ЖУРНАЛ ВЫСШЕЙ НЕРВНОЙ ДЕЯТЕЛЬНОСТИ, 2022, том 72, № 6, с. 800–825

ОБЗОРЫ И ТЕОРЕТИЧЕСКИЕ СТАТЬИ

УДК 159.91

КОДИРОВАНИЕ СМЫСЛА В АКТИВНОСТИ МОЗГА

© 2022 г. Г. Г. Князев*

Научно-исследовательский институт нейронаук и медицины, Новосибирск, Россия

**e-mail: knyazev@physiol.ru*

Поступила в редакцию 01.12.2021 г.
После доработки 26.04.2022 г.
Принята к публикации 26.04.2022 г.

Вопрос о природе отношения между психическими процессами и активностью мозга не является прерогативой философов. Все профессиональные ученые, занимающиеся изучением мозга человека, должны так или иначе решать для себя этот вопрос. Доминирующим в современном научном мировоззрении является редукционистский подход, согласно которому психические состояния можно в принципе свести к активности мозга. В этой статье я рассматриваю некоторые полученные в последние годы данные, которые проливают свет на природу связи между содержанием психических процессов и активностью мозга. Эти данные, касающиеся кодирования сенсорной и семантической лингвистической информации, а также более сложной информации, относящейся к содержанию абстрактных концепций, показывают, что активность мозга, сопровождающая извлечение смысла, имеет широко распределенный вероятностный характер, не соответствующий природе ментального содержания, которое, как правило, определенно и имеет холистический характер. Таким образом, на основе имеющихся в настоящее время эмпирических данных редукционизм вряд ли можно рассматривать как жизнеспособную опцию и единственный возможный в рамках материализма вариант, это признание ментальности эмерджентной сущностью.

Ключевые слова: содержание сознания, активность мозга, редукционизм, семантика, фМРТ, многомерный анализ паттернов

DOI: 10.31857/S000446772206003X

ВВЕДЕНИЕ

С появлением и развитием все более изощренных методов регистрации активности мозга нейробиология начинает проникать во все вообразимые области гуманитарных наук. Уже трудно найти сферу наук о человеке, в которой не было бы подраздела с приставкой “нейро-” (социальная нейронаука, культурная нейронаука, нейроэкономинка, нейроэтика, нейроэстетика, нейротеология и т.д.). Эмпирические данные, лежащие в основе этих вновь возникших дисциплин, состоят в экспериментально выявляемых корреляциях между психическими (чаще всего варианты самоотчета или параметры поведения) и физическими (активность мозга) переменными. Интерпретация этих корреляций, как правило, базируется на редукционистском мировоззрении, согласно которому психические переменные являются производной активности мозга. То есть предполагается, что, хотя мы пока и не знаем, как это сделать, в принципе, содержание психики может быть выведено из активности мозга.

В этой статье я хочу рассмотреть некоторые данные о природе связи между содержанием психических процессов и активностью мозга, чтобы оценить, в какой степени эти данные согласуются с редукционистским взглядом на природу психических процессов. Я не буду останавливаться на всех вариантах и тонкостях философской проработки вопроса (см., например, (Robinson, 2020; Schaffer, 2018)), так же как и на полном обзоре научных теорий и эмпирических данных. В плане философии я ограничусь лишь онтологической стороной проблемы: являются ли субъективно воспринимаемые психические состояния подклассом состояний мозга, или психические состояния и состояния мозга фундаментально

различны (Robinson, 2020)? В плане нейробиологии я лишь выборочно рассмотрю некоторые работы, которые, на мой взгляд, в наибольшей степени проливают свет на природу связи между активностью мозга и содержанием сознания.

Философские аспекты проблемы

При описании феноменологии сознания философы часто используют термин "феноменальное сознание" (Chalmers, 1995; Dennett, 1991; Frankish, 2016), подразумевающий наличие определенных свойств субъективного опыта, таких как квалиа (Nagel, 1974; Chalmers, 1996). Эти характеристики являются предметом серьезных споров среди философов (см., например, (Hacker, 2012)), и я не буду здесь на них останавливаться. Говоря о содержании сознания, я буду иметь в виду то, что Блок называет доступным сознанием – информацию, которая доступна для сознательного 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, стр. 227). Поскольку это содержание доступно для интроспекции и использования в поведении, оно доступно также и для научного исследования и подавляющее большинство работ по нейробиологии сознания исследуют именно его.

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 современной науки трудно представить, какова может быть природа ментального мира и как он может взаимодействовать с материальным. Доминирует в современном мировоззрении материалистический монизм, или физикализм, отрицающий существование чего-либо помимо материи. Поскольку содержание сознания не кажется материальным, а материализм отрицает существование нематериальных явлений или считает, что они определяются движением материальных объектов (Smart, 2016), то с позиций материализма нужно либо отрицать существование психических процессов, либо считать их побочным продуктом материальных процессов. Поэтому в рамках материализма спектр возможных интерпретаций природы этих явлений ограничен тремя вариантами. Первое направление называют "элиминативный материализм", потому что оно отрицает реальность ментальных явлений (Ramsey, 2020). Второе, редукционизм, утверждает, что эти 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 являются результатом активности мозга, их нельзя "свести" к этой активности, то есть они являются "эмерджентной" сущностью. Наконец, в последние годы все большую популярность у специалистов, занимающихся изучением сознания, приобретают альтернативные физикализму философские позиции, такие как панпсихизм (Chalmers, 2017), нейрофеноменология (Varela, 1996) или нейтральный монизм (Stubenberg, 2018), но в этой статье я не буду на них останавливаться.

Элиминативисты считают, что обыденные представления о психических процессах (в том числе их субъективное восприятие) не имеют нейробиологического базиса. На основании этого делается вывод, что эти явления просто не существуют, так как критерием реальности существования ментального явления, с их точки зрения, является его редуцируемость до нейробиологического уровня (Churchland, 1986). Вариантом элиминативизма можно считать направление, которое рассматривает ментальные явления как иллюзии. Так, Деннетт и некоторые другие авторы считают, что феномены субъективного восприятия, такие как чувство боли, цвета и прочее, в корне ошибочны, потому что не соответствуют реальным свойствам объектов и процессам в мозге, и потому могут считаться иллюзией (Dennett, 1978; Frankish, 2016). Критики элиминативизма считают его логически непоследовательным и чрезмерно радикальным вариантом материализма (см., например, (Reppert, 1992; Searle, 1997)).

Редукционизм, в отличие от элиминативизма, признает, что ментальные явления реально существуют, но они могут быть "сведены" к физическим процессам в мозге. Редукционизм, как тенденция объяснять свойства сложных объектов свойствами составляющих их более простых объектов, возник и укрепился в процессе появления и развития есте-

ственных наук. В основе редукционизма лежит естественное и рациональное стремление к экономности и универсальности научных теорий. Во многих случаях редукционистский подход оказался плодотворным и в значительной степени определил прогресс науки за последние 200 лет. В других случаях, однако, редукция оказывается невозможной, так как поведение некоторых сложных систем подчиняется законам, которые не выводятся понятным образом из законов, управляющих поведением их компонентов.

Пожалуй, нигде вопрос редукции не являлся предметом таких дебатов, как при обсуждении природы ментальных явлений. В качестве примера крайнего редукционизма можно привести точку зрения Френсиса Крика, который после получения Нобелевской премии за открытие структуры ДНК стал заниматься нейробиологией: “Вы, ваши радости и печали, ваши воспоминания и ваши амбиции, ваше чувство личной идентичности и свободы воли – на самом деле не более чем поведение огромного скопления нервных клеток и связанных с ними молекул” (Crick, 1994). В 90-х – начале 2000-х годов Крик в соавторстве с Кристофом Кохом опубликовал серию статей, в которых они очерчивают программу научного решения проблемы сознания и советуют отбросить философские мудрствования (Crick, Koch, 1998, 2003): “Мы считаем, что философские аспекты проблемы (сознания) следует оставить в стороне, так как пришло время для ее научной атаки” (Crick, Koch, 1998, стр. 97). “Мы не будем описывать различные мнения философов, разве что скажем, что ... исторически у них очень плохой послужной список в получении достоверных научных ответов” (Crick, Koch, 1998, стр. 103). В основе этой критической в отношении философии позиции лежит глубоко редукционистское убеждение, согласно которому философия себя изжила, и физика и другие естественные науки способны дать ответы на все философские вопросы (Hawking, 1988).

Центральным звеном предложенной Криком и Кохом программы является поиск нервных коррелятов сознания (НКС). В частности, в последней совместной статье, опубликованной уже после смерти Крика, авторы в качестве центральной структуры сознания предлагают клауструм (Crick, Koch, 2005). Интересно, что Кристоф Кох, который с 2011 года является президентом Института Аллена по изучению мозга, с течением времени отошел от редукционизма и в настоящее время является адептом современного варианта панпсихизма, согласно которому ментальное не может быть сведено к материальному и является фундаментальным свойством всего сущего (Koch, 2021).

С прагматической точки зрения знание НКС можно использовать для предсказания ментальных феноменов на основе нервной активности, однако достаточно ли этого для понимания природы сознания? Наличие НКС можно объяснить даже в рамках дуализма. Какова природа этих корреляций? Как из работы мозга “возникает” содержание сознания? Есть ли причинная связь между активностью мозга, содержанием сознания и поведением? На субъективном уровне нам кажется, что содержание мыслей в значительной степени определяет наше поведение. Редукционистский подход предполагает, что поведение определяется исключительно активностью мозга, а каузальная эффективность ментального содержания – лишь кажущаяся. Чтобы объяснить эту кажущуюся эффективность, необходимо объяснить, как одно и то же состояние мозга порождает содержание сознания и соответствующее ему поведение. Предполагается, что содержание сознания “закодировано” в активности мозга и прогресс нейробиологии позволит в конечном итоге взломать этот код и “читать” содержание мыслей прямо по активности мозга.

Давайте рассмотрим, как философы определяют понятие редукции. “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, стр. 1). “Утверждение, что *x* сводится к *y*, обычно означает, что *x* есть не что иное, как *y*” (van Riel, Van Gulick, 2019, стр. 1). В данном случае это означает, что концепции психологии должны непосредственно выводиться из концепций нейрофизиологии и содержание мыслей есть не что иное, как активность соответствующих нейронов.

Третья, потенциально совместимая с материализмом концепция состоит в том, что сознание – это эмерджентная сущность. Разные теоретики вкладывают разный смысл в

понятие эмерджентности, но общим во всех определениях является то, что эмерджентная сущность “возникает” из более фундаментальных сущностей, но не “сводится” к ним (O’Connor, 2020). 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). Слабый эмерджентизм отрицает существование нисходящей каузальности. Сильный эмерджентизм предполагает меньшую степень согласованности между явлениями низшего и высшего уровней и допускает наличие нисходящей каузальности (Bedau, 1997).

