and action III.* Abingdon: Routledge, 2019. P. 389–392.
- Kraut R.** Plato // Zalta E.N. (ed.) *The Stanford encyclopedia of philosophy.* 2017. URL: <https://plato.stanford.edu/archives/fall2017/entries/plato/>.
- Kravitz D., Behrmann M.** Space-, object-, and feature-based attention interact to organize visual scenes // *Atten. Percept. Psychophys.* 2011. V. 73. P. 2434–2447.
- Kravitz D.J., Saleem K.S., Baker C.I., Mishkin M.** A new neural framework for visuospatial processing // *Nat. Rev. Neurosci.* 2011. V. 12. P. 217–230.
- Kreiman G., Hung C.P., Kraskov A., Quian Quiroga R., Poggio T., DiCarlo J.J.** Object selectivity of local field potentials and spikes in the macaque inferior temporal cortex // *Neuron.* 2006. V. 49. P. 433–445.
- Krolak-Salmon P., Henaff M.A., Isnard J., Tallon-Baudry C., Guenet M., Vighetto A. et al.** An attention modulated response to disgust in human ventral anterior insula // *Ann. Neurol.* 2003. V. 53. P. 446–453.
- Krotov D., Hopfield J.** Unsupervised learning by competing hidden units // *PNAS.* 2019. V. 116. P. 7723–7731.

- Kühn S., Gallinat J.** The neural correlates of subjective pleasantness // *NeuroImage*. 2012. V. 61. P. 289–294.
- Kumaran D. et al.** What learning systems do intelligent agents need? Complementary learning systems theory updated // *Trends Cogn. Sci.* 2016. V. 20. P. 512–534.
- Kwantes P.J.** Using context to build semantics // *Psy. Bull. Rev.* 2005. V. 12. P. 703–710.
- Kwisthout J., van Rooij I.** Computational resource demands of a predictive bayesian brain // *Comput. Brain Behav.* 2020. V. 3. P. 174–188.
- Lacey S., Hagtvéd H., Patrick V.M., Anderson A., Stilla R., Deshpande G. et al.** Art for reward's sake: visual art recruits the ventral striatum // *NeuroImage*. 2011. V. 55. P. 420–433.
- Ladd K.L., Spilka B.** Inward, outward, and upwards: Cognitive aspects of prayer // *J. Sci. Stud. Relig.* 2002. V. 41. P. 475–484.
- Lamme V.** Toward a true neural stance on consciousness // *Trends Cogn. Sci.* 2006. V. 10. P. 494–501.
- Lamme V.** How neuroscience will change our view on consciousness // *Cogn. Neurosci.* 2010. V. 1. P. 204–220.
- Landau M.J., Greenberg J., Solomon S., Pyszczynski T., Martens A.** Windows into nothingness: Terror management, meaninglessness, and negative reactions to modern art // *J. Pers. Soc. Psychol.* 2006. V. 90. P. 879–892.
- Landau M.J., Solomon S., Pyszczynski T., Greenberg J.** On the compatibility of terror management theory and perspectives on human evolution // *Evol. Psychol.* 2007. V. 5. P. 476–519.
- Lang S., Kanngieser N., Jaskowski P., Haider H., Rose M., Verleger R.** Precursors of insight in event-related brain potentials // *J. Cogn. Neurosci.* 2006. V. 18. P. 2152–2166.
- Larson E.J., Witham L.** Leading scientists still reject God // *Nature*. 1998. V. 394, N 6691. P. 313–313.
- Latimer B., Ohman J.** Axial dysplasia in *Homo erectus* // *J. Hum. Evol.* 2001. V. 40. P. A12.
- Lau H.C., Passingham R.E.** Relative blindsight in normal observers and the neural correlate of visual consciousness // *PNAS*. 2006. V. 103, N 49. P. 18763–18768.
- Lau H., Rosenthal D.** Empirical support for higher-order theories of conscious awareness // *Trends Cogn. Sci.* 2011. V. 15. P. 365–373.
- Lavenex P., Amaral D.G.** Hippocampal-Neocortical interaction: A hierarchy of associativity // *Hippocampus*. 2000. V. 10. P. 420–430.
- Lavric A., Forstmeier S., Rippon G.** Differences in working memory involvement in analytical and creative tasks: an ERP study // *NeuroReport*. 2000. V. 11. P. 1613–1618.
- Lazarus R.S.** *Emotion and adaptation*. N.Y.: Oxford Univ. Press, 1991.

- Leakey R., Lewin R.** Origins reconsidered. N.Y.: Doubleday, 1992.
- Leal S., Yassa M.** Integrating new findings and examining clinical applications of pattern separation // *Nat. Neurosci.* 2018. V. 21. P. 163–173.
- LeCun Y., Bengio Y., Hinton G.** Deep learning // *Nature.* 2015. V. 521. P. 436–444.
- LeDoux J.E.** Semantics, surplus meaning, and the science of fear // *Trends Cogn. Sci.* 2017. V. 21. P. 303–306.
- Lee J., Maunsell J.H.R.** A normalization model of attentional modulation of single unit responses // *PLoS ONE.* 2009. V. 4. P. e4651.
- Legare C.H., Evans E.M., Rosengren K.S., Harris P.L.** The coexistence of natural and supernatural explanations across cultures and development // *Child Dev.* 2012. V. 83. P. 779–793.
- Leiter B.** Nietzsche's theory of the will // *Philos. Impr.* 2007. V. 7. P. 1–15.
- Lerner J.S., Goldberg J.H., Tetlock P.E.** Sober second thought: The effects of accountability, anger, and authoritarianism on attributions of responsibility // *Pers. Soc. Psychol. Bull.* 1998. V. 24. P. 563–574.
- Levine J.** Out of the closet: a qualophile confronts qualophobia // *Philos. Top.* 1994. V. 22. P. 107–126.
- Lewes G.H.** Problems of life and mind. L.: Kegan Paul, Trench, Turbner, and Co, 1875. V. 2.
- Li R., Johansen J.S., Ahmed H., Ilyevsky T.V., Wilbur R.B., Bharadwaj H.M., Siskind J.M.** The perils and pitfalls of block design for EEG classification experiments // *IEEE Trans. Pattern Anal. Mach. Intell.* 2021. V. 43. P. 316–333.
- Li H.H., Sprague T.C., Yoo A.H., Ma W.J., Curtis C.E.** Joint representation of working memory and uncertainty in human cortex // *Neuron.* 2021a. V. 109, N 22. P. 3699–3712.
- Lieberman P.** The unpredictable species: what makes humans unique. Princeton, N.J.: Princeton Univ. Press, 2013.
- Libet B.** Unconscious cerebral initiative and the role of conscious will in voluntary action // *Beh. Brain Sci.* 1985. V. 8. P. 529–566.
- Libet B.** Do we have free will? // Kane R. (ed.) *Oxford handbook of free will.* N.Y.: Oxford Univ. Press, 2002. P. 551–564.
- Lindquist K.A., Wager T.D., Kober H., Bliss-Moreau E., Barrett L.F.** The brain basis of emotion: A meta-analytic review // *Beh. Brain Sci.* 2012. V. 35. P. 121–202.
- Lindell A.K., Mueller J.** Can science account for taste? Psychological insights into art appreciation // *J. Cogn. Psychol.* 2011. V. 23. P. 453–475.
- Lipps T.** *Aesthetik.* Hamburg: Voss Verlag, 1905. V. 2.
- Lipps T.** *Das Wissen von Fremden Ichen* // *Psychologische Untersuchungen.* 1907. V. 1. P. 694–722.
- Llinas R.** *I of the vortex: From neurons to self.* Cambridge, MA: MIT Press, 2001.