В последующем тексте я попытаюсь проанализировать некоторые данные о характере нервной активности, сопровождающей некоторые ментальные процессы, с целью оценки правдоподобия выполнимости выше представленных требований редукционизма. В последние десятилетия проведено большое количество исследований кодирования/декодирования информации в активности мозга. В этих исследованиях под кодированием/декодированием обычно подразумевается использование математических алгоритмов классификации и распознавания паттернов. Под кодированием подразумевается предсказание параметров активности мозга исходя из семантики сенсорного восприятия или лингвистических переменных, а под декодированием – обратная процедура (см., например, (Naselaris et al., 2011; Wen et al., 2018)). Большая часть этих исследований выполнена с использованием фМРТ и относится к кодированию сенсорной (преимущественно зрительной) информации. Немало данных накоплено и о механизмах кодирования в мозге семантической лингвистической информации. Есть исследования о кодировании и более сложной информации, относящейся к содержанию мыслей, в том числе абстрактных научных концепций. В последующем тексте я выборочно рассмотрю наиболее представительные работы по каждому из этих направлений. Я ни в коей мере не претендую на полный обзор соответствующих исследований и выберу лишь отдельные, наиболее типичные и представительные, на мой взгляд, работы. Такой выборочный подход диктуется не только обилием исследований и невозможностью их полного обзора в этой короткой статье, но и тем, что для понимания сути эмпирических данных наиболее представительные работы должны быть рассмотрены более детально, чем это обычно делается в обзорах. Нужно отметить также, что все рассматриваемые процессы неизбежно связаны с извлечением информации из семантической памяти, и соответствующие исследования и теоретические модели во многом перекрываются с исследованиями и моделями семантической памяти. Я, однако, не буду рассматривать специфически связанные с памятью проблемы, такие как механизмы запоминания, хранения и извлечения следа, различия между эпизодической и семантической памятью и так далее.

Кодирование семантики зрительного восприятия

Под семантикой зрительного восприятия я в данном случае понимаю восприятие смысла увиденного, в противоположность, например, восприятию бессмысленной мозаики цветовых пятен. Обычно это связано с узнаванием объектов, что неизбежно требует участия памяти, но для начала я хочу рассмотреть простейший пример – восприятие 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 нейронов 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). Однако содержание чувственного образа не имеет вероятностного характера – мы определяем направление движения объекта однозначно, а не в вероятностных терминах. Теннисист во время матча не сомневается даже долю секунды, летит ли мяч в правый или в левый сектор. 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 все" (Block, 2018). Остается неясным, как и где происходит выявление "победителя". В последних строках своей статьи Блок подчеркивает, что он не допускает, что в мозге есть конечная стадия процессинга, на которой принимается решение, – "картезианский театр" в терминологии Деннетта (Dennett, 1991), однако альтернативного объяснения не приводит (Block, 2018).

Если моделировать процесс соревнования между нейронными популяциями механистически, в инженерных терминах, то выявление "победителя" требует, как минимум, наличия некоего "компаратора", на котором должна сходиться активность всех соревнующихся модулей. Если придерживаться эпифеноменологической интерпретации сознания, то этим компаратором может быть только некая популяция нейронов, которая должна отвечать на приходящие от первичных модулей сигналы разной интенсивности по принципу "все 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 этом случае можно было бы думать, что 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. Эмпирически ничего похожего на такого рода компаратор в мозге пока обнаружить не удалось.

На людях механизмы представления семантической информации в мозге изучаются преимущественно с использованием фМРТ. Замечу в скобках, что, строго говоря, эти 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 используется обычная линейная регрессия или итеративная Байесовская модель, а в последние годы обычным 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 обычно 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 соответствует яркости 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 – это 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 представляет собой уникальную комбинацию 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 versus здания, или различные сегменты пространства в задачах на ориентацию).

В исследовании Назелариса с соавторами сравнивалась предсказательная сила извлеченных разными способами признаков в от-

ношении 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) лучше предсказывали активность более высоко в иерархии расположенных 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 определенной 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, то есть 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 (то есть те, которые отвечали на семантические признаки) были объединены в общую модель. Когда использовались лишь две семантические категории (одушевленные и неодушевленные объекты), максимальная точность предсказания этой модели достигала 90%; когда были использованы 23 категории, максимальная точность была 40% (Naselaris et al., 2009).

Я относительно подробно описал это исследование потому, что его можно считать типичным. Из него 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 информацией, хранящейся 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. То есть по этим данным активность мозга лишь очень грубо соответствует семантической информации, предположительно представленной в сознании.

Представленные выше данные подчеркивают сейчас уже хорошо известный факт: кодирование зрительных образов организовано в виде многоуровневой системы по ходу вентрального зрительного тракта. 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). В противоположность этому нейроны конечного звена вентрального зрительного пути в нижней височной извилине (НВИ) отвечают на принадлежность стимула к той или иной семантической категории (лица, объекты, части тела) (Gross, 2008). 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 аналогичные исследования кодирования и слуховой информации, например, фрагментов музыки (Hoefle et al., 2018). Иерархическая послойная организация зрительного анализатора послу-

жила прототипом для создателей нейронных сетей глубокого обучения. Исходя из успешности использования этих сетей для решения задач категоризации и распознавания образов, можно думать, что подобная архитектура позволяет решать такие задачи, однако конкретные механизмы извлечения атрибутов высшего уровня в мозге неизвестны. Неизвестно также, как на основе этих атрибутов происходит категоризация объектов.

В любом случае в пределах вентрального зрительного пути обработка стимула заканчивается отнесением его к той или иной семантической категории. Степень селективности и инвариантности нейронов НВИ ограничена. Эти нейроны реагируют на большое количество разных стимулов, в основном в рамках предпочитаемой категории (Tsao et al., 2006). 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 (то есть по срабатыванию нейрона мы не можем сказать, чье это лицо) и распределяется в популяции нейронов (Quian Quiroga, Kreiman, 2010).

Кажется странным, что активность нейронов НВИ позволяет определить лишь, к какой категории относится объект. Если распознавание 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 содержать. Действительно, в одной из недавних работ с регистрацией активности нейронов НВИ у обезьян, которым показывали изображения человеческих лиц, показано, что можно с достаточной точностью определить идентичность лица по активности лишь 200 нейронов НВИ, если каждое лицо представлено в виде точки в 50-мерном пространстве (Chang, Tsao, 2017). Для формирования этого пространства каждое лицо представлялось в виде показателей формы и яркости, 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? Авторы предполагают, что это достигается с помощью архитектуры аналогичной иерархической глубокой сети с прямой связью, но это пока лишь гипотеза. Важно подчеркнуть, что активность нейронов НВИ кодирует не образ лица, а выраженность определенных характеристик. “Цель” нейронов НВИ состоит в том, чтобы настроить систему координат для измерения лиц, а не в том, чтобы идентифицировать лица (Chang, Tsao, 2017). Как из этой системы координат возникает образ лица, представленный в сознании, остается загадкой.

Перцептивная осведомленность о том, к какой категории принадлежит стимул, появляется уже через 100–170 мс после его предъявления и связана с активностью НВИ, а узнавание конкретного лица происходит через 300 мс и связано с активностью так называемых “концептуальных клеток” в гиппокампе (Quian Quiroga, 2016). Концептуальные клетки отвечают на определенные концепции (например, знакомый человек), а не на отдельные атрибуты. 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). Концептуальные клетки найдены только у людей, но не у животных. У людей показано также наличие нейронов, кодирующих ассоциации между концепциями. Такие нейроны могли отвечать, например, на изображение актера и изображение аэроплана, если этот актер сни-

мался в фильме под названием “Аэроплан” (Rey et al., 2020). Таким образом, в то время как зрительная кора высокого уровня участвует в начальной дифференциации, разделяющей стимулы на категории примерно через 100 мс после их предъявления, извлечение смысла и узнавание происходит в течение последующих 200 мс, после чего активация передается нейронам гиппокампа для кодирования концепций и запоминания ассоциаций между ними.

На примере зрительного анализатора можно видеть, что процесс анализа информации по мере продвижения от низших уровней к высшим состоит в последовательно сужающейся семантической категоризации. Адаптивный смысл этого очевиден. Например, если в течение 100 мс зрительный анализатор способен определить, что воспринимаемый объект является лицом или зданием, это резко снижает количество возможных вариантов развития событий и организации поведения. Остается неясным, как конкретно эта категоризация осуществляется в мозге. На входе, в первичной зрительной коре нейроны отвечают на зрительную стимуляцию распределенной активацией, которая вероятностно 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 со входной комбинацией. В результате сравнения гипотеза либо принимается и сигнал передается в верхние отделы для проверки уточняющих гипотез, либо отвергается и заменяется альтернативной гипотезой. Показано, что точность предсказания семантической категории стимула на основе фМРТ-данных улучшается, когда для классификации используются двунаправленные рекуррентные нейронные сети, в конструкцию которых заложены как восходящие, так и нисходящие потоки информации (Qiao et al., 2019). Предполагается, что мозг отдает приоритет декодированию высокоуровневых атрибутов, потому что они более релевантны с точки зрения поведения и категоризации и более инвариантны, соответственно, их легче удерживать в рабочей памяти. Поэтому декодирование более высокого уровня налагает нисходящие ограничения на менее надежное декодирование нижнего уровня (Ding et al., 2017).

Для того чтобы генерировать гипотезы, мозг должен извлекать информацию о сходных ситуациях из семантической и/или эпизодической 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 в явном виде кодируется принадлежность воспринимаемого объекта к какой-либо семантической категории (например, лица). Далее, по теории Кироги, в течение последующих 200 мс происходит определение субъективного смысла воспринимаемой информации (т.е., например, что это не просто лицо, а лицо моего соседа). Процесс извлечения смысла заканчивается активацией ансамбля нейронов гиппокампа, соответствующего концепции моего соседа. Однако как происходит извлечение смысла? В отношении этого Кирога говорит лишь, что это происходит, вероятно, не в зрительном анализаторе и не в гиппокампе, а требует участия других областей коры (Quian Quiroga, 2020).

В какой-то степени на этот вопрос отвечает недавнее исследование узнавания лиц близко или поверхностно знакомых людей (di Oleggio Castello et al., 2021). 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, такие как лицевая область в веретенообразной извилине (Grill-Spector et al., 2004). Однако для извлечения всей необходимой информации о человеке активации лишь веретенообразной извилины недостаточно. Авторы рассматриваемой работы задались вопросом, какие области мозга участвуют в опознании близко и поверхностно знакомых людей и одинакова ли топография этих областей у разных людей, знающих одного и того же человека. Для ответа на последний вопрос авторы использовали несколько искусственный прием “гипервыравнивания” фМРТ-данных всех испытуемых. Испытуемые предварительно просматривали один и тот же фильм, и у каждого из них для последующего анализа были выбраны лишь

те воксели, которые у всех испытуемых одинаково активировались при просмотре этого фильма. При анализе данных использовали методы многомерного анализа паттернов (МАП). Было обнаружено, что у всех испытуемых личности как близко, так и поверхностно знакомых людей можно было декодировать из активности базовой системы визуальной обработки лиц, включающей помимо лицевой области веретенообразной извилины большой набор структур в теменной, височной и лобной коре и преклиновидной полоске. Для представления близких знакомых дополнительно требовалась активация расширенной системы обработки невизуальной информации социального характера, включающей правую височно-теменную связку, медиальный префронтальный кортекс, преклиновидную полоску и правый островок (di Oleggio Castello et al., 2021). Когда далее мы будем рассматривать работы по кодированию концепций, мы увидим, что такое широко распределенное представительство характерно для репрезентации практически любого смысла. Как из этой распределенной активности возникает смысл – остается загадкой.