- Lloyd S.** Programming the universe. N.Y.: Vintage, 2007.
- Logothetis N.K., Leopold D.A., Sheinberg D.L.** What is rivalling during binocular rivalry? // *Nature*. 1996. V. 380, N 6575. P. 621–624.
- London F., Bauer E.** La théorie de l'observation en mécanique quantique. Paris: Hermann, 1939. English translation, The theory of observation in quantum mechanics // Wheeler J.A., Zurek W.H. (eds.) Quantum theory and measurement. Princeton: Princeton Univ. Press, 1983. P. 217–259.
- Lovejoy C.O.** Reexamining human origins in light of *Ardipithecus ramidus* // *Science*. 2009. V. 326, N 5949. P. 74–74e8.
- Lucas J.R.** Minds, machines, and Gödel // *Philosophy*. 1961. V. 36. P. 112–137.
- Machery E., Mallon R.** Evolution of morality // Doris J. (ed.) The moral psychology handbook. Oxford: Oxford Univ. Press, 2010.
- Mackie J.L.** Ethics: Inventing right and wrong. L.: Penguin Books, 1977.
- MacLean P.D.** Some psychiatric implications of physiological studies on frontotemporal portion of limbic system (visceral brain) // *EEG Clin. Neurophysiol.* 1952. V. 4. P. 407–418.
- MacLean P.D.** The triune brain in evolution. N.Y.: Plenum Press, 1990.
- Mai X., Luo J., Wu J., Luo Y.** «Aha!» effects in a guessing riddle task: An event related potential study // *Hum. Brain Mapp.* 2004. V. 22. P. 261–270.
- Maibom H.L.** Empathy and morality. Oxford: Oxford Scholarship Online, 2014.
- Malinowski B.** Magic, science, and religion and other essays. Prospect Heights, IL: Waveland Press, 1992.
- Manktelow K.I.** Thinking and reasoning. Hove: Psychology Press, 2012.
- Manza L., Reber A.S.** Representing artificial grammars: transfer across stimulus forms and modalities // Berry D.C. (ed.) How implicit is implicit learning? Oxford: Oxford Univ. Press, 1997. P. 73–106.
- Marblestone A., Wayne G., Kording K.** Toward an integration of deep learning and neuroscience // *Front. Comput. Neurosci.* 2016. V. 10. P. 1–41.
- Martin A.** The representation of object concepts in the brain // *Ann. Rev. Psychol.* 2007. V. 58. P. 25–45.
- Masci D., Smith G.A.** Is God dead? No, but belief has declined slightly // Pew Research Center report, 2016. URL: <https://www.pewresearch.org/fact-tank/2016/04/07/is-god-dead-no-but-belief-has-declined-slightly/>
- Mayr E.** Cold Spring Harbor Symposium // *Quant. Biol.* 1950. V. 25. P. 109
- McCulloch W., Pitts W.** A logical calculus of the ideas immanent in nervous activity // *Bull. Math. Biol.* 1943. V. 7. P. 115–133.
- McFadden J.** Integrating information in the brain's EM field: The cemi field theory of consciousness // *Neurosci. Conscious.* 2020. V. 6. P. niaa016.
- McClelland J.L., Rumelhart D.E.** An interactive activation model of context effects in letter perception: I. An account of basic findings // *Psychol. Rev.* 1981. V. 88. P. 375–407.

- Mechelli A., Price C.J., Friston K.J., Ishai A.** Where bottom-up meets top-down: Neuronal interactions during perception and imagery // *Cereb. Cortex*. 2004. V. 14. P. 1256–1265.
- Mednick S.A., Mednick M.T.** Remote associates test. College and adult form. USA: ERIC, 1967.
- Mele A.R.** Free will and luck. Oxford: Oxford Univ. Press, 2006.
- Mele A.R.** Effective intentions: The power of conscious will. Oxford: Oxford Univ. Press, 2009.
- Mele A.** Unconscious decisions and free will // *Philos. Psychol.* 2013. V. 26. P. 777–789.
- Meltzoff A., Brooks R.** ‘Like me’ as a building block for understanding other minds: Bodily acts, attention, and intention // Malle B., Moses L., Baldwin D. (eds.) *Intentions and intentionality*. Cambridge, MA: MIT Press, 2001. P. 171–191.
- Mesquita B., Frijda N.H.** Cultural variations in emotions: A review // *Psychol. Bull.* 1992. V. 112. P. 179–204.
- Metzinger T.** Neural correlates of consciousness: Empirical and conceptual questions. Cambridge, MA: MIT Press, 2000.
- Mikhail J.** Chomsky and moral philosophy // McGilvray J. (ed.) *The Cambridge companion to chomsky*. Cambridge: Cambridge Univ. Press, 2005. P. 235–253.
- Miller J.D.** The mating mind: how sexual choice shaped the evolution of human nature. Portsmouth: Heinemann, 2000.
- Milner A.D., Goodale M.A.** The visual brain in action. Oxford: Oxford Univ. Press, 1995.
- Mitchell T.M., Shinkareva S., Carlson A., Chang K.M., Malave V.L., Mason R.A., Just M.A.** Predicting human brain activity associated with the meanings of nouns // *Science*. 2008. V. 320. P. 1191–1195.
- Mole C.** Attention // Zalta E.N. (ed.) *The Stanford encyclopedia of philosophy*, 2017. URL: <https://plato.stanford.edu/archives/fall2017/entries/attention/>.
- Moles A.** Théorie de l’information et perception esthétique. (Information theory and aesthetical perception). Paris: Denoël, 1973.
- Momennejad I., Haynes J.D.** Encoding of prospective tasks in the human prefrontal cortex under varying task loads // *J. Neurosci.* 2013. V. 33. P. 17342–17349.
- Moore G.E.** Principia ethica. N.Y.: Cambridge Univ. Press, 1903.
- Moore D.W.** Measuring new types of question-order effects // *Public Opin. Q.* 2002. V. 66. P. 80–91.
- Moors A.** Integration of two skeptical emotion theories: Dimensional appraisal theory and Russell’s psychological construction theory // *Psychol. Inq.* 2017. V. 28. P. 1–19.
- Morais F.B.** Vision and the Nobel prize // *Arq. Bras. Oftalmol.* 2017. V. 81. P. 161–165.

- Morales J., Solovey G., Maniscalco B., Rahnev D., de Lange F., Lau H.** Low attention impairs optimal incorporation of prior knowledge in perceptual decisions // *Atten. Percept. Psychophys.* 2015. V. 77. P. 2021–2036.
- Morrison J.** Perceptual confidence and categorization // *Anal. Philos.* 2017. V. 58. P. 71–85.
- Moscovitch M. et al.** Episodic memory and beyond: The hippocampus and neocortex in transformation // *Annu. Rev. Psychol.* 2016. V. 67. P. 105–134.
- Moser E. et al.** Spatial representation in the hippocampal formation: A history // *Nat. Neurosci.* 2017. V. 20. P. 1448–1464.
- Murphy K.** *Machine learning: A probabilistic perspective.* Cambridge, MA: MIT Press, 2012.
- Murphy M.** Theological voluntarism // Zalta E.N. (ed.) *The Stanford encyclopedia of philosophy.* 2019. URL: <https://plato.stanford.edu/archives/sum2019/entries/voluntarism-theological/>.
- Nagel T.** What is it like to be a bat? // *Philos. Rev.* 1974. V. 83. P. 435–450.
- Nagel T.** *Panpsychism* // Nagel T. (ed.) *Mortal questions.* Cambridge: Cambridge Univ. Press, 1979a. P. 181–195.
- Nagel T.** *Ethics without biology* // Nagel T. (ed.) *Mortal questions.* Cambridge: Cambridge Univ. Press, 1979b. P. 142–146.
- Nagel E., Newman J.R.** *Gödel's proof.* N.Y.: N.Y. Univ. Press, 1958.
- Nahmias E.** Is free will an illusion? Confronting challenges from the modern mind sciences // Sinnott-Armstrong W. (ed.) *Moral psychology.* Cambridge, MA: MIT Press, 2014. V. 4: Free will and moral responsibility. P. 1–25.
- Nakahashi W.** Evolution of improvement and cumulative culture // *Theor. Popul. Biol.* 2013. V. 83. P. 30–38.
- Nalbantian S.** Neuroaesthetics: Neuroscientific theory and illustration from the arts // *Interdiscip. Sci. Rev.* 2008. V. 33. P. 357–368.
- Naselaris T., Kay K.N., Nishimoto S., Gallant J.L.** Encoding and decoding in fMRI // *Neuroimage.* 2011. V. 56. P. 400–410.
- Naselaris T., Prenger R.J., Kay K.N., Oliver M., Gallant J.L.** Bayesian reconstruction of natural images from human brain activity // *Neuron.* 2009. V. 63. P. 902–915.
- Natsoulas T.** Concepts of consciousness // *J. Mind Behav.* 1983. V. 4. P. 13–59.
- Navajas J., Ahmadi M., Quiñero R.** Uncovering the mechanisms of conscious face perception: A single-trial study of the N170 responses // *J. Neurosci.* 2013. V. 33. P. 1337–1343.
- Nehamas A.** *Only a promise of happiness: the place of beauty in a world of art.* Princeton: Princeton Univ. Press, 2007.
- Neisser U.** *Cognition and reality.* San Francisco: Freeman, 1976.

- Nenciovici L., Allaire-Duquette G., Masson S.** Brain activations associated with scientific reasoning: A literature review // *Cogn. Process.* 2019. V. 20. P. 139–161.
- Neubauer R.L.** Prayer as an interpersonal relationship: A neuroimaging study // *Relig. Brain Behav.* 2014. V. 4. P. 92–103.
- Newberg A.B., D’Aquili E.G., Rause V.** *Why God won’t go away* // *Brain Science and the Biology of Belief*. N.Y. City: Ballantine Books, 2002.
- Ney A.** Convergence on the problem of mental causation: shoemaker’s strategy for (nonreductive?) physicalists // *Philos. Iss.* 2010. V. 20. P. 438–445.
- Ngo T.T., Miller S.M., Liu G.B., Pettigrew J.D.** Binocular rivalry and perceptual coherence // *Curr. Biol.* 2000. V. 10. P. R134–R136.
- Nicki R.M., Moss V.** Preference for non-representational art as a function of various measures of complexity // *Can. J. Psychol.* 1975. V. 29. P. 237–249.
- Nikolić D., Singer W.** Creation of visual long-term memory // *Percept. Psychophys.* 2007. V. 69. P. 904–912.
- Norenzayan A.** *Big Gods: How religion transformed cooperation and conflict*. Princeton, N.J.: Princeton Univ. Press, 2013.
- Noroozian M.** The role of the cerebellum in cognition: Beyond coordination in the central nervous system // *Neurol. Clin.* 2014. V. 32. P. 1081–104.
- Northoff G., Heinzl A., de Greck M., BERPohl F., Dobrowolny H., Panksepp J.** Self-referential processing in our brain—a meta-analysis of imaging studies on the self // *NeuroImage*. 2006. V. 31. P. 440–457.
- Nussbaum M.C.** *Upheavals of thought: The intelligence of emotions*. Cambridge: Cambridge Univ. Press, 2001.
- Oaksford M.** Imaging deductive reasoning and the new paradigm // *Front. Hum. Neurosci.* 2015. V. 8. P. 1061.
- Oaksford M., Chater N.** *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford: Oxford Univ. Press, 2007.
- Oaksford M., Chater N.** Precise of bayesian rationality: The probabilistic approach to human reasoning // *Behav. Brain Sci.* 2009. V. 32. P. 69–120.
- Oatley K., Jenkins J.M.** *Understanding emotions*. Malden, MA: Blackwell Publishers, 1996.
- O’Connor T.** Degrees of freedom // *Philos. Explor.* 2009. V. 12. P. 119–125.
- O’Connor T.** Conscious willing and the emerging sciences of brain and behavior // Ellis G.F.R., Murphy N., O’Connor T. (eds.) *Downward causation and the neurobiology of free will*. N.Y.: Springer Publications, 2009b. P. 173–186.
- O’Connor T.** Emergent properties // Zalta E.N. (ed.) *The Stanford encyclopedia of philosophy*. 2020. URL: <https://plato.stanford.edu/archives/fall2020/entries/properties-emergent/>.

- O'Connor T., Franklin C.** Free will // Zalta E.N. (ed.) The Stanford encyclopedia of philosophy. 2020. URL: <https://plato.stanford.edu/archives/fall2020/entries/freewill/>.
- O'Connor D.H., Fukui M.M., Pinsk M.A., Kastner S.** Attention modulates responses in the human lateral geniculate nucleus // *Nat. Neurosci.* 2002. V. 5. P. 1203–1209.
- O'Connor T., Churchill J.R.** Is non-reductive physicalism viable within a causal powers metaphysic? // Macdonald C., Macdonald G. (eds.) *Emergence in mind*. Oxford: Oxford Univ. Press, 2010. P. 43–60.
- Odegaard B., Knight R.T., Lau H.** Should a few null findings falsify prefrontal theories of conscious perception? // *J. Neurosci.* 2017. V. 37, N 40. P. 9593–9602.
- Okasha S.** Biological Altruism // Zalta E.N. (ed.) The Stanford encyclopedia of philosophy. 2020. URL: <https://plato.stanford.edu/archives/sum2020/entries/altruism-biological/>.
- Olds J., Milner P.** Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain // *J. Comp. Physiol. Psychol.* 1954. V. 47. P. 419–427.
- Orlandi N.** Bayesian perception is ecological perception // *Philos. Top.* 2016. V. 44. P. 327–351.
- Ossorio P.N.** Chapter 9: myth and mystification: The science of race and IQ // Krimsky S., Sloan K. (eds.) *Race and the genetic revolution: Science, myth, and culture*. N.Y. City: Columbia Univ. Press, 2011. P. 189.
- Oswald F.L., Mitchell G., Blanton H., Jaccard J., Tetlock P.E.** Predicting ethnic and racial discrimination: A meta-analysis of IAT criterion studies // *J. Pers. Soc. Psychol.* 2013. V. 105. P. 171–192.
- Oxnard C.E.** The place of the australopithecines in human evolution: Grounds for doubt? // *Nature*. 1975. V. 258. P. 389–395.
- Palacio N., Cardenas F.** A systematic review of brain functional connectivity patterns involved in episodic and semantic memory // *Rev. Neurosci.* 2019. V. 30. P. 889–902.
- Panichello M.F., Buschman T.J.** Shared mechanisms underlie the control of working memory and attention // *Nature*. 2021. V. 592, N 7855. P. 607–605.
- Panksepp J.** *Affective neuroscience: The foundation of human and animal emotions*. N.Y.: Oxford Univ. Press, 1998.
- Panksepp J.** Foreword // Cory G., Gardner R. (eds.) *The evolutionary neuroethology of Paul MacLean: Convergences and frontiers*. Westport, Connecticut: Greenwood Publ. Group, 2002.
- Panksepp J.** Affective consciousness: Core emotional feelings in animals and humans // *Conscious. Cogn.* 2005. V. 14. P. 30–80.
- Papez J.W.** A proposed mechanism of emotion // *J. Neuropsychiatry Clin. Neurosci.* 1995. V. 7. P. 103–112.

- Papo D.** How can we study reasoning in the brain? // *Front. Hum. Neurosci.* 2015. V. 9. P. 222.
- Parkin A.J.** The central executive does not exist // *J. Internat. Neuropsychol. Soc.* 1998. V. 4. P. 518–522.
- Patterson K. et al.** Where do you know what you know? The representation of semantic knowledge in the human brain // *Nat. Rev. Neurosci.* 2007. V. 8. P. 976–987.
- Pattyn N., Neyt X., Henderickx D., Soetens E.** Psychophysiological investigation of vigilance decrement: Boredom or cognitive fatigue? // *Physiol. Behav.* 2008. V. 93. P. 369–378.
- Paul E., Harding E., Mendl M.** Measuring emotional processes in animals: The utility of a cognitive approach // *Neurosci. Biobehav. Rev.* 2005. V. 29. P. 469–491.
- Pedreira C. et al.** Responses of human medial temporal lobe neurons are modulated by stimulus repetition // *J. Neurophysiol.* 2010. V. 103. P. 97–107.
- Peirce C.S., Jastrow J.** On small differences in sensation // *Biogr. Mem. Natl. Acad. Sci.* 1884. V. 3. P. 73–83.
- Penfield W., Faulk M.E.** The insula: Further observations on its function // *Brain.* 1955. V. 78. P. 445–470.
- Penrose R.** The emperor's new mind: Concerning computers, minds and the laws of physics. Oxford UK: Oxford Univ. Press, 1989.
- Pereira F., Lou B., Pritchett B., Ritter S., Gershman S.J., Botvinick M., Fedorenko E.** Toward a universal decoder of linguistic meaning from brain activation // *Nat. Commun.* 2018. V. 9. P. 1–13.
- Pessoa L.** The cognitive-emotional brain: From interactions to integration. Oxford: Oxford Univ. Press, 2013.
- Peters M.A.K. et al.** Perceptual confidence neglects decision-incongruent evidence in the brain // *Nat. Hum. Behav.* 2017. V. 1. P. 0139.
- Petersen S.E., Posner M.I.** The attention system of the human brain: 20 years after // *Annu. Rev. Neurosci.* 2012. V. 35. P. 73–89.
- Pfeifer R., Bongard J.** How the body shapes the way we think: A new view of intelligence. Cambridge, Massachusetts: MIT Press, 2006.
- Phillips D.Z.** Religion without explanation. Oxford: Blackwell, 1976.
- Piaget J.** Piaget's theory // Mussen P.H. (ed.) *Handbook of child psychology.* N.Y.: Wiley, 1983. P. 103–128. V. 1.
- Pizarro D., Bloom P.** The intelligence of the moral intuitions: Comments on Haidt (2001) // *Psychol. Rev.* 2003. V. 110. P. 193–196.
- Plate J.** An analysis of the binding problem // *Philos. Psychol.* 2007. V. 20. P. 773–792.
- Plato.** *Collected dialogues* // Hamilton E., Huntington C. (eds.) Princeton: Princeton Univ. Press, 1961.
- Plotinus.** *The six enneads.* McKenna S., Page B.S., trans. Chicago: Encyclopedia Britannica Publ., 1952.