Моделирование содержания рабочей памяти

Как видно из предыдущей секции, различные характеристики стимула можно предсказать из активности мозга. Есть данные, что можно предсказать даже субъективную уверенность в наличии той или иной характеристики (Burge, 2010; Peters et al., 2017; van Bergen et al., 2015). В последние годы стало популярным применять для описания поведения популяций нейронов Байесовскую модель. Согласно этой модели, извлечение информации из активности популяции нейронов дает не характеристику стимула (например, направление движется слева направо), а функцию распределения вероятностей в пространстве возможных направлений движения. Нейроны как бы рассчитывают апостериорную вероятность с учетом эффекта стимула и вероятности события на основе предшествующего знания (Rescorla, 2015). По этой теории активность совокупности нейронов содержит совместное представление как самого стимула, так и его неопределенности и, возможно, даже распределение полной байесовской апостериорной вероятности (Jazayeri, Movshon, 2006).

Субъективное ощущение неуверенности, связанное с неопределенностью информации, может иметь место при наличии помех в процессе восприятия (например, густой туман или плохое зрение). Точно так же мы можем быть не уверены в содержании информации, хранящейся в памяти, например, когда нужно удерживать большой объем информации в рабочей памяти. То есть, по крайней мере в некоторых случаях, появление в сознании информации сопровождается субъективной уверенностью/неуверенностью в точности этой информации. Чему соответствует это чувство в активности мозга? Выяснению этого вопроса посвящена недавно опубликованная работа, в которой тестировались предсказания Байесовской модели в отношении содержания рабочей памяти о локализации точки в пространстве. Было показано, что среднее значение распределения вероятностей, декодированных из BOLD-активности вокселей ретинотопных кортикальных областей, предсказывало поведенческие ошибки при выполнении теста на рабочую память, а ширина этого распределения предсказывала субъективную неуверенность в результате (Li et al., 2021). С позиций редукционизма этот результат можно интерпретировать как доказательство того, что не только содержание рабочей памяти, но и субъективная уверенность в ее надежности “записаны” в активности нейронов. Но так ли это? В работе Ли с соавторами содержание ментального домена предсказывается из параметров активности мозга с помощью генеративной Байесовской модели BOLD-активности (van Bergen et al., 2015). Эта модель позволяет (с некоторой, не очень высокой степенью надежности) “извлечь” из активности нейронов содержание сознания. Но может ли такое извлечение происходить в самом мозге? Согласно теории кодирования распределения вероятности в популяции нейронов, мозг знает генеративную модель, которая описывает активность нейронной популяции как функцию характеристик стимула (например, его местоположения или ориентации), включая распределение шума. Использование этих знаний позволяет мозгу оценить соответствующий уровень неопределенности, связанный с функцией стимула (Jazayeri, Movshon, 2006; Ma et al., 2006). Ли с соавторами предполагают, что использованная ими для декодирования распределения вероятностей Байесовская модель должна быть похожа на ту, которую мозг может ис-

пользовать для принятия решений (Li et al., 2021). Однако где в мозге записана эта модель, откуда она появилась, и где и как производится расчет распределения вероятностей и вычисляется его среднее значение и вариация? На все эти вопросы теория ответов не дает. Если расчеты делает мозг, то результат этого расчета должен появиться в виде активности какой-то другой популяции нейронов. Ничего подобного в работе Ли с соавторами не обнаружено. Например, можно было бы думать, что активность ретинотопных кортикальных областей – это низший уровень процессинга, из которого на более высоких уровнях в ассоциативной коре как раз и извлекается нужная информация. Однако в работе Ли с соавторами показано, что в отношении предсказания содержания сознания активность ассоциативных зон затылочной и лобной коры ничем не отличается от активности ретинотопной коры – из нее также можно извлечь это содержание лишь с помощью той же математической модели, но само содержание там не представлено. Таким образом, в работе Ли с соавторами, как и во многих других похожих работах, показано, что с помощью математического анализа исследователь может извлечь из активности мозга информацию, коррелирующую с информацией, содержащейся в сознании, однако остается загадкой, как и где это происходит в реальности. Как содержание сознания извлекается из широко распределенной активности мозга и где (помимо сознания) оно представлено?

Примеры из нейролингвистики

Цель нейролингвистики – расшифровать нейрональную основу знания и использования языка. В связи с обсуждаемыми в этой статье вопросами наибольший интерес представляет кодирование в мозге семантической, то есть смысловой информации. Лингвисты 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 является 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 является частью значения слова. Максималисты же, наоборот, считают, 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 это значение можно определить путем простого подсчета всех возможных контекстов. Этот подход называют дистрибутивной семантической моделью (ДСМ). ДСМ позволяет представить значение слова в виде реального числа, путем подсчета количества контекстов, в которых это слово используется в достаточно большой базе письменных текстов. ДСМ часто используется в когнитивных и нейролингвистических исследованиях. Существует два основных варианта ДСМ. В первом варианте составляется контекстуальная матрица слов. В этой матрице каждое слово задано рядом значений, каждое из которых равно количеству случаев, когда это слово встречается в контексте с каким-либо другим словом. В последние годы чаще используется другой вариант ДСМ, в котором с помощью нейронных сетей рассчитывается вероятность найти слово в определенном контексте. Используется прямая (feed-forward) нейронная сеть, у которой в 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 (МВС, word embedding). МВС можно получать и другими методами, основная идея остается той же и была сформулирована уже в 50-х годах прошлого века: “слово характеризуется его компанией” (Firth, 1957). Слова, относящиеся к одному и тому же семантическому домену, чаще встречаются в одинаковых контекстах. Степень сходства двух слов можно оценить корреляцией их 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 ДСМ использовалась в нейролингвистическом исследовании, была опубликована в 2008 году (Mitchell et al., 2008). Была использована фМРТ, записанная в эксперименте с предъявлением 60 конкретных существительных, принадлежащих 12 семантическим кате-

гориям, которые предъявлялись девяти испытуемым в случайном порядке 6 раз каждое. фМРТ-ответ на каждое слово рассчитывался 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 так далее) 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), авторы которой 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 (таких как “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. Анализ фМРТ-ответов методом главных компонент выявил четыре компонента, объясняющих большую часть вариации. Первый, наиболее сильный в плане объясненного разнообразия компонент одним своим полюсом представлял категории, относящиеся к людям и социальным взаимодействиям, а другим – перцептивные, количественные и пространственные дескрипторы. Остальные три компонента труднее поддавались интерпретации. Описанные в статье семантические категории охватывали 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 сентенций на любую тему, включая абстрактные идеи. Семантические векторы были посчитаны для всех слов базового словаря (~30000 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. Затем декодер тестировали на новом лингвистическим 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). Далее авторы показывают, что декодер можно использовать и для предсказания смысла фраз и сентенций, 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 предшествующих 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.

Давайте теперь, после рассмотрения этих этапных и представительных исследований в области кодирования лингвистической семантической информации, попробуем разобраться в том, как эти данные можно интерпретировать в плане соответствия представ-

ленного в сознании смысла слов выявляемым с помощью фМРТ паттернам активности мозга. Прежде всего нужно отметить, что использование модели встраивания слова (МВС) в качестве идентификатора его смысла имеет свои ограничения. МВС зависит от размеров и содержания базы текстов, использованных для ее построения. Показано, что 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, МВС 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 повторяет организацию 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, чтобы тематические категории каким-то образом распознавались на предшествующем этапе анализа, однако механизмы этого распознавания пока неизвестны. Можно думать, что это распознавание должно включать извлечение информации из семантической памяти и сравнение признаков воспринимаемых объектов с признаками, характерными для разных семантических категорий. 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 мы видели, 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 на основе активности мозга семантической категории, к которой относится слово, была достоверно выше уровня случайных совпадений, но не очень от него отличалась (0.7 и 0.5 соответственно (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, так же как кодирование отдельного лица, кодирование отдельного слова или фразы представлено в виде активности, широко распределенной по разным отделам мозга. Механизмы извлечения смысла из этой широко распределенной и вероятностной по своей природе активности остаются непонятными.

Подходы и общие данные по кодированию концепций

Вместе с рассмотренными выше данными о "кодировании" в мозге семантики слов, механизмы кодирования мыслей, и особенно абстрактных концепций, представляют, пожалуй, наибольший интерес, так как они связаны со специфически человеческими когнитивными процессами, без которых невозможно было бы существование человеческой культуры. Можно выделить два основных методологических подхода к вопросу о репрезентации мыслей в активности мозга. Классический когнитивизм, в основе которого лежит вычислительная теория разума, исходит из представления, согласно которому мозг принципиально не отличается от компьютера и устроен по принципу функционально специализированных модулей (Lilienfeld et al., 2010). Модули, специализирующиеся на восприятии сенсорной информации или организации движения, отделены от модулей, участвующих в мыслительной деятельности, в качестве которых чаще всего рассматриваются области височной коры, связанные с лингвистическими процессами. Соответственно, смысл символов и концепций, в том числе извлеченный из сенсорной информации, обрабатывается в специальном центре, в качестве которого чаще всего рассматривают передний височный полюс (Lambon Ralph et al., 2016).

Альтернативой традиционному когнитивизму является популярное сейчас направление, которое называют "воплощенная когниция" (embodied cognition). Концепция воплощенной когниции (КВК) постулирует, что смысл символов и концепций имеет корни в нашем опыте взаимодействия с внешним миром и доступ к концептуальному знанию задействует те же процессы, которые активны при получении или непосредственном использовании этого знания (Shapiro, 2019). КВК предсказывает, что понимание смысла возникает в результате "проигрывания" чувств, испытанных при восприятии объекта, на который ссылается слово. Таким образом, сенсомоторные области, участвующие в восприятии и действии, должны перекрываться с областями мозга, которые активны во время понимания речи (Barsalou, 2008; Bergen, 2012). Промежуточный вариант (слабая версия КВК) предполагает иерархическую организацию семантического процессинга и существование зон конвергенции, связанных с сенсорными и моторными областями (Damasio, Damasio, 1994; Papeo et al., 2014).

Кроме теоретических моделей организации когнитивных процессов, большое значение в этой области исследований, которую иногда называют нейросемантикой, имеют методы анализа данных. Большая часть наиболее интересных результатов получена в последние годы с использованием методов многомерного анализа паттернов (МАП), которые использовались и в некоторых уже описанных в предыдущих секциях исследованиях распознавания лиц и семантики слов (di Oleggio Castello et al., 2021; Pereira et al., 2018). Прежде чем рассмотреть последние данные о репрезентации концепций в активности мозга, полученные методами МАП, я вкратце рассмотрю данные тестирования предсказаний КВК традиционными методами статистического параметрического картирования (СПК).