- Pockett S.** The nature of consciousness: A hypothesis. Lincoln, NE: Writers Club Press, 2000.
- Poeppel D., Embick D.** Defining the relation between linguistics and neuroscience // Cutler A. (ed.) Twenty-first century psycholinguistics: Four cornerstones. Hillsdale, N.J.: Erlbaum, 2005. P. 103–118.
- Poincaré H.** The foundations of science. Lancaster, PA: Science Press, 1913.
- Polonsky A., Blake R., Braun J., Heeger D.J.** Neuronal activity in human primary visual cortex correlates with perception during binocular rivalry // Nat. Neurosci. 2000. V. 3. P. 1153–1159.
- Popper K.** Objective knowledge: An evolutionary approach. Oxford UK: Oxford Univ. Press, 1975.
- Posner M.I., Petersen S.E.** The attention system of the human brain // Annu. Rev. Neurosci. 1990. V. 13. P. 25–42.
- Postle B.R.** Working memory as an emergent property of the mind and brain // Neurosci. 2006. V. 139. P. 23–38.
- Potter M.C., Staub A., O'Connor D.H.** Pictorial and conceptual representation of glimpsed pictures // J. Exp. Psychol. Hum. Percept. Perform. 2004. V. 30. P. 478–489.
- Pozzi I., Bohté S., Roelfsema P.** A biologically plausible learning rule for deep learning in the brain // arXiv. 2018. 1811.01768.
- Prado J., Chadha A., Booth J.R.** The brain network for deductive reasoning: A quantitative meta-analysis of 28 neuroimaging studies // Cogn. Neurosci. 2011. V. 23. P. 3483–3497.
- Preston S., de Waal F.** Empathy: Its ultimate and proximate bases // Behav. Brain Sci. 2002. V. 25. P. 1–72.
- Price C.J.** The anatomy of language: a review of 100 fMRI studies published in 2009 // Ann. N.Y. Acad. Sci. 2010. V. 1191. P. 62–88.
- Prinz J.** Gut reactions: A perceptual theory of emotion. Oxford: Oxford Univ. Press, 2004.
- Prinz J.** The emotional construction of morals. Oxford: Clarendon Press, 2007.
- Prinz J.** Is empathy necessary for morality? // Coplan A., Goldie P. (eds.) Empathy: Philosophical and psychological perspectives. Oxford: Oxford Univ. Press, 2011. P. 211–229.
- Prinz J.** The conscious brain. Oxford: Oxford Univ. Press, 2012.
- Prinz J.** Sentimentalism and the moral brain // Liao S.M. (ed.) Moral brains: The neuroscience of morality. Oxford: Oxford Scholarship Online, 2016.
- Proulx T., Heine S.J., Vohs K.D.** When is the unfamiliar the uncanny? Meaning affirmation after exposure to absurdist literature, humor, and art // Pers. Soc. Psychol. Bull. 2010. V. 36. P. 817–829.
- Pseudo-Dionysius.** Works of Dionysius the Areopagite / Parker J. (trans.). L.: James Parker and Co., 1897.
- Purves D.** Principles of cognitive neuroscience. Sunderland, Mass.: Sinauer Associates, 2008.

- Putnam H.** Psychophysical predicates // Capitan W., Merrill D. (eds.) *Art, mind, and religion*. Reprinted in Putnam 1975 as «The nature of mental states». Pittsburgh: Univ. of Pittsburgh Press, 1967. P. 429–440.
- Pylyshyn Z.** *Computation and cognition*. Cambridge, MA: MIT Press, 1984.
- Qiu J., Luo Y., Wu Z., Zhang Q.** A further study of ERP effects of «insight» in a riddle guessing task // *Acta Psychol. Sin.* 2006. V. 38. P. 507–514.
- Qiu J., Li H., Yang D., Luo Y., Li Y., Wu Z., Zhang Q.** The neural basis of insight problem solving: An event-related potential study // *Brain Cogn.* 2008a. V. 68. P. 100–106.
- Qiu J., Li H., Yang D., Luo Y., Li Y., Wu Z., Zhang Q.** Spatiotemporal cortical activation underlies mental preparation for successful riddle solving: An event-related potential study // *Exp. Brain Res.* 2008b. V. 186. P. 629–634.
- Quian Quiroga R.** Concept cells: The building blocks of declarative memory functions // *Nat. Rev. Neurosci.* 2012. V. 13. P. 587–597.
- Quian Quiroga R.** Neuronal codes for visual perception and memory // *Neuropsychologia.* 2016. V. 83. P. 227–241.
- Quian Quiroga R.** No pattern separation in the human hippocampus // *Trends Cogn. Sci.* 2020. V. 24. P. 994–1007.
- Quian Quiroga R., Kraskov A., Koch C., Fried I.** Explicit encoding of multimodal percepts by single neurons in the human brain // *Curr. Biol.* 2009. V. 19. P. 1308–1313.
- Quian Quiroga R., Kraskov A., Mormann F., Fried I., Koch C.** Single-cell responses to face adaptation in the human medial temporal lobe // *Neuron.* 2014. V. 84. P. 363–369.
- Quian Quiroga R., Kreiman G.** Measuring sparseness in the brain: Comment on Bowers // *Psychol. Rev.* 2010. V. 117. P. 291–299.
- Quian Quiroga R., Mukamel R., Isham E.A., Malach R., Fried I.** Human single-neuron responses at the threshold of conscious recognition // *PNAS.* 2008. V. 105. P. 3599–3604.
- Quian Quiroga R., Reddy L., Koch C., Fried I.** Decoding visual inputs from multiple neurons in the human temporal lobe // *J. Neurophysiol.* 2007. V. 98. P. 1997–2007.
- Quian Quiroga R., Reddy L., Kreiman G., Koch C., Fried I.** Invariant visual representation by single neurons in the human brain // *Nature.* 2005. V. 435. P. 1102–1107.
- Quine W.V.O.** *Word and object*. Cambridge, MA: MIT Press, 1960.
- Rachels J.** *Created from animals: The moral implications of darwinism*. Oxford: Oxford Univ. Press, 1990.
- Raichle M.E.** The brain's default mode network // *Ann. Rev. Neurosci.* 2015. V. 38. P. 433–447.
- Raju T.N.** The Nobel chronicles. 1967: George Wald (1906–97); Ragnar A. Granit (1900–91); and Haldan Keffer Hartline (1903–83) // *Lancet.* 1999. V. 354, N 9178. P. 605.

- Ramachandran V.S., Hirstein W.** The science of art: A neurological theory of aesthetic experience // *J. Conscious. Stud.* 1999. V. 6. P. 15–51.
- Ramon y Cajal S.** The croonian lecture: la fine structure des centres nerveux // *Proc. R. Soc. Lond.* 1894. V. 5. P. 444–468.
- Ramsey W.** Eliminative materialism // Zalta E.N. (ed.) *The Stanford encyclopedia of philosophy.* 2020. URL: <https://plato.stanford.edu/archives/sum2020/entries/materialism-eliminative/>.
- Ramsey W., Stic S., Garon J.** Connectionism, eliminativism and the future of folk psychology // *Philos. Persp.* 1990. V. 4. P. 499–533.
- Ransom M., Fazelpour S., Mole C.** Attention in the predictive mind // *Conscious. Cogn.* 2017. V. 47. P. 99–112.
- Reber A.S.** The cognitive unconscious: an evolutionary perspective // *Conscious. Cogn.* 1992. V. 1. P. 93–133.
- Reber R., Schwarz N., Winkielman P.** Processing fluency and aesthetic pleasure: Is beauty in the perceiver’s processing experience? // *Pers. Soc. Psychol. Rev.* 2004. V. 8. P. 364–382.
- Reber R., Brun M., Mitterndorfer K.** The use of heuristics in intuitive mathematical judgment // *Psychon. Bull. Rev.* 2008. V. 15. P. 1174–1178.
- Reisenzein R.** Emotions as metarepresentational states of mind: Naturalizing the belief-desire theory of emotion // *Cogn. Syst. Res.* 2009. V. 10. P. 6–20.
- Rensink R.** Change detection // *Ann. Rev. Psychol.* 2002. V. 53. P. 4245–4277.
- Reppert V.** Eliminative materialism, cognitive suicide, and begging the question // *Metaphilos.* 1992. V. 23. P. 378–392.
- Rescorla M.** Bayesian perceptual psychology // Mohan M. (ed.) *The Oxford handbook of philosophy of perception.* Oxford: Oxford Univ. Press, 2015. P. 694–716.
- Rescorla M.** A realist perspective on Bayesian cognitive science // Chan T., Nes A. (eds.) *Inference and consciousness.* Abingdon: Routledge, 2019.
- Rey G.** A reason for doubting the existence of consciousness // Davidson R., Schwartz G., Shapiro D. (eds.) *Consciousness and self-regulation.* N.Y.: Plenum, 1983. P. 1–39. V. 3.
- Rey G.** A question about consciousness // Otto H., Tuedio J. (eds.) *Perspectives on mind.* Dordrecht: D. Reidel, 1988. P. 5–24.
- Rey H. et al.** Single neuron coding of identity in the human hippocampal formation // *Curr. Biol.* 2020. V. 30. P. 1152–1159.
- Rey H., Fried I., Quian Quiroga R.** Timing of single neuron and local field potential responses in the human medial temporal lobe // *Curr. Biol.* 2014. V. 24. P. 299–304.
- Reynolds J., Desimone R.** Competitive mechanisms subserve selective visual attention // Marantz A., Miyashita Y., O’Neil W. (eds.) *Image, language, brain: Papers from the first mind articulation project symposium.* Cambridge, MA: The MIT Press, 2000. P. 233–247.

- Rhodes G.** The evolutionary psychology of facial beauty // *Ann. Rev. Psychol.* 2006. V. 57. P. 199–226.
- Ricciardi L.M., Umezawa H.** Brain and physics of many-body problems // *Kybernetik.* 1967. V. 4. P. 44–48.
- Richards B.A., Frankland P.W.** The conjunctive trace // *Hippocampus.* 2013. V. 23. P. 207–212.
- Richter D. et al.** The age of the hominin fossils from Jebel Irhoud, Morocco, and the origins of the middle stone age // *Nature.* 2017. V. 546, N 7657. P. 293–296.
- Rifkin J.** The empathic civilization: the race to global consciousness in a world in crisis. N.Y.: Tarcher/Penguin, 2009.
- Rilling J.K., Gutman D.A., Zeh T.R., Pagnoni G., Berns G.S., Kilts C.D.** A neural basis for social cooperation // *Neuron.* 2002. V. 35. P. 395–405.
- Rios K., Cheng Z.H., Totton R.R., Shariff A.F.** Negative stereotypes cause christians to underperform in and disidentify with science // *Soc. Psychol. Pers. Sci.* 2015. V. 6. P. 959–967.
- Rips L.** The psychology of proof. Cambridge, MA: MIT Press, 1994.
- Rizzolatti G., Sinigaglia C.** The mirror mechanism: A basic principle of brain function // *Nat. Rev. Neurosci.* 2016. V. 17. P. 757–765.
- Robb D., Heil J.** Mental causation // Zalta E.N. (ed.) *The Stanford encyclopedia of philosophy.* 2021. URL: <<https://plato.stanford.edu/archives/spr2021/entries/mental-causation/>>.
- Roberts R.C.** Emotions: An essay in aid of moral psychology. N.Y.: Cambridge Univ. Press, 2003.
- Roberts M.J., Newton E.J.** Methods of thought: Individual differences in reasoning strategies. Hove, UK: Psychology Press, 2005.
- Robinson W.** Epiphenomenalism // Zalta E.N. (ed.) *The Stanford encyclopedia of philosophy,* 2019. URL: <https://plato.stanford.edu/entries/epiphenomenalism/#SelStu>.
- Robinson H.** Dualism // Zalta E.N. (ed.) *The Stanford encyclopedia of philosophy.* 2020. URL: <https://plato.stanford.edu/archives/fall2020/entries/dualism/>.
- Rolls E.T., Grabenhorst F.** The orbitofrontal cortex and beyond: from affect to decision-making // *Progr. Neurobiol.* 2008. V. 86. P. 216–244.
- Rosenberg G.** A place for consciousness: probing the deep structure of the natural world. N.Y.: Oxford Univ. Press, 2004.
- Rosenthal D.** Two concepts of consciousness // *Philos. Stud.* 1986. V. 49. P. 329–359.
- Rosenthal D.M.** Thinking that one thinks // Davies M., Humphreys G. (eds.) *Consciousness: psychological and philosophical essays.* Oxford: Blackwell, 1993.
- Roser M.E., Gazzaniga M.S.** The interpreter in human psychology // Preuss T.M., Kaas J.H. (eds.) *The evolution of primate nervous systems.* Oxford, UK: Academic Press, 2006. P. 503–508.

- Roskies A.** Can neuroscience resolve issues about free will? // Sinnott-Armstrong W. (ed.) *Moral psychology*. Cambridge, MA: MIT Press, 2014. V. 4: Free will and moral responsibility. P. 103–126.
- Ross D.** Game theory // Zalta E.N. (ed.) *The Stanford encyclopedia of philosophy*. 2019. URL: <https://plato.stanford.edu/archives/win2019/entries/game-theory/>.
- Roughgarden J.** *Evolution's rainbow: Diversity, gender and sexuality in nature and people*. Los Angeles: Univ. of California Press, 2004.
- Ruch S., Markes O., Duss B.S., Oppliger D.R., Koenig T., Mathis J., Roth C., Henke K.** Sleep stage II contributes to the consolidation of declarative memories // *Neuropsychologia*. 2012. V. 50. P. 2389–2396.
- Rudner M., Fransson P., Ingvar M., Nyberg L., Rönnerberg J.** Neural representation of binding lexical signs and words in the episodic buffer of working memory // *Neuropsychologia*. 2007. V. 45. P. 2258–2276.
- Rumelhart D., Hinton G., Williams R.** Learning representations by back-propagating errors // *Nature*. 1986a. V. 323. P. 533–536.
- Rumelhart D., McClelland J.** The PDP Research Group. *Parallel distributed processing*. Cambridge: MIT Press, 1986b. V. 1.
- Russell B.** *The analysis of matter*. L.: Kegan Paul, 1927.
- Russell J.A.** A circumplex model of affect // *J. Pers. Soc. Psychol.* 1980. V. 39. P. 1161–1178.
- Russell J.A.** Is there universal recognition of emotion from facial expressions? A review of the cross-cultural studies // *Psychol. Bull.* 1994. V. 115. P. 102–141.
- Russell J.A.** Core affect and the psychological construction of emotion // *Psychol. Rev.* 2003. V. 110. P. 145–172.
- Ryglu R., Pluta H., Popik P.** Laughing rats are optimistic // *PLOS ONE*. 2012. V. 7. P. e51959.
- Sadato N., Nakamura S., Oohashi T., Nishina E., Fuwamoto Y., Waki A., Yonekura Y.** Neural networks for generation and suppression of alpha rhythm: A PET study // *NeuroReport*. 1998. V. 9. P. 893–897.
- Sakai K., Passingham R.T.** Prefrontal interactions reflect future task operations // *Nat. Neurosci.* 2003. V. 6. P. 75–81.
- Salmela M.** Can emotion be modelled on perception? // *Dialectica*. 2011. V. 65. P. 1–29.
- Sample I.** Oldest Homo sapiens bones ever found shake foundations of the human story // *The Guardian*. 2017. June 7.
- Sandkuhler S., Bhattacharya J.** Deconstructing insight: EEG correlates of insightful problem solving // *PLOS ONE*. 2008. V. 3. P. e1459.
- Sanfey A.G., Rilling J.K., Aronson J.A., Nystrom L.E., Cohen J.D.** The neural basis of economic decision-making in the ultimatum game // *Science*. 2003. V. 300, N 5626. P. 1755–1758.
- Santayana G.** *The sense of beauty*. N.Y.: Scribner's. 1896.

- Sartwell C.** Six names of beauty. N.Y.: Routledge, 2004.
- Sartwell C.** Beauty // Zalta E.N. (ed.) The Stanford encyclopedia of philosophy. 2017. URL: <https://plato.stanford.edu/archives/win2017/entries/beauty/>.
- Scarantino A.** The motivational theory of emotions // D'Arms J., Jacobson D. (eds.) Moral psychology and human agency. Oxford: Oxford Univ. Press, 2014. P. 156–185.
- Scarantino A.** Basic emotions, psychological construction and the problem of variability // Russell J., Barrett L. (eds.) The psychological construction of emotion. N.Y.: The Guilford Press, 2015. P. 334–376.
- Scarantino A., de Sousa R.** Emotion // Zalta E.N. (ed.) The Stanford encyclopedia of philosophy, 2021. URL: <https://plato.stanford.edu/archives/sum2021/entries/emotion/>.
- Schacter D.L.** Implicit knowledge: new perspectives on unconscious processes // PNAS. 1992. V. 89. P. 11113–11117.
- Schaich Borg J., Lieberman D., Kiehl K.A.** Infection, incest, and iniquity: investigating the neural correlates of disgust and morality // J. Cogn. Neurosci. 2008. V. 20. P. 1529–1546.
- Schenk T., McIntosh R.D.** Do we have independent visual streams for perception and action? // Cogn. Neurosci. 2010. V. 1. P. 52–78.
- Scherer K.R., Schorr A., Johnstone T. (eds.)** Appraisal processes in emotion: Theory, methods, research (series in affective science). Oxford: Oxford Univ. Press, 2001.
- Schiller F.** On the aesthetic education of man. N.Y.: Dover, 2004.
- Schlosberg H.** A scale for the judgment of facial expression // J. Exp. Psychol. 1941. V. 29. P. 497–510.
- Schmidhuber J.** Low-complexity art // Leonardo. 1997. V. 30. P. 97–103.
- Schmidhuber J.** Developmental robotics, optimal artificial curiosity, creativity, music, and the fine arts // Connect. Sci. 2006. V. 18. P. 173–187.
- Schmidhuber J.** Simple algorithmic principles of discovery, subjective beauty, selective attention, curiosity & creativity // Proc. 10th Intl. Conf. on Discovery Science. 2007. P. 26–38.
- Schmitt P., Sandner G., Karli P.** Escape and approach induced by brain stimulation: A parametric analysis // Behav. Brain Res. 1981. V. 2. P. 49–79.
- Schnall S., Haid J., Clore A.** Disgust as embodied moral judgment // Pers. Soc. Psychol. Bull. 2008. V. 34. P. 1096–1109.
- Schneirla T.C.** A theoretical consideration of the basis for approach-withdrawal adjustments in behavior. Psychol. Bull. 1939. V. 37. P. 501–502. Reprinted in Aronson L.R., Tobach E., Lehrman D.S., Rosenblatt J.S. (eds.) Selected writings of T.C. Schneirla. San Francisco: Freeman, 1972a. P. 295–296.
- Schneirla T.C.** An evolutionary and developmental theory of biphasic processes underlying approach and withdrawal // Jones M.R. (ed.) Nebraska