Данные исследований, изучающих доступ к одной и той же семантической информации при использовании разных стимулов и разных типов задач, показывают, 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 показано 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, что согласуется с предсказаниями КВК (Yee et al., 2013).

Хотя фМРТ-данные, в целом согласующиеся с КВК, многочисленны (см., например, обзор (Barsalou et al., 2003)), их анализ показывает, что доступ к знанию часто активирует области коры, расположенные немного впереди областей, активируемых при получении этого знания, причем этот сдвиг тем больше, чем более абстрактная форма знания извлекается (Rugg, Thompson-Schill, 2013). В целом, если связь конкретных семантических концепций с активацией сенсорной и моторной коры многократно показана, эта связь менее очевидна для абстрактных концепций, которые могут быть в большей степени связаны с активацией лингвистических зон коры (Hart, Kraut, 2007). Desai и соавт. (Desai et al., 2013) обнаружили, что сенсомоторная активация в ответ на слова, означающие действия, сильно зависит от контекста. Они наблюдали постепенное снижение вовлеченности сенсомоторных областей при представлении буквальных, метафорических, идиоматических и абстрактных значений этих слов. Кодирование абстрактных концепций, которые не имеют очевидных сенсомоторных компонентов, таких как "числа", "демократия" или "истина", представляет собой серьезную проблему для КВК (Dove, 2016). Нарушения сенсомоторного функционирования регулярно сопровождаются нарушениями процессинга конкретных синтаксических знаний, но абстрактная концептуальная информация в гораздо меньшей степени связана с сенсомоторными областями (Hauk, Tschentscher, 2013). Так, Вукович с соавторами показали, что блокирование моторных областей коры с помощью транскраниальной магнитной стимуляции замедляло понимание слов, связанных с действием, но понимание абстрактных слов даже ускорялось (Vukovic et al., 2017).

Слабый вариант КВК, предполагающий существование зон конвергенции и иерархи-

ческую организацию семантического процессинга, считается в настоящее время наиболее перспективным (Galetzka, 2017), однако конкретные механизмы мультимодальной интеграции неизвестны (Binder, Desai, 2011). Исследователи в области КВК предполагают, что крупномасштабные нейронные сети, включающие области мозга, участвующие в восприятии и движении, необходимы для начального обучения и последующего применения семантических понятий, и что специфическое взаимодействие между системами, поддерживающими обработку речи, и системами, поддерживающими механизмы восприятия и движения, позволяет формировать абстрактные концепции (Tschentscher, 2017). Но, может быть, связь с системами, поддерживающими механизмы восприятия и движения, важна лишь на стадии формирования концепции? Азиз-Заде и Дамасио (Aziz-Zadeh, Damasio, 2008) считают, например, что хорошо знакомые метафоры могут не использовать те же сенсомоторные ресурсы, которые нужны для недавно созданных метафор. Это предположение подтверждается данными, показывающими, что абстрактные концепции ассоциируются с сенсомоторными переживаниями лишь на ранних стадиях развития. Например, у детей чувство привязанности и любви часто сопровождается ощущением физического тепла. Эта ассоциация может сохраняться долго, но у взрослых абстрактные концепции уже могут быть отделены от сенсомоторной основы (Casasanto, 2017).

Недавние обзоры последних данных предоставляют доказательства в пользу того, что нейронные репрезентации социальных и эмоциональных знаний, психических состояний и концепций величины вовлекают системы мозга, участвующие в соответствующих переживаниях, и, таким образом, подтверждают расширение воплощенных моделей организации семантической памяти на несколько типов абстрактных знаний. В целом, однако, можно заключить, что определенные данные о связи абстрактных концепций с активацией сенсомоторной коры отсутствуют. В частности, ни одно исследование с использованием стимуляции мозга не продемонстрировало наличие причинной связи между активностью моторной коры и процессингом абстрактных слов. Нет доказательств того, что временное блокирование моторной коры с помощью транскраниальной магнитной стимуляции нарушает семантическую обработку абстрактных понятий. Также нет никаких доказательств в пользу прямого действия стимуляции моторной коры на обработку чисел (Tschentscher, 2017).

Далее я рассмотрю недавние работы по нервным репрезентациям концепций, которые позволили существенно продвинуться вперед благодаря использованию методов многомерного анализа паттернов (МАП) с использованием машинного обучения (см. (Bauer, Just, 2019; Vargas, Just, 2020) для более полного обзора этих работ). Эти методы более чувствительны, чем традиционно использовавшиеся в нейровизуализационных исследованиях методы статистического параметрического картирования (СПК). Кроме того, в них заложена другая идеология. В основе СПК лежит предположение об участии в когнитивных процессах определенных центров мозга. Поэтому одним из критериев выявления достоверных вокселей является их группирование в пространственно-ограниченные кластеры. Методы МАП позволяют выявить достоверные воксели независимо от того, группируются ли они друг с другом или разбросаны по всему мозгу.

Одним из подходов в рамках МАП является анализ сходства репрезентации (АСР) разных концепций. АСР переводит паттерн активации одной концепции в паттерн ее похожести на другую концепцию. Другой подход – использование факторного анализа или анализа главных компонент для извлечения “нейронально-значимых” размерностей из многомерного паттерна активации. Под нейронально-значимыми подразумевается то, что набор концепций может систематически активировать набор релевантных вокселей. Например, конкретные объекты, которые влекут за собой взаимодействие с частями человеческого тела (например, ручные инструменты), вызывают активацию моторных и премоторных областей, так что возникает нейрональная размерность взаимодействия тела и объекта (Just et al., 2010).

Области мозга, соответствующие размерности, могут быть локализованы по факторным нагрузкам различных кластеров вокселей. После того как выявлены размерности и связанные с ними концепции и воксели, эти размерности нуждаются в интерпретации, которая часто делается на основе прошлого знания о функциональной роли вовлеченных регионов и характера концепций, тесно связанных с размерностью. Один из подходов к

оценке правдоподобия интерпретации состоит в том, чтобы получить от независимой группы экспертов оценки выраженности постулируемой характеристики в каждой из концепций. Корреляция между рейтингом экспертов и факторными оценками позволит оценить, насколько хорошо интерпретация размерности соответствует данным активации. Этот метод использовался для извлечения и интерпретации семантически значимых размерностей, лежащих в основе репрезентации как конкретных существительных, так и абстрактных понятий (Just et al., 2010; Vargas, Just, 2019).

Основная теоретическая концепция, которая лежит в основе интерпретации пространственно распределенных нейронных репрезентаций, соответствует KBK и состоит в том, что области мозга, которые вместе представляют данную концепцию, соответствуют системам, которые участвуют в физическом и умственном взаимодействии с референтами концептов. Например, понятия, относящиеся к физически манипулируемым объектам, таким как нож, включают то, как они выглядят, для чего используются, как их держат и используют и т.д., в результате чего возникает нейронная репрезентация, распределенная по сенсорным, перцептивным, моторным и ассоциативным областям (Bauer, Just, 2019).

В качестве размерностей можно также рассматривать разные референты концепта, которые можно извлечь из описаний людей или из базы, содержащей большое количество текстов. В нескольких исследованиях использовались регрессионные модели для прогнозирования паттернов активации мозга, связанных с концепцией данного объекта, в соответствии с тем, как разные воксели "настроены" на различные размерности объекта и насколько важны эти размерности для определения концепции данного объекта. Были сделаны достаточно точные прогнозы с использованием свойств объектов, описанных участниками (Chang et al., 2011) или извлеченных из текстовых баз данных, таких как статьи в Интернете (Pereira et al., 2013). Обычная интерпретация этих данных предполагает, что сенсомоторные системы в мозге хранят или как-то обрабатывают информацию, которая является неотъемлемой частью понимания концепции объекта и включает информацию о взаимодействии тела с объектом, в соответствии с KBK. Альтернативные теории утверждают, что активация, наблюдаемая в сенсомоторных областях, отражает воображаемые образы, или моделирование движения, которое происходит уже после концептуальной обработки, а фундаментальное значение концепции кодируется только в ассоциативных областях, таких как передняя медиальная височная доля. Однако исследования с использованием слов, относящихся к конкретным объектам, показали, что сенсомоторная активация появляется слишком рано, чтобы происходить из образов, генерируемых воображением после извлечения смысла (Kieferet al., 2008), что свидетельствует в пользу первой интерпретации.

Кодирование эмоций

В качестве примера типичного подхода к анализу данных с использованием МАП рассмотрим работу (Kassam et al., 2013), авторы которой поставили себе целью выявить паттерны кодирования в мозге основных эмоций. В качестве испытуемых использовали 10 студентов, обучающихся драматическому мастерству. До эксперимента их просили написать сценарии для каждого из 18 эмоциональных слов и оценить каждый сценарий по семибалльной шкале на валентность, уровень возбуждения (arousal), уверенности, контроля и внимания, а также по 9 эмоциональным категориям. Во время эксперимента с записью фМРТ задачей испытуемых было активно испытывать эмоции в ответ на предъявленное на мониторе ключевое слово в соответствии с созданными ранее сценариями. Каждое из 18 слов предъявлялось 6 раз в случайном порядке. Анализ фМРТ-данных начинался с выбора вокселей с наиболее стабильным профилем активации в ответ на стимулы. У каждого испытуемого было выбрано 240 таких вокселей. Далее на первом шаге был использован один из методов машинного обучения (Gaussian Naïve Bayes) на части данных (4 из 6 предъявлений) для тренировки классификатора, задачей которого было выявить паттерн активации выбранных вокселей, соответствующий каждой из 18 эмоциональных категорий. На втором шаге классификатор тестировался на оставшейся части данных. Оба шага повторялись для всех возможных сочетаний обучающих и тестовых трайлов, что позволило оценить внутрииндивидуальную (within-subject) стабильность выявляемых паттернов. Для тестирования соответствия паттернов у разных участников (be-

tween-subject) классификатор тренировался на всех минус один участниках и тестировался на оставшемся с повторением этой процедуры соответствующее количество раз (Leave-One-Out Cross-Validation). Результаты внутриндивидуального тестирования показали среднюю точность предсказаний эмоциональной категории 0.84 (при 0.5-уровне случайных совпадений). При межиндивидуальном тестировании средняя точность была 0.7 (варьировала от 0.51 до 0.81) и предсказания были достоверно лучше уровня случайных совпадений для 8 из 10 участников. Воксели, внесшие вклад в успешные предсказания, были широко распределены по всему мозгу, включая большую их часть во фронтальных и орбитальных областях коры (что нетипично для нейросемантической классификации физических объектов). В частности, успешность классификации не зависела от включения или исключения затылочной коры, свидетельствуя о том, что визуальная форма слова (например, его длина) не играла принципиальной роли в классификации. Хотя наиболее точная классификация наблюдалась при включении в анализ вокселей из всех областей мозга, классификация достоверно выше уровня случайных совпадений была возможна при использовании лишь вокселей из любой изолированной области, включая лобную, теменную, височную, затылочную кору или подкорковые области, что согласуется с данными метаанализов, показывающих, что переживание эмоций связано с активацией широко распределенных сетей мозга (Lindquist et al., 2012).