- symposium on motivation. Lincoln: Univ. of Nebraska Press, 1959. Reprinted in: Aronson L.R., Tobach E., Lehrman D.S., Rosenblatt J.S. (eds.) Selected writings of T. C. Schneirla. San Francisco: Freeman, 1972b. V. 7.
- Schoenemann P.T.** Evolution of the size and functional areas of the human brain // *Annu. Rev. Anthropol.* 2006. V. 35. P. 379–406.
- Schrenk F., Kullmer O., Bromage T.** Chapter 9: The earliest putative homo fossils // Henke W., Tattersall I. (eds.) *Handbook of Paleoanthropology*, 2007. P. 1611–1631.
- Schulz W.** Predictive reward signal of dopamine neurons // *J. Neurophysiol.* 1998. V. 80. P. 1–27.
- Schwitzgebel E.** *Perplexities of consciousness*. Cambridge, MA: MIT Press, 2011.
- Scruton R.** *Beauty*. Oxford: Oxford Univ. Press, 2009.
- Searle J.** Minds, brains and programs // *Behav. Brain Sci.* 1980. V. 3. P. 417–457.
- Searle J.** *Intentionality*. Cambridge; N.Y.: Cambridge Univ. Press, 1983.
- Searle J.** The mystery of consciousness. N.Y.: The N.Y. Review of Books, 1997.
- Searle J.** Can information theory explain consciousness? N.Y.: The N.Y. Review, 2013.
- Seeley W.W., Menon V., Schatzberg A.F., Keller J., Glover G.H., Kenna H., Reiss A.L., Greicius M.D.** Dissociable intrinsic connectivity networks for salience processing and executive control // *J. Neurosci.* 2007. V. 27. P. 2349–2356.
- Seife C.** *Decoding the universe*. L.: Penguin, 2007.
- Shanks D.R., St. John M.F.** Characteristics of dissociable human learning systems // *Behav. Brain Sci.* 1994. V. 17. P. 367–448.
- Shannon C.E.** A mathematical theory of communication // *Bell Labs Tech. J.* 1948. V. 27. P. 379–423.
- Sharot T., Velasquez C.M., Dolan R.J.** Do decisions shape preference? Evidence from blind choice // *Psychol. Sci.* 2010. V. 21. P. 1231–1235.
- Shea N.** Neural signalling of probabilistic vectors // *Philos. Sci.* 2014. V. 81. P. 902–913.
- Shields C.** Aristotle // Zalta E.N. (ed.) *The Stanford encyclopedia of philosophy*. 2020. URL: <https://plato.stanford.edu/archives/fall2020/entries/aristotle/>.
- Shipman P.** Those ears were made for walking // *New Sci.* 1994. V. 143. P. 26–29.
- Siegler R.** Strategic development // *Trends Cogn. Sci.* 1999. V. 3. P. 430–435.
- Siewert C.** *The significance of consciousness*. Princeton, N.J.: Princeton Univ. Press, 1998.

- Simonite T.** Facebook creates software that matches faces almost as well as you do. Cambridge, Massachusetts: MIT Technology Review, 2014. URL: <https://www.technologyreview.com/2014/03/17/13822/facebook-creates-software-that-matches-faces-almost-as-well-as-you-do/>
- Simpson G.G.** Classification and human evolution // Washburn S.L. (ed.). N.Y.: Routledge, 1963. P. 4.
- Singer W.** Neuronal synchrony: A versatile code for the definition of relations // *Neuron*. 1999. V. 24. P. 49–65.
- Singer P.** Ethics and intuitions // *J. Ethics*. 2005. V. 9. P. 331–352.
- Singer T., Seymour B., O’Doherty J., Kaube H., Dolan R.J., Frith C.D.** Empathy for pain involves the affective but not sensory components of pain // *Science*. 2004. V. 303. P. 1157–1162.
- Singer T., Seymour B., O’Doherty J., Stephan K.E., Dolan R.J., Frith C.D.** Empathic neural responses are modulated by the perceived fairness of others // *Nature*. 2006. V. 439. P. 466–469.
- Sircello G.** Love and beauty. Princeton, N.J.: Princeton Univ. Press, 1989.
- Slote M.** Moral sentimentalism. Oxford: Oxford Univ. Press, 2010.
- Small D.M., Gregory M.D., Mak Y.E., Gitelman D., Mesulam M.M., Parrish T.** Dissociation of neural representation of intensity and affective valuation in human gustation // *Neuron*. 2003. V. 39. P. 701–711.
- Smart J.J.C.** Materialism. L.: Encyclopedia Britannica, 2016. URL: <https://www.britannica.com/topic/materialism-philosophy>.
- Smith A.** The theory of moral sentiments. N.Y.: August M. Kelley Publishers, 1966.
- Sober E., Wilson D.S.** Unto others: The evolution and psychology of unselfish behavior. Cambridge, MA: Harvard Univ. Press, 1998.
- Solomon S., Pyszczynski T., Greenberg J.** The worm at the core: On the role of death in life. N.Y.: Random House, 2015.
- Soon C.S., Brass M., Heinze H.J., Haynes J.D.** Unconscious determinants of free decisions in the human brain // *Nat. Neurosci*. 2008. V. 11. P. 543–545.
- Soon C.S., He A.H., Bode S., Haynes J.D.** Predicting free choices for abstract intentions // *PNAS*. 2013. V. 110. P. 6217–6222.
- Sosis R., Alcorta C.** Signaling, solidarity, and the sacred: the evolution of religious behavior // *Evol. Anthropol*. 2003. V. 12. P. 264–274.
- Spaniol J., Madden D.J., Voss A.** A diffusion model analysis of adult age differences in episodic and semantic long-term memory retrieval // *J. Exp. Psychol.: Learn. Mem. Cogn*. 2006. V. 32. P. 101–117.
- Spencer H.** The data of ethics. N.Y.: Thomas Y. Crowell & Company, 1879.
- Spener M.** Calibrating introspection // *Philos. Iss*. 2015. V. 25. P. 300–321.
- Spooner W.A.** The golden rule // Hastings J. (ed.) *Encyclopedia of religion and ethics*. N.Y.: Charles Scribner’s Sons, 1914. V. 6. P. 310–312. Quoted

- in Rushworth M. Kidder, *How good people make tough choices: Resolving the dilemmas of ethical living*, N.Y.: Harper, 2003.
- Sprevak M.** Computation, individuation, and the received view on representation // *Stud. Hist. Philos. Sci.* 2010. V. 41. P. 260–270.
- Squire L. et al.** Memory consolidation // *Cold Spring Harb. Perspect. Biol.* 2015. V. 7. P. a021766.
- Stanovich K.E.** *Rationality and the reflective mind*. Oxford: Oxford Univ. Press, 2011.
- Stapp H.P.** *Mind, matter and quantum mechanics*, 3-rd edition. Berlin: Springer-Verlag, 2009.
- Stark R.** Atheism, faith, and the social scientific study of religion // *J. Contemp. Relig.* 1999. V. 14. P. 41–61.
- Steele R.S., Morawski J.G.** Implicit cognition and the social unconscious // *Theory Psychol.* 2002. V. 12. P. 37–54.
- Stephan A.** Emergentism, irreducibility, and downward causation // *Grazer Philosophische Studien.* 2002. V. 65. P. 77–93.
- Stevenson C.** The emotive meaning of ethical terms // *Mind.* 1937. V. 46. P. 14–31.
- Stewart I.** *Why beauty is truth: the history of symmetry*. N.Y.: Basic Books, 2007.
- Stowe L.A., Haverkort M., Zwarts F.** Rethinking the neurological basis of language // *Lingua.* 2005. V. 115. P. 997–1042.
- Strawson P.F.** Social morality and individual ideal // *Philos.* 1961. V. 36. P. 1–17.
- Strawson G.** *Mental reality*. Cambridge, Mass: MIT Press, Bradford Books, 1994a.
- Strawson G.** The impossibility of moral responsibility // *Philos. Stud.* 1994b. V. 75. P. 5–24.
- Strickberger M.W.** *Evolution* (3rd ed.). Sudbury, MA: Jones and Bartlett Publishers, 2000.
- Striedter G.F.** *Principles of brain evolution*. Sunderland, MA: Sinauer Associates, 2005.
- Stubenberg L.** Neutral Monism // Zalta E.N. (ed.) *The Stanford encyclopedia of philosophy*. 2018. URL: <https://plato.stanford.edu/archives/fall2018/entries/neutral-monism/>.
- Stueber K.** *Rediscovering empathy: agency, folk psychology, and the human sciences*. Cambridge, MA: MIT Press, 2006.
- Stueber K.** Empathy // Zalta E.N. (ed.) *The Stanford encyclopedia of philosophy*. 2019. URL: <https://plato.stanford.edu/archives/fall2019/entries/empathy/>.
- Swinburne R.** *The evolution of the soul*. Oxford: Oxford Univ. Press, 1986.
- Swinburne R.** *Mind, brain, and free will*. Oxford: Oxford Univ. Press, 2013.