Отдельной задачей работы было определение смысловых размерностей выявленной активации мозга. Эта задача решалась с помощью двухуровневого эксплораторного факторного анализа паттернов активации. На первом уровне факторный анализ выполнялся для каждого испытуемого отдельно, используя матрицу корреляций между активностью 600 стабильных вокселей и принадлежностью стимула к одной из 18 эмоциональных категорий. 600 вокселей выбирались из шести билатеральных областей мозга (лобная, височная, теменная и затылочная доля, поясная кора и подкорковые области, за исключением мозжечка). Факторные оценки 10 первых факторов от каждого испытуемого использовались для факторного анализа второго уровня на всех испытуемых вместе. Анализ выявил четыре фактора, кодирующие релевантные эмоциям концепции, а также пятый фактор, кодирующий длину стимулирующего слова (были извлечены два дополнительных фактора, но оказалось, что их трудно интерпретировать). Для интерпретации смысла этих факторов использовались четыре источника информации: 1) факторные оценки, показывающие, как 18 эмоций были ранжированы по данному фактору; 2) расположение вокселей, лежащих в основе каждого фактора, и предыдущие фМРТ-исследования, показывающие дифференциальную активацию в этих местах; 3) корреляция оценок факторов с оценками эмоций участниками по параметрам валентности и возбуждения; и 4) корреляция оценок факторов и оценок слов-эмоций независимой онлайн-выборкой из 60 участников. Как и в любом факторном анализе, интерпретация смысла факторов в некоторой степени субъективна. Фактор, объясняющий наибольший процент вариации, по-видимому, кодировал валентность эмоции. Положительные слова имели положительную оценку этого фактора, а отрицательные – отрицательную. Его факторные оценки почти идеально коррелировали с оценками приятности эмоциональных сценариев, сделанными участниками вне сканера. Области мозга, лежащие в основе этого фактора, также соответствовали интерпретации валентности, включая медиальные лобные области, участвующие в основных аффектах и регуляции эмоций, а также орбитальные лобные и средние области мозга, связанные с оценкой аффективной значимости. Второй фактор, по-видимому, коррелировал с возбуждением или подготовкой к действию. Его факторные оценки коррелировали с субъективными оценками возбуждения. Категории гнева, страха и похоти получили самые высокие оценки, тогда как грусть, стыд и гордость получили самые низкие оценки. В мозге этот фактор соответствовал активности базальных ганглиев и прецентральной извилины, участвующих в подготовке к действию, а также медиальной лобной области. Третий фактор кодировал, есть ли у эмоции социальный элемент (то есть другой человек). Физическое отвращение, которое с меньшей вероятностью касается другого человека, имело самые высокие оценки, тогда как ревность, зависть и похоть, требующие участия другого человека, имели самые низкие оценки. Воксели, лежащие в основе социального фактора, в основном локализовались в передней и задней частях по-

ясной коры, принадлежащих к дефолтной сети. Четвертый фактор, по-видимому, однозначно определял вожделение, отделяя его от других категорий эмоций. Нервные области, связанные с этим фактором, включали веретенообразную извилину и нижние лобные области, участвующие в восприятии лиц, а также области, о которых ранее сообщалось, что они участвуют в обработке сексуальных стимулов. Наконец, пятый фактор кодировал длину слова. Его факторные оценки коррелировали с длиной стимулирующих слов, а нервные области, лежащие в его основе, локализовались только в затылочной коре. Данные этого исследования вносят некоторый вклад в разрешение давнего спора в психологии эмоций между сторонниками концепции базовых эмоций и конструктивистами, рассматривающими эмоции как эмерджентные сущности (Barrett, Russell, 2015). Хотя, в соответствии с теорией базовых эмоций, в этой работе показано существование специфических для каждой эмоции паттернов активации в мозге, факторная структура этой активации и ее широкий охват разных областей мозга больше согласуются с пониманием эмоций как эмерджентных сущностей в рамках теории конструктивизма.

Кодирование конкретных и абстрактных концепций

Похожий анализ паттернов активации, вызванной обработкой 60 объектных концепций, выявил три главные размерности: манипулируемость (например, орудия труда), еда (например, овощи или кухонная утварь) и укрытия (например, жилища и автомобили), – которые были связаны с активацией левой постцентральной и нижней височной извилины, левой нижней височной и нижней лобной извилины, и билатеральной парагиппокампальной извилины соответственно (Just et al., 2010). В другом исследовании анализировали фМРТ-данные, записанные в процессе многочасового просмотра фильмов. Факторный анализ паттернов активации мозга выявил четыре интерпретируемые размерности: подвижность, социальность, цивилизация и биологические объекты (Huth et al., 2012).

Важный вопрос – где кодируется соотношение семантических размерностей, определяющее смысл индивидуальной концепции. По логике КВК не должно быть отдельного места для кодирования этого смысла. Слабый вариант КВК допускает, однако, существование зон конвергенции. Эмпирические данные не дают определенного ответа на этот вопрос. В некоторых работах объединение размерностей выявлено в тех же областях, которые кодируют и сами размерности (например, (Seymouret al., 2009)), в других – в передней части височной доли (Coutanche, Thompson-Schill, 2014).

Одним из самых значимых открытий нейросемантики является то, что паттерн активации, соответствующий данному понятию, во многом одинаков у разных людей. Когда два человека думают о понятии “яблоко”, их паттерны активации распределяются в одинаковых областях мозга и очень похожи. Этот феномен общности был продемонстрирован для нейронных репрезентаций конкретных объектов (Just et al., 2010), эмоций (Kassam et al., 2013), чисел (Damarla, Just, 2013) и социальных взаимодействий (Just et al., 2014). Способность людей точно классифицировать концепции предполагает, что они используют одни и те же свойства данного понятия. Если бы это было не так, людям было бы трудно найти общий язык. Показано, что наиболее определяющие свойства концепции автоматически активируются в любом случае вызова этой концепции, даже во время задач, для которых эта информация не имеет отношения к делу (Hsu et al., 2014).

Наибольший интерес и наибольший вызов для КВК представляет кодирование абстрактных концепций. Традиционный взгляд на абстрактность предполагает отсутствие перцептивной основы, в противоположность конкретности (Brysbaert et al., 2017). Можно ли тогда ожидать, что нейронные репрезентации абстрактных, так же как и конкретных концепций, будут базироваться в сенсорных и моторных областях мозга? Прямое сравнение кодирования конкретных и абстрактных концепций показало, что левая нижняя лобная извилина (ЛНЛИ) была среди небольшого набора областей мозга, которые позволяли классификатору распознавать паттерны активации, соответствующие абстрактным понятиям (Wang et al., 2013). Этот результат согласуется с данными метаанализа ранних работ, выполненных методами СПК, показывающими, что различие в картинах активации мозга при процессинге конкретных и абстрактных концепций наиболее выражено в ЛНЛИ, которая более активна именно в случае абстрактных концепций (Wang et al., 2010). Этот ре-

зультат оставляет открытым вопрос, какую информацию собственно кодирует ЛНЛИ, которая, помимо ее роли в лингвистических процессах, связана с фонологической рабочей памятью и процессами торможения (Fecteau et al., 2007).

Первые исследования, выполненные методами МАП, показывают, что паттерны активации мозга позволяют классифицировать набор абстрактных концепций в соответствии с их таксономическими категориями (Anderson et al., 2017). Показано, что для абстрактных концепций извлеченные из активации мозга семантические размерности отличаются от таковых, выявленных при изучении конкретных концепций. Так, Vargas и Just (2019), исследуя по фМРТ-данным паттерны активации, соответствующие 28 абстрактным понятиям (например, “этика”, “правда”, “духовность”), выявили три основные семантические размерности, которые соответствовали: 1) степени вербальной репрезентации концепции; 2) тому, была ли концепция внешней (или внутренней) по отношению к человеку; и 3) тому, имела ли концепция социальное содержание. Области мозга, связанные с первой размерностью, были те же (в основном ЛНЛИ), что и выявленные в метаанализе (Wang et al., 2010) при сравнении конкретных и абстрактных концепций. По гипотезе Vargas и Just (2020), ЛНЛИ может участвовать в интеграции смыслов данной концепции в множестве контекстов.

Некоторые авторы выделяют особый тип концепций, который они называют “гибридные” концепции, и которые лежат между конкретными и абстрактными концепциями (Vargas, Just, 2020). Эти концепции связаны с психологическими состояниями, которые можно испытывать, но не с непосредственным восприятием информации через органы чувств. К ним относятся эмоции, социальные и некоторые физические концепции. Нейросемантические исследования этих концепций выявляют паттерны активации, близкие к тем, которые выявляются при исследовании объектных концепций (Kassam et al., 2013; Mason, Just, 2016; Just et al., 2014).

Научные концепции – это особый тип абстрактных понятий, изучаемых только в процессе формального образования. Нейронные репрезентации абстрактных концепций физики (например, гравитация, крутящий момент, частота) можно разложить на значимые базовые нейронные и семантические размерности, несмотря на их абстрактность. В исследовании Mason и Just (2016) студентов физических факультетов просили думать о смысле 30 концепций из области классической физики. Использование мультивоксельного классификатора на базе машинного обучения позволило идентифицировать отдельные концепции со средней ранговой точностью 0.75. Классификатор, обученный на данных всех, кроме одного, участников идентифицировал концепции по данным оставшегося участника со средней точностью 0.71. Факторный анализ фМРТ-данных позволил выделить четыре основные размерности, лежащие в основе нейронного представления этих понятий: причинность, периодичность, алгебраическое представление (наличие количественных отношений между понятиями) и поток энергии, все из которых, по данным других работ, используются и для представления конкретных концепций. В работе (Mason et al., 2021) в качестве испытуемых использовали профессиональных физиков и в набор концепций включили в высшей степени абстрактные понятия из квантовой механики и астрофизики. Помимо уже описанных в предыдущем исследовании размерностей периодичности и алгебраического представления, были выделены дополнительные размерности, такие как нематериальность (невозможность непосредственного наблюдения), согласованность с классическими концепциями, размышление о каузальности в отношении ненаблюдаемых объектов и многоуровневая организация знания. В работе (Wang et al., 2021) показано, что математические, физические и химические концепции были связаны с похожими паттернами активации в зрительно-пространственной и семантической сетях, указывая на то, что эти сети играют ключевую роль в процессинге как математической, так и научной информации. Эти данные свидетельствуют о том, что абстрактные научные концепции представлены перепрофилированными нейронными структурами, которые первоначально развивались для более общих целей. Основные возможности мозга, лежащие в основе физических концепций, существовали задолго до того, как были развиты знания физики и математики (Bauer, Just, 2019). В целом, несмотря на достигнутые успехи, механизмы нейронной репрезентации абстрактных концепций и соответствие этих механизмов слабой или сильной версии

КВК остаются предметом дебатов (Conca et al., 2021; Dove, 2021).

Итоги исследований нейросемантики

Итак, если подвести итог обзору исследований нейросемантики, то можно выделить три темы: 1) теоретическая основа – когнитивизм, КВК и их варианты; 2) методические подходы к сбору и анализу данных; 3) основные результаты. Теоретической основой исследований в нейросемантике является КВК, которая постулирует, что корни наших знаний о мире лежат во взаимодействии с этим миром. Соответственно, ожидается, что концепция некоего объекта представлена в мозге в виде распределенной активности всех областей мозга, которые активируются при восприятии и взаимодействии с этим объектом. Это ожидание кажется менее оправданным в случае абстрактных концепций, связь которых с сенсорными и двигательными процессами менее очевидна. Альтернативная КВК позиция, которая сейчас в чистом виде уже всерьез не рассматривается, утверждает, что извлечение смысла связано со специализированными модулями мозга, отделенными от сенсорных и двигательных модулей. Чаще всего в качестве такого специализированного модуля рассматривается передняя часть височной доли (ПЧВД). Гибридная теория, которую называют слабым вариантом КВК, признает роль так называемых зон конвергенции, в которых интегрируется информация от сенсорных и двигательных зон коры. Эти зоны могут находиться в ПЧВД или в лингвистических областях мозга (например, левая нижняя лобная извилина), или в смежных с сенсорными областями, более фронтально расположенных участках мозга. Считается, что роль зон конвергенции особенно велика в случае абстрактных концепций. Разделение между абстрактными и конкретными концепциями не абсолютно, признается существование гибридных концепций, которые не связаны напрямую с сенсорными и двигательными процессами, но связаны с состояниями, которые человек может чувствовать в виде некоторых телесных проявлений. То есть конкретные и гибридные концепции более явно “воплощены”, чем чисто абстрактные.

Методы анализа данных в рамках МАП более чувствительны, чем традиционные методы СПК, но их главная особенность состоит в том, что они основаны на классификации паттернов активации мозга при обработке определенного набора концепций. То есть классификатор выбирает такие паттерны, которые лучше всего позволяют отличить друг от друга концепции из данного набора. Например, воксели, которые активируются при обработке всех или большинства концепций из набора, не будут выявляться классификатором, хотя они и участвуют в обработке каждой концепции. То же можно сказать про факторный анализ, результаты которого определяются матрицей корреляций между активацией вокселей и данным набором концепций. Достаточно посмотреть на интерпретацию выявляемых размерностей (например, манипулируемость, еда и укрытия (Just et al., 2010)), чтобы увидеть, что они отнюдь не универсальны и имеют смысл только в пределах рассмотренного в данном исследовании набора объектных концепций. Вряд ли в пространстве этих размерностей можно найти место для концепций справедливости или добра. Таким образом, получаемые результаты зависят от того, какие концепции были использованы в наборе, и для какой-то определенной концепции эти результаты могут быть другими, если ее исследовать в наборе с другими концепциями. Это в какой-то степени видно из сравнения результатов работ (Mason, Just, 2016) и (Mason et al., 2021), где добавление в набор концепций понятий из квантовой механики вызвало исчезновение двух ранее описанных и появление четырех новых размерностей. Кроме того, надо помнить, что интерпретация результатов факторного анализа неизбежно субъективна и всегда есть искушение подогнать ее под существующие теории и концепции. Тем не менее методы МАП, в отличие от СПК, не ограничены идеологией нервных центров и во многом произвольным выбором размеров достоверных кластеров.

О чем же говорят результаты этих исследований, если забыть обо всех перечисленных ограничениях и интерпретировать их “в лоб”? Прежде всего эти данные говорят о том, что информация, необходимая для идентификации отдельных концепций из использованного набора, в мозге присутствует, но связь между семантикой и активностью мозга имеет вероятностный характер. Точность идентификации концепций на основе активности мозга в среднем равна 0.7 при уровне случайности 0.5. Эта вариативность

видна при анализе как внутри- так и межиндивидуальной воспроизводимости результатов. Несмотря на отмечаемую во всех работах схожесть паттернов активности у разных испытуемых, средняя точность предсказания опять равна 0.7, а в некоторых случаях (2 из 10 в работе (Kassam et al., 2013)) предсказания были недостоверны.

Во-вторых, в активности мозга любая концепция представлена в виде широко распределенного ансамбля вокселей, охватывающего и сенсомоторные (в соответствии с КВК), и вербальные (в соответствии с когнитивизмом), и многие другие области ассоциативной коры. Не надо забывать, что в процессе факторного анализа извлекается лишь небольшое количество факторов, поддающихся интерпретации и объясняющих меньше половины вариации активности вокселей. Большая часть этой активности не поддается интерпретации. Эти данные вместе с теорией КВК предполагают, что воспроизведение любой концепции сопровождается активацией всех чувственных, моторных и прочих знаний, имеющих к ней отношение. Предполагается, что эта активация необходима для извлечения смысла концепции. Однако в повседневной жизни, если не ставится задача вспомнить и прочувствовать все аспекты концепции, смысл извлекается мгновенно и незаметно. На уровне сознания мы не испытываем все те ощущения, которые должны были бы испытывать исходя из активации всех этих областей мозга, иначе мысль о лимоне сопровождалась бы галлюцинацией лимона, а мысль об ударе ногой произвела бы этот удар. Концепции извлекаются, как правило, не сами по себе, а в связи со многими другими концепциями в процессе обсуждения или обдумывания какой-то проблемы с привлечением логических связей и причинно-следственных отношений. Вряд ли это было бы возможно, если бы понимание смысла каждой концепции требовало “прочувствования” всех ее аспектов. Вся эта массивная нервная активация остается за пределами сознания, и в сознании возникает лишь смысл, непонятным способом извлеченный из этой активности. По одной из гипотез отсутствие ощутимого “проигрывания” соответствующих ощущений можно объяснить тем, что основная активация происходит в зонах конвергенции, а сами сенсомоторные области активируются гораздо слабее (Yee et al., 2013), однако эмпирические данные в поддержку этой гипотезы пока отсутствуют.

По наиболее популярным теориям, для извлечения смысла важна не активация всех этих нейронных популяций сама по себе, а их взаимодействие друг с другом, то есть возникновение сети взаимосвязанных посредством функциональных связей ансамблей (см., например, обзор (Palacio, Cardenas, 2019)). Это направление развивается в рамках коннекционистских моделей семантических репрезентаций (см., например, обзор (Jones et al., 2015)), параллельно с дистрибутивными моделями, которые мы частично рассматривали при разборе примеров из нейролингвистики. Эти модели используются при разработке архитектуры нейронных сетей. Я не буду здесь рассматривать эту тему, так как, несмотря на немалое количество эмпирических исследований, тестирующих адекватность этих моделей для описания процессов в мозге, пока нет понимания того, как язык компьютерных нейронных сетей можно транспонировать в язык динамики процессов в мозге.

По одной из гипотез, объединение нейронных ансамблей в мозге может достигаться синхронизацией осцилляций электрической активности, например, в диапазоне гамма-ритма, обеспечивая таким образом пресловутое связывание (binding) и интеграцию разнородной информации в единый холистический смысл (Varela, 2000). Подобные теории кажутся привлекательными и действительно дают намек на объяснение того, как единый смысл может возникать из разнородной активности нейронных популяций. Эти теории, однако, помимо отсутствия однозначной эмпирической поддержки, оставляют неотвеченным целый ряд сопутствующих вопросов. Каков механизм этой синхронизации, что ее вызывает и что ей управляет? Как и где представлен смысл, извлеченный из синхронной активности нейронных популяций?

Общее обсуждение и философские выводы

Таким образом, мы рассмотрели несколько примеров репрезентации семантической информации. Во всех этих примерах есть немало общего, что позволяет увидеть общие принципы организации отражения семантики в активности мозга. Прежде всего, активность мозга, сопровождающая извлечение смысла, имеет вероятностный характер, и речь

тут идет не об ошибках, связанных с погрешностью методов, а о природе самой нервной активности. Как из этой вероятностной картины возникает определенность, присущая нашему восприятию реального мира и наших действий в нем, остается неясным.

Конечно, можно думать, что и содержание ментальности в процессе одного и того же эксперимента может варьировать как у разных испытуемых, так и у одного и того же человека при повторных тестированиях. Однако эта вариация не связана с содержанием экспериментального задания и, в принципе, может контролироваться, например, путем опроса испытуемых после эксперимента (например, (Knyazev et al., 2012)). Связанное с экспериментальным заданием 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 определенно и лучше связано с поведением, чем активность мозга.

Вторая характеристика активности мозга, сопровождающей ментальные операции, – это ее распределенный характер. Чем сложнее когнитивный процесс, тем большее количество областей мозга активируется. Эти области, как правило, далеко друг от друга расположены и участвуют в большом количестве разнородных процессов, включая сенсорное восприятие, движение и лингвистические процессы. Метаанализ большого 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 подавляющее большинство областей мозга активируется в подавляющем 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 показывают, что средняя вовлеченность для 78 больших областей мозга составляет 0.7. Авторы заключают, что локальные нейронные цепи не очень избирательны и обычно участвуют во множестве задач из разных когнитивных доменов. Для функциональных связей показатель вовлеченности был меньше, чем для активации, то есть при выполнении задач из разных когнитивных доменов одна и та же область мозга может взаимодействовать с разным набором других областей. Однако часто наблюдается похожесть паттернов коактивации при выполнении задач, казалось бы, очень непохожих друг на друга. По предположению авторов, это свидетельствует о том, что 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). Как эта, по видимости неспецифическая, вероятностно распределенная активность порождает содержание сознания, которое, как правило, определенно и имеет холистический характер (Curby, Moerel, 2019; Jin et al., 2021; Poltoratski et al., 2021)?

Можно думать, что содержание каждого состояния сознания определяется уникальным паттерном активности широко распределенных нейронных ансамблей, разные единицы которых (это могут быть отдельные колонки или даже отдельные пирамидные нейроны) связаны друг с другом сетью рекуррентных возбуждающих и тормозных влияний с разными весовыми коэффициентами (наподобие того, как это устроено в искусственных нейронных сетях). Топографию и поведение таких ансамблей можно в некотором приближении описать с помощью теоретических (таких как КВК) и математических (коннекционистские, распределенные, Байесовские) вероятностных моделей. Нет, однако, даже намека на то, что из этих описаний можно “вычислить” конкретное содержание состояния сознания. Еще раз подчеркну, что относительная успешность методов МАП в распознавании паттернов активности, соответствующих определенной концепции из данного набора концепций, никак не приближает нас к тому, чтобы теоретически предсказывать содержание сознания из известных параметров активности мозга.

Если вспомнить предложенную Криком и Кохом примерно четверть столетия назад программу научного решения проблемы сознания, то можно констатировать, что наука далеко продвинулась в плане выявления нервных коррелятов (отдельных элементов) сознания. Однако выявление коррелятов в

принципе не может ответить на вопрос о причинной связи коррелирующих явлений. Как процессы, происходящие в мозге, "порождают" содержание сознания? Как происходит категоризация сенсорных образов? Как извлекается смысл и происходит узнавание образа? Как принимается сознательное решение? Безусловно, существует немало гипотез и теорий (Байесовская модель, нейронные сети, предсказывающее кодирование и так далее), которые пытаются объяснить, как мозг производит "вычисления", необходимые для обеспечения этих процессов, но эти теории (помимо скудости подтверждающих их эмпирических данных) сталкиваются с рядом принципиальных трудностей, одна из которых – высокая скорость и эффективность протекания этих процессов, которую трудно объяснить на основе известной нам в настоящее время физиологической базы, предположительно лежащей в их основе (импульсная активность нейронов и синаптическая передача) (Gallistel, King, 2009).