- Swinney D.** Aphasia // Wilson R., Keil F. (eds.) The MIT encyclopedia of the cognitive sciences. Cambridge, MA: MIT Press, 1999. P. 31–32.
- Szpak P.** Evolution of the Australopithecines. Tree of life web project, 2007. URL: http://tolweb.org/treehouses/?treehouse_id=4438
- Taliaferro C.** Philosophy of religion // Zalta E.N. (ed.) The Stanford encyclopedia of philosophy. 2019. URL: <https://plato.stanford.edu/archives/fall2019/entries/philosophy-religion/>.
- Tappolet C.** Emotions, values, and agency. Oxford: Oxford Univ. Press, 2016.
- Taren A.A., Gianaros P.J., Greco C.M., Lindsay E.K., Fairgrieve A., Brown K.W., Rosen R.K., Ferris J.L., Julson E., Marsland A.L., Creswell J.D.** Mindfulness meditation training and executive control network resting state functional connectivity: a randomized controlled trial // Psychosom. Med. 2017. V. 79. P. 674–683.
- Tattersall I., Schwartz J.H.** Extinct humans. Boulder, CO: Westview Press, 2000.
- Tattersall I.** Human origins: Out of Africa // PNAS. 2009. V. 106, N 38. P. 16018–16021.
- Tettamanti M., Perani D.** The neurobiology of structure-dependency in natural language grammar // Faust M. (ed.) The handbook of the neuropsychology of language. Oxford, UK: Wiley-Blackwell, 2012. P. 229–251.
- Thakral P.P., Moo L.R., Slotnick S.D.** A neural mechanism for aesthetic experience // NeuroReport. 2012. V. 23. P. 310–313.
- Thompson E.** Mind in life: Biology, phenomenology, and the sciences of mind. Cambridge, MA: Harvard Univ. Press, 2007.
- Thompson V.A., Prowse Turner J.A., Pennycook G.** Intuition, reason and metacognition // Cogn. Psychol. 2011. V. 63. P. 107–140.
- Tipper S.P., Driver J.** Negative priming between pictures and words in a selective attention task: Evidence for semantic processing of ignored stimuli // Mem. Cogn. 1988. V. 16. P. 64–70.
- Tomkins S.S.** Affect imagery consciousness: The complete edition. N.Y.: Springer Publishing Company, 2008. V. 1–4.
- Tonegawa S., Pignatelli M., Roy D.S., Ryan T.J.** Memory engram storage and retrieval // Curr. Opin. Neurobiol. 2015. V. 35. P. 101–109.
- Tong F., Meng M., Blake R.** Neural Bases of Binocular Rivalry // Trends Cogn. Sci. 2006. V. 10. P. 502–511.
- Tononi G.** Consciousness as integrated information: A provisional manifesto // Biol. Bull. 2008. V. 215. P. 216–242.
- Treisman A.** Feature binding, attention and object perception // Humphries G.W., Duncan J., Treisman A. (eds.) Attention, space and action. Oxford: Oxford Univ. Press, 1999. P. 91–111.
- Trevena J., Miller J.** Brain preparation before a voluntary action: Evidence against unconscious movement initiation // Conscious. Cogn. 2010. V. 19. P. 447–456.

- Trut L.N., Plyusnina I., Oskina I.N.** An experiment on fox domestication and debatable issues of evolution of the dog // *Russ. J. Gen.* 2004. V. 40. P. 644–655.
- Tsao D.Y., Freiwald W.A., Tootell R.B., Livingstone M.** A cortical region consisting entirely of face-selective cells // *Science*. 2006. V. 311. P. 670–674.
- Tsukiura T., Cabeza R.** Shared brain activity for aesthetic and moral judgments: Implications for the Beauty-is-Good stereotype // *Soc. Cogn. Affect. Neurosci.* 2011. V. 6. P. 138–148.
- Tulving E.** Episodic memory: from mind to brain // *Annu. Rev. Psychol.* 2002. V. 53. P. 1–25.
- Turing A.** On computable numbers, with an application to the entscheidungsproblem // *Proc. Lond. Math. Soc. Ser. 2.* 1936. V. 42. P. 230–265.
- Turing A.** Computing machinery and intelligence // *Mind*. 1950. V. 49. P. 433–460.
- Turk M., Turk I., Dimkaroski L.** The mousterian musical instrument from the Divje babe I cave (Slovenia): Arguments on the material evidence for neanderthal musical behavior // *L'Anthropologie*. 2018. V. 122. P. 1–28.
- Tversky A., Kahneman D.** Extension versus intuitive reasoning: The conjunction fallacy in probability judgment // *Psychol. Rev.* 1983. V. 90. P. 293–315.
- Tversky A., Shafir E.** The disjunction effect in choice under uncertainty // *Psychol. Sci.* 1992. V. 3. P. 305–309.
- Tye M.** *Consciousness, color, and content.* Cambridge, MA: MIT Press, 2000.
- Ungerleider L.G., Mishkin M.** Two cortical systems // Ingle D.J., Goodale M.A., Mansfield R.J.W. (eds.) *Analysis of visual behavior.* Cambridge, MA: MIT Press, 1982. P. 549–586.
- Valentine T., Darling S., Donnelly M.** Why are average faces attractive? The effect of view and averageness on the attractiveness of female faces // *Psychon. Bull. Rev.* 2004. V. 11. P. 482–487.
- Van Bergen R.S., Ma W.J., Pratte M.S., Jehee J.F.** Sensory uncertainty decoded from visual cortex predicts behavior // *Nat. Neurosci.* 2015. V. 18. P. 1728–1730.
- Van Gulick R.** *Consciousness* // Zalta E.N. (ed.) *The Stanford encyclopedia of philosophy.* 2014. URL: <https://plato.stanford.edu/entries/consciousness/#DivExpPro>.
- Van Riel R., Van Gulick R.** *Scientific reduction* // Zalta E.N. (ed.) *The Stanford encyclopedia of philosophy.* 2019. URL: <https://plato.stanford.edu/archives/spr2019/entries/scientific-reduction/>.
- Van Roojen M.** *Moral cognitivism vs. non-cognitivism* // Zalta E.N. (ed.) *The Stanford encyclopedia of philosophy.* 2018. URL: <https://plato.stanford.edu/archives/fall2018/entries/moral-cognitivism/>.

- Vartanian O., Navarrete G., Chatterjee A., Fich L.B., Leder H., Modroño C. et al.** Impact of contour on aesthetic judgments and approach-avoidance decisions in architecture // *PNAS*. 2013. V. 110. Suppl. 2. P. 10446–10453.
- Vataneşev D., Smallwood J., Jefferies E.** Varying demands for cognitive control reveals shared neural processes supporting semantic and episodic memory retrieval // *Nat. Commun.* 2021. V. 12. P. 2134.
- Veblen T.** The theory of the leisure class: An economic study of institutions. N.Y.: MacMillan, 1899.
- Vedral V.** Decoding reality. Oxford: Oxford Univ. Press, 2010.
- Vernot B., Akey J.M.** Resurrecting surviving Neandertal lineages from modern human genomes // *Science*. 2014. V. 343, N 6174. P. 1017–1021.
- Verplaetse J., De Schrijver J., Vanneste S., Braeckman J.** The moral brain. Heidelberg; London; N.Y.: Springer Dordrecht, 2009.
- Vigliocco G., Vinson D.P., Druks J., Barber H., Cappa S.F.** Nouns and verbs in the brain: A review of behavioural, electrophysiological, neuropsychological and imaging studies // *Neurosci. Biobehav. Rev.* 2011. V. 35. P. 407–426.
- Viskontas I., Quian Quiroga R., Fried I.** Human medial temporal lobe neurons respond preferentially to personally-relevant images // *PNAS*. 2009. V. 106. P. 21329–21334.
- Vitiello G.** Fractals as macroscopic manifestation of squeezed coherent states and brain dynamics // *J. Phys.* 2012. V. 380. P. 012021.
- Vogt B.A., Laureys S.** Posterior cingulate, precuneal and retrosplenial cortices: Cytology and components of the neural network correlates of consciousness // *Prog. Brain Res.* 2005. V. 150. P. 205–217.
- Von Neumann J.** Mathematical foundations of quantum mechanics. R.T. Beyer (trans.). Princeton: Princeton Univ. Press, 1996.
- Wallas G.** The art of thought. N.Y.: Harcourt Brace, 1926.
- Wang Z., Busemeyer J.R., Atmanspacher H., Pothos E.M.** The potential of using quantum theory to build models of cognition // *Top. Cogn. Sci.* 2013. V. 5. P. 672–688.
- Wang L., Zhang M., Zou F., Wu X., Wang Y.** Deductive-reasoning brain networks: A coordinate-based meta-analysis of the neural signatures in deductive reasoning // *Brain Behav.* 2020. V. 10. P. e01853.
- Watson J.B.** Behaviorism (Revised ed.). Chicago: Univ. of Chicago Press, 1930.
- Watzl S.** Structuring mind: The nature of attention and how it shapes consciousness. Oxford: Oxford Univ. Press, 2017.
- Waydo S., Kraskov A., Quian Quiroga R., Fried I., Koch C.** Sparse representation in the human medial temporal lobe // *J. Neurosci.* 2006. V. 26. P. 10232–10234.