Сейчас накоплены данные, показывающие, что ментальные процессы и процессы в мозге подчиняются разным законам и описываются разными математическими и логическими моделями. Классическая теория вероятности и булева логика хорошо описывают активность мозга, но для описания когнитивных процессов лучше подходит формализм квантовой теории и не-булева 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* (то есть ментальная сфера) есть не что иное, как *y* (то есть активность мозга). Выполнение этих требований необходимо для утверждения о сводимости ментального к физическому (Brigandt, Love, 2017; van Riel, Van Gulick, 2019). Можно заключить поэтому, что на основе имеющихся в настоящее время данных редукционизм вряд ли можно рассматривать как жизнеспособную опцию. Слабый эмерджентизм также кажется малоправдоподобным, так как он предполагает высокую степень соответствия между состояниями мозга и состояниями сознания, которая в свете хорошо известной как внутри-, так и межиндивидуальной вариативности этих процессов явно не выполняется. Единственная оставшаяся опция – сильный эмерджентизм – в принципе совместима с имеющимися эмпирическими данными, однако ее совместимость с парадигмальным материализмом вызывает сомнения (Bedau, 1997).

ФИНАНСИРОВАНИЕ

При написании этой статьи автор получал финансовую поддержку из бюджетной темы НИИНМ № АААА-А21-121011990039-2 и Российского фонда фундаментальных исследований (проект № 20-013-00404).

СПИСОК ЛИТЕРАТУРЫ

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.

Anderson A.J., Kiela D., Clark S., Poesio M. Visually grounded and textual semantic models differentially decode brain activity associated with concrete and abstract nouns. Transactions of the Association for Computational Linguistics. 2017. 5: 17–30.

Anderson M.L., Penner-Wilger M. Neural reuse in the evolution and development of the brain: Evidence for developmental homology? Developmental Psychobiology. 2013. 55: 42–51.

Barrett L.F., Russell J.A. The Psychological Construction of Emotion. 2015. New York: Guilford Press.

Barsalou L.W. Grounded cognition. Annu. Rev. Psychol. 2008. 59: 617–645.

Barsalou L.W., Simmons W.K., Barbey A.K., Wilson C.D. Grounding conceptual knowledge in modality-specific systems. Trends in Cognitive Sciences. 2003. 7: 84–91.

Bauer A.J., Just M.A. Neural representations of concept knowledge. The oxford handbook of neurolinguistics. 2019. 1–21.

Bedau M. Weak Emergence. Philosophical Perspectives, Mind, Causation, and World. 1997. Oxford: Blackwell, pp. 375–399.

Bergen B.K. Louder than Words: The New Science of how the Mind makes Meaning. 2012. New York City, NY: Basic Books.

Binder J.R., Desai R.H. The neurobiology of semantic memory. Trends Cogn. Sci. 2011. 15: 527–536.

Block N. On a Confusion about a Function of Consciousness. Behavioral and Brain Sciences. 1995. 18: 227–247.

Block N. If Perception Is Probabilistic, Why Does It Not Seem Probabilistic? Philosophical Transactions of the Royal Society B: Biological Sciences. 2018. 373(1755): 20170341.

Brigandt I., Love A. Reductionism in Biology. The Stanford Encyclopedia of Philosophy. 2017. Edward N.

Zalta (ed.): https://plato.stanford.edu/archives/spr2017/entries/reduction-biology/.

Brysbaert M., Warriner A.B., Kuperman V. Concreteness ratings for 40 thousand generally known English word lemmas. Behavior Research Methods. 2014. 46(3): 904–911.

Burge T. Origins of objectivity. 2010. Oxford, UK: Oxford University Press.

Casasanto D. The hierarchical structure of mental metaphors. In Metaphor: Embodied Cognition and Discourse. 2017. Ed. *B. Hampe*, Cambridge: Cambridge University Press.

Chalmers D.J. The Conscious Mind. 1996. Oxford University Press.

Chalmers D.J. Strong and Weak Emergence. In (*P. Clayton and P. Davies*, eds.) The Re-emergence of Emergence. 2006. Oxford University Press.

Chang K.K., Mitchell T., Just M.A. Quantitative modeling of the neural representation of objects: How semantic feature norms can account for fMRI activation. NeuroImage. 2011. 56(2): 716–727.

Chang L., Tsao D.Y. The Code for Facial Identity in the Primate Brain. Cell. 2017. 169: 1013–1028.

Chao L.L., Martin A. Representation of manipulable man-made objects in the dorsal stream. NeuroImage. 2000. 12: 478–484.

Churchland P.S. Neurophilosophy: Toward a Unified Science of the Mind-Brain. 1986. Cambridge, Massachusetts: The MIT Press.

Conca F., Borsa V.M., Cappa S.F., Catricalà E. The multidimensionality of abstract concepts: A systematic review. Neuroscience & Biobehavioral Reviews. 2021. https://doi.org/10.1016/j.neubiorev.2021.05.004

Coutanche M.N., Thompson-Schill S.L. Creating concepts from converging features in human cortex. Cerebral Cortex. 2014. 25: 2584–2593.

Crick F. The Astonishing Hypothesis: The Scientific Search for the Soul. Scribner: 1994.

Crick F., Koch C. Consciousness and neuroscience. Cerebral Cortex. 1998. 8(2): 97–107.

Crick F., Koch C. A framework for consciousness. Nature neuroscience. 2003. 6(2): 119–126.

Crick F.C., Koch C. What is the function of the claustrum? Phil. Trans. R. Soc. Lond. B. 2005. 380(1458): 1271–1279.

Curby K.M., Moerel D. Behind the face of holistic perception: Holistic processing of Gestalt stimuli and faces recruit overlapping perceptual mechanisms. Attention, Perception, & Psychophysics. 2019. 81(8): 2873–2880.

Damarla S.R., Just M.A. Decoding the representation of numerical values from brain activation patterns. Human Brain Mapping. 2013. 34(10): 2624–2634.

Damasio A.R., Damasio H. Cortical systems for retrieval of concrete knowledge: the convergence zone framework. In: Large-scale Neuronal Theories of the Brain. 1994. Ed. *C. Koch* (Cambridge, MA: MIT Press): 61–74.

Dennett D. Why You Can't Make a Computer that Feels Pain. Brainstorms. 1978. Cambridge, MA, MIT Press: 190–229.

Dennett D. Consciousness Explained. 1991. The Penguin Press.

Desai R.H., Conant L.L., Binder J.R., Park H., Seidenberg M.S. A piece of the action: modulation of sensory-motor regions by action idioms and metaphors. Neuroimage. 2013. 83: 862–869.

di Oleggio Castello M.V., Haxby J.V., Gobbini M.I. Shared neural codes for visual and semantic information about familiar faces in a common representational space. Proceedings of the National Academy of Sciences. 2021. 118(45).

Ding S., Cueva C.J., Tsodyks M., Qian N. Visual perception as retrospective Bayesian decoding from high-to low-level features. Proceedings of the National Academy of Sciences. 2017. 114(43): E9115–E9124.

Dove G. Three symbol ungrounding problems: abstract concepts and the future of embodied cognition. Psychon. Bull. Rev. 2016. 23: 1109–1121.

Dove G. The Challenges of Abstract Concepts. Handbook of Embodied Psychology. 2021. 171–195.

Fecteau S., Pascual-Leone A., Zald D.H., Liguori P., Théoret H., Boggio P.S., Fregni F. Activation of prefrontal cortex by transcranial direct current stimulation reduces appetite for risk during ambiguous decision making. J Neurosci. 2007. 27(23): 6212–6218.

Firth J.R. A synopsis of linguistic theory 1930–1955. Studies in Linguistic Analysis. 1957. 1–32.

Frankish K. Illusionism as a Theory of Consciousness. Journal of Consciousness Studies. 2016. 23: 11–39.

Friston K., Kilner J., Harrison L. A free energy principle for the brain. Journal of Physiology. 2006. Paris, 100(1–3): 70–87.

Galetzka C. The story so far: How embodied cognition advances our understanding of meaning-making. Frontiers in psychology. 2017. 8: 1315.

Gallistel C.R., King A. Memory and the Computational Brain. 2009. Malden: Wiley-Blackwell.

Grill-Spector K., Knouf N., Kanwisher N. The fusiform face area subserves face perception, not generic within-category identification. Nature Neuroscience. 2004. 7(5): 555–562.

Gross C. Single neuron studies of inferior temporal cortex. Neuropsychologia. 2008. 46: 841–852.

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. Supplementary volume 70: 1–21.

Hart J., Kraut M.A. Neural hybrid model of semantic object memory. In *J. Hart & M. A. Kraut* (Eds.),

Neural basis of semantic memory (pp. 331–359). 2007. New York, Cambridge University Press.

Hauk O., Tschentscher N. The body of evidence: what can neuroscience tell us about embodied semantics? Front. Psychol. 2013. 4: 50.

Hawking S.W. A Brief History of Time. Bantam Dell Publishing Group. 1988. NY.

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. Journal of Cognitive Neuroscience. 2011. 23: 2554–2557.

Hsu N.S., Schlichting M.L., Thompson-Schill S.L. Feature diagnosticity affects representations of novel and familiar objects. Journal of Cognitive Neuroscience. 2014. 26: 2735–2749.

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. 532(7600): 453–458.

Huth A.G., Nishimoto S., Vu A.T., Gallant J.L. A continuous semantic space describes the representation of thousands of object and action categories across the human brain. Neuron. 2012. 76: 1210–1224.

Jazayeri M., Movshon J.A. Optimal representation of sensory information by neural populations. Nat. Neurosci. 2006. 9: 690–696.

Jin H., Oxner M., Corballis P.M., Hayward W.G. Holistic face processing is influenced by non-conscious visual information. British Journal of Psychology. 2021.
https://doi.org/10.1111/bjop.12521

Jones M.N., Willits J., Dennis S., Jones M. Models of semantic memory. Oxford handbook of mathematical and computational psychology. 2015. 232–254.

Just M.A., Cherkassky V.L., Aryal S., Mitchell T.M. A neurosemantic theory of concrete noun representation based on the underlying brain codes. PLOS One. 2010. 5(1): e8622.

Just M.A., Cherkassky V.L., Buchweitz A., Keller T.A., Mitchell T.M. Identifying autism from neural representations of social interactions: Neurocognitive markers of autism. PLOS One. 2014. 9(12): e113879.

Kanwisher N., McDermott J., Chun M.M. The fusiform face area: a module in human extrastriate cortex specialized for face perception. The Journal of Neuroscience. 1997. 17(11): 4302–4311.