- Wayman E.** Killer whales, grandmas and what men want: Evolutionary biologists consider menopause // *Sci. News*. 2013. V. 8.
- Webb T.L., Sheeran P.** Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence // *Psychol. Bull.* 2006. V. 132. P. 249–268.
- Wegner D.** *The illusion of conscious will*. Cambridge, MA: MIT Press, 2002.
- Weisberg R.W.** Toward an integrated theory of insight in problem solving // *Think. Reason*. 2014. V. 21. P. 5–39.
- Weiskrantz L.** *Blindsight: A case study and implications*. Oxford: Clarendon Press, 1986.
- Wen H., Shi J., Zhang Y., Lu K.-H., Cao J., Liu Z.** Neural encoding and decoding with deep learning for dynamic natural vision // *Cereb. Cortex*. 2018. V. 28. P. 4136–4160.
- Wheeler J.A.** Information, physics, quantum: The search for links // *Proc. III Int. Symp. on Foundations of Quantum Mechanics*. Tokyo, 1989. P. 354–368.
- Wheeler A.L. et al.** Identification of a functional connectome for long-term fear memory in mice // *PLOS Comput. Biol.* 2013. V. 9. P. e1002853.
- White T.** Early hominids diversity or distortion? // *Science*. 2003. V. 299, N 5615. P. 1994–1997.
- Wigner E.** *Symmetries and reflections. Scientific essays*. Bloomington: Indiana Univ. Press, 1967.
- Wilson E.O.** The biological basis of morality // *Atlantic Monthly*. 1998. V. 4. P. 53–70.
- Wölfflin H.** *Principles of art history*. Hottinger M.D. (trans.). N.Y.: Dover, 1950.
- Wolpert D.M.** Probabilistic models in human sensorimotor control // *Hum. Mov. Sci.* 2007. V. 26. P. 511–524.
- Wolpert D.M., Landy M.S.** Motor Control is Decision-Making // *Curr. Opin. Neurobiol.* 2012. V. 22. P. 996–1003.
- Wolpert S.** Neuroscientists can predict your behavior better than you can // *ScienceDaily*. 2010. V. 6. URL: <https://www.sciencedaily.com/releases/2010/06/100623110114.htm>.
- Wong D.B.** Integrating philosophy with anthropology in an approach to morality // *Anthropol. Theory*. 2014. V. 14. P. 336–355.
- Wood B.** The age of australopithecines // *Nature*. 1994. V. 372. P. 31–32.
- Wood B.** Human evolution: Fifty years after *Homo habilis* // *Nature*. 2014. V. 508. P. 31–33.
- Woodward J.** Emotion versus cognition in moral decision-making // Liao S.M. (ed.) *Moral brains: The neuroscience of morality*. Oxford Scholarship Online, 2016. V. 9. DOI: 10.1093/acprof:oso/9780199357666.001.0001.
- Wrangham R.W.** Two types of aggression in human evolution // *PNAS*. 2018. V. 115. P. 245–253.

- Wrangham R.W.** Hypotheses for the evolution of reduced reactive aggression in the context of human self-domestication // *Front. Psychol.* 2019. V. 10. P. 1914.
- Wu W.** *Attention.* L.: Routledge, 2014.
- Wu W.** The neuroscience of consciousness // Zalta E.N. (ed.) *The Stanford encyclopedia of philosophy.* 2018. URL: <https://plato.stanford.edu/archives/win2018/entries/consciousness-neuroscience/>.
- Wu M.C., David S.V., Gallant J.L.** Complete functional characterization of sensory neurons by system identification // *Ann. Rev. Neurosci.* 2006. V. 29. P. 477–505.
- Wurtz R.H.** Recounting the impact of Hubel and Wiesel // *J. Physiol.* 2009. V. 587. P. 2817–2823.
- Yancey G.** Neither Jew nor Gentile // *The SAGE handbook of intercultural competence.* Thousand Oaks, CA: SAGE Publications, 2012. P. 374–542.
- Yee E., Chrysikou E.G., Thompson-Schill S.L.** Semantic memory // Ochsner K., Kosslyn S. (eds.). *The Oxford handbook of cognitive neuroscience.* Oxford: Oxford Univ. Press, 2013. V. 1: Core topics. P. 353–374.
- Yue X., Vessel E.A., Biederman I.** The neural basis of scene preferences // *NeuroReport.* 2007. V. 18. P. 525–529.
- Zaki J., Ochsner K.** The cognitive neuroscience of sharing and understanding others' emotions // Decety J. (ed.) *Empathy: From bench to bedside.* Cambridge, MA: MIT Press, 2012. P. 207–226.
- Zeki S.** *Inner vision: An exploration of art and the brain.* Oxford: Oxford Univ. Press, 1999.
- Zeki S.** Essays on science and society. Artistic creativity and the brain // *Science.* 2001. V. 293, N 5527. P. 51–52.
- Zeki S., Romaya J.P., Benincasa D.M., Atiyah M.F.** The experience of mathematical beauty and its neural correlates // *Front. Hum. Neurosci.* 2014. V. 8. P. 68.
- Zhaoping L.** From the optic tectum to the primary visual cortex: Migration through evolution of the saliency map for exogenous attentional guidance // *Curr. Opin. Neurobiol.* 2016. V. 40. P. 94–102.
- Zihl J., Von Cramon D., Mai N.** Selective disturbance of movement vision after bilateral brain damage // *Brain.* 1983. V. 106. P. 313–340.

СПИСОК СОКРАЩЕНИЙ

ВВИ	– верхняя височная извилина
ВЗТ	– вентральный зрительный тракт
ВМПФК	– венстромедиальная префронтальная кора
ВТР	– вычислительная теория разума
ГСМ	– гипотеза соматического маркера
ДЛПК	– дорсолатеральная префронтальная кора
ДСМ	– дефолтная сеть
ИАТ	– имплицитный ассоциативный тест
ИИ	– искусственный интеллект
ИИТ	– теория интегрированной информации
КВТР	– Классическая вычислительная теория разума
КонТ	– коннекционистская теория
МВК	– медиальная височная кора
МВС	– модель встраивания слова
МПК	– медиальная префронтальная кора
МТ	– машина Тьюринга
НВИ	– нижняя височная извилина
ОФК	– орбитофронтальная кора
ПЛП	– потенциал локального поля
ППК	– передняя часть поясной извилины
РСМ	– распределительная семантическая модель
РП	– разделение паттернов
ТГРП	– теория глобального рабочего пространства
ТММ	– теория ментальной модели
ТМС	– транскраниальная магнитная стимуляция
ТРС	– теория регуляции страха
ТФП	– теория формальных правил
УМТ	– универсальная машина Тьюринга
УОГА	– устройство обнаружения гиперактивного агента
фМРТ	– функциональная магниторезонансная томография
ФПС	– фронто-париетальная сеть
ЦИ	– центральный исполнитель

Оглавление

Предисловие	3
Глава 1. Эволюция человека	8
1.1. Большой мозг	11
Глава 2. Психическое и мозг – философские аспекты	21
2.1. Элиминативизм и иллюзионизм	22
2.2. Эпифеноменализм	27
2.2.1. Редукционизм	28
2.2.2. Эмерджентизм	30
Глава 3. Сознание	33
3.1. «Сознание объяснено»	34
3.2. Сознание как эпифеномен активности мозга	39
3.3. Теории сознания	41
3.3.1. «Квантовое познание»	49
3.3.2. Вычислительная теория разума	55
3.3.3. Феноменология сознания	65
3.4. Эмпирические данные о нервном субстрате сознания	70
3.4.1. Феноменология и неврология зрительного восприятия	72
3.4.2. Байесовская модель	80
3.4.3. Нейролингвистика	84
3.5. Как же семантика «кодируется» в мозге?	90
Глава 4. Неосознанное	94
Глава 5. Когнитивные процессы и мозг	103
5.1. Память	—
5.2. Внимание	123
5.3. Мышление	129
5.4. Функции контроля в мозге	142
Глава 6. Свобода воли	145
Глава 7. Предварительный итог	158
Глава 8. Ценности	177
8.1. Поведение приближения и отступления/избегания	178
8.2. Эмоции	181
8.3. Мораль	192
8.3.1. Философия морали	193
8.3.2. Биология морали	202
8.3.3. Эмпатия	215
8.3.4. Мораль и мозг	222
8.4. Красота	231
8.4.1. Эволюционные теории чувства красоты	241
8.4.2. Современные эмпирические данные и теории	244
8.4.3. Нейроэстетика	249
8.5. Религия	253
8.6. Итоги по ценностным ориентациям	263
Заключение	265
Список литературы	268
Список сокращений	314

Научное издание

Тематический план выпуска изданий
Сибирского отделения РАН на 2022 г.

Князев Геннадий Георгиевич

**ЯВЛЯЕТСЯ ЛИ ПСИХИЧЕСКОЕ
ЭПИФЕНОМЕНОМ АКТИВНОСТИ МОЗГА?
Философские и нейробиологические аспекты проблемы**

Подготовлено к печати Сибирским отделением РАН

Редактор *Ф.Х. Сагалаева*
Дизайн обложки *В.Ю. Антонов*
Оригинал-макет *Н.М. Райзвих*

Подписано в печать 22.12.2022. Формат 60×90 1/16.
Уч.-изд. л. 18,2. Усл. печ. л. 19,8. Тираж 300 экз. Заказ № 368.

Сибирское отделение РАН
630090, Новосибирск, просп. акад. Лаврентьева, 17
Отпечатано в Сибирском отделении РАН
630090, Новосибирск, Морской просп., 2
Тел. (383) 330-84-66, e-mail: e.lyannaya@sb-ras.ru