Kassam K.S., Markey A.R., Cherkassky V.L., Loewenstein G., Just M.A. Identifying Emotions on the Basis of Neural Activation. PLoS ONE. 2013. 8(6): e66032.

Kay K.N., Naselaris T., Prenger R.J., Gallant J.L. Identifying natural images from human brain activity. Nature. 2008. 452(7185): 352–355.

Kiefer M., Sim E.-J., Herrnberger B., Grothe J., Hoenig K. The sound of concepts: Four markers for a link between auditory and conceptual brain systems. The Journal of Neuroscience. 2008. 28(47): 12224–12230.

Knyazev G.G., Slobodskoj-Plusnin J.Y., Bocharov A.V. Subject's State Modulates Oscillatory Responses to Emotional Facial Expressions. Journal of Psychophysiology. 2012. 26: 83–91.

Koch C. Reflections of a Natural Scientist on Panpsychism. Journal of Consciousness Studies. 2021. 28(9-10): 65–75.

Lambon Ralph M.A., Jefferies E., Patterson K., Rogers T.T. The neural and computational bases of semantic cognition. Nat. Rev. Neurosci. 2016. 18: 42–55.

Lee J., Maunsell J.H.R. A Normalization Model of Attentional Modulation of Single Unit Responses. PLoS ONE. 2009. 4(2): e4651.

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. 2021.
https://doi.org/10.1016/j.neuron.2021.08.022

Lilienfeld S., Lynn S.J., Namy L., Woolf N. Psychology: A Framework for Everyday Thinking. 2010. Pearson.

Lindquist K.A., Wager T.D., Kober H., Bliss-Moreau E., Barrett L.F. The brain basis of emotion: A meta-analytic review. Behav. Brain Sci. 2012. 35: 121–143.

Ma W.J., Beck J.M., Latham P.E., Pouget A. Bayesian inference with probabilistic population codes. Nat. Neurosci. 2006. 9: 1432–1438.

Mason R.A., Just M.A. Neural representations of physics concepts. Psychological Science. 2016. 27(6): 904–913.

Mason R.A., Schumacher R.A., Just M.A. The neuroscience of advanced scientific concepts. NPJ science of learning. 2021. 6(1): 1–12.

Metzinger T. Neural Correlates of Consciousness: Empirical and Conceptual Questions. 2000. Cambridge, MA: MIT Press.

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. 320: 1191–1195.

Nagel T. What Is It Like to Be a Bat? The Philosophical Review, 1974. 83(4): 435–450.

Naselaris T., Kay K.N., Nishimoto S., Gallant J.L. Encoding and decoding in fMRI. Neuroimage. 2011. 56(2): 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. 63(6): 902–915.

O'Connor T. Emergent Properties. The Stanford Encyclopedia of Philosophy. 2020. E.N. Zalta (ed.),

URL = <https://plato.stanford.edu/archives/fall2020/entries/properties-emergent/>.

Palacio N., Cardenas F. A systematic review of brain functional connectivity patterns involved in episodic and semantic memory. Reviews in the Neurosciences. 2019. 30: 889–902.

Papeo L., Lingnau A., Agosta S., Pascual-Leone A., Battelli, L. Caramazza A. The origin of word-related motor activity. Cereb. Cortex. 2014. 25, 1668–1675.

Pereira F., Botvinick M., Detre G. Using Wikipedia to learn semantic feature representations of concrete concepts in neuroimaging experiments. Artificial Intelligence. 2013. 194: 240–252.

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. Nature Communications. 2018. 9(1): 1–13.

Peters M.A.K., Thesen T., Ko Y.D., Maniscalco B., Carlson C., Davidson M., Doyle W., Kuzniecky R., Devinsky O., Halgren E., Lau H. Perceptual confidence neglects decision-incongruent evidence in the brain. Nat. Hum. Behav. 2017. 1: 0139.

Poltoratski S., Kay K., Finzi D., Grill-Spector K. Holistic face recognition is an emergent phenomenon of spatial processing in face-selective regions. Nature Communications. 2021. 12(1): 1–13.

Potter M.C., Staub A., O'Connor D.H. Pictorial and conceptual representation of glimpsed pictures. Journal of Experimental Psychology: Human Perception and Performance. 2004. 30: 478–489.

Purves D. Principles of Cognitive Neuroscience. 2008. Sunderland, Mass.: Sinauer Associates.

Quian Quiroga R. Neuronal codes for visual perception and memory. Neuropsychologia. 2016. 83: 227–241.

Quian Quiroga R. No Pattern Separation in the Human Hippocampus. Trends in Cognitive Sciences. 2020. 24: 994–1007.

Quian Quiroga R., Kreiman G. Measuring sparseness in the brain: comment on Bowers. 2009. Psychological Review. 2010. 117: 291–299.

Qiao K., Chen J., Wang L., Zhang C., Zeng L., Tong L., Yan B. Category decoding of visual stimuli from human brain activity using a bidirectional recurrent neural network to simulate bidirectional information flows in human visual cortices. Frontiers in neuroscience. 2019. 13: 692.

Ramsey W. Eliminative Materialism. The Stanford Encyclopedia of Philosophy. 2020. E.N. Zalta (ed.): https://plato.stanford.edu/archives/sum2020/entries/materialism-eliminative/.

Reppert V. Eliminative Materialism, Cognitive Suicide, and Begging the Question. Metaphilosophy. 1992. 23: 378–92.

Rescorla M. Bayesian Perceptual Psychology. In: The Oxford Handbook of Philosophy of Perception. 2015. *Mohan, M.* (ed.): Oxford, Oxford University Press, 694–716.

Rey H.G., Gori B., Chaure F.J., Collavini S., Blenkmann A.O., Seoane P., Seoane E., Kochen S., Quiroga R.Q. Single neuron coding of identity in the human hippocampal formation. Curr. Biol. 2020. 30: 1152–1159.

Richards B.A., Frankland P.W. The conjunctive trace. Hippocampus. 2013. 23: 207–212.

Robinson H. Dualism. The Stanford Encyclopedia of Philosophy. 2020. E.N. Zalta (ed.): https://plato.stanford.edu/archives/fall2020/entries/dualism/.

Rugg M.D., Thompson-Schill S.L. Moving forward with fMRI data. Perspectives on Psychological Science. 2013. 8(1): 84–87.

Schaffer J. Monism. The Stanford Encyclopedia of Philosophy. 2018. E.N. Zalta (ed.): https://plato.stanford.edu/archives/win2018/entries/monism/.

Searle J. The Mystery of Consciousness. 1997. New York: The New York Review of Books.

Seymour K., Clifford C.W.G., Logothetis N.K., Bartels A. The coding of color, motion, and their conjunction in the human visual cortex. Current Biology. 2009. 19(3): 177–183.

Shapiro L. Embodied cognition. 2019. Routledge.

Simonite T. Facebook Creates Software That Matches Faces Almost as Well as You Do. MIT Technology Review. 2014. https://www.technologyreview.com/2014/03/17/13822/facebook-creates-software-that-matches-faces-almost-as-well-as-you-do/

Smart J.J.C. Materialism. Encyclopedia Britannica. 2016. https://www.britannica.com/topic/materialism-philosophy.

Stubenberg L. Neutral Monism. The Stanford Encyclopedia of Philosophy. 2018. E.N. Zalta (ed.): https://plato.stanford.edu/archives/fall2018/entries/neutral-monism/.

Tsao D.Y., Freiwald W.A., Tootell R.B., Livingstone M. A cortical region consisting entirely of face-selective cells. Science. 2006. 311: 670–674.

Tschentscher N. Embodied Semantics: Embodied Cognition in Neuroscience. German Life and Letters. 2017. 70(4): 423–429.

van Bergen R.S., Ma W.J., Pratte M.S., Jehee J.F. Sensory uncertainty decoded from visual cortex predicts behavior. Nat. Neurosci. 2015. 18: 1728–1730.

van Riel R., Van Gulick R. Scientific Reduction. The Stanford Encyclopedia of Philosophy. 2019. E.N. Zalta (ed.): https://plato.stanford.edu/archives/spr2019/entries/scientific-reduction/.

Varela F.J. Neurophenomenology: A methodological remedy for the hard problem. Journal of Consciousness Studies. 1996. 3: 330–335.

Varela F. Upward and downward causation in the brain: Case studies on the emergence and efficacy of consciousness. No Matter, Never Mind. (2000). Ed. *K. Yasue, M. Jibu and T. Della Senta* (Amsterdam/Philadelphia: John Benjamins Publishing Co.).

Vargas R., Just M.A. Neural representations of abstract concepts: Identifying underlying neurosemantic dimensions. Cerebral Cortex. 2019. 30(4): 2157–2166.

Vargas R., Just M.A. Neural representations of abstract concepts: Identifying underlying neurosemantic dimensions. Cerebral Cortex. 2020. 30(4): 2157–2166.

Vukovic N., Feurra M., Shpektor A., Myachykov A., Shtyrov Y. Primary motor cortex functionally contributes to language comprehension: an online rTMS study. Neuropsychologia. 2017. 96: 222–229.

Wang J., Baucom L.B., Shinkareva S.V. Decoding abstract and concrete concept representations based on single-trial fMRI data. Human Brain Mapping. 2013. 34(5): 1133–1147.

Wang J., Conder J.A., Blitzer D.N., Shinkareva S.V. Neural representation of abstract and concrete concepts: A meta-analysis of neuroimaging studies. Human Brain Mapping. 2010. 31(10): 1459–1468.

Wang Z., Busemeyer J.R., Atmanspacher H., Pothos E.M. The Potential of Using Quantum Theory to Build Models of Cognition. Topics in Cognitive Sciences. 2013. 5: 672–688.

Wang L., Li M., Yang T., Zhou X. Mathematics Meets Science in the Brain. Cerebral Cortex. 2021. https://doi.org/10.1093/cercor/bhab198

Wen H., Shi J., Zhang Y., Lu K.-H., Cao J., Liu Z. Neural Encoding and Decoding with Deep Learning for Dynamic Natural Vision. Cerebral Cortex. 2018. 28: 4136–4160.

Wu M.C., David S.V., Gallant J.L. Complete functional characterization of sensory neurons by system identification. Ann. Rev. Neurosci. 2006. 29: 477–505.

Wurtz R.H. Recounting the impact of Hubel and Wiesel. J. Physiol. 2009. 587(12): 2817–2823.

Yee E., Chrysikou E.G., Thompson-Schill S.L. Semantic Memory. In Kevin Ochsner and Stephen Kosslyn (Eds), The Oxford Handbook of Cognitive Neuroscience, Volume 1: Core Topics (pp. 353–374). 2013. Oxford University Press.

Zihl J., Von Cramon D., Mai N. Selective Disturbance of Movement Vision after Bilateral Brain Damage. Brain. 1983. 106(2): 313–340.

CODING THE MEANING IN BRAIN ACTIVITY

G. G. Knyazev[#]

Federal State Budgetary Scientific Institution "Scientific Research Institute of Neurosciences and Medicine", Novosibirsk, Russia

[#]*e-mail: knyazev@physiol.ru*

The question of the nature of the relationship between mental processes and brain activity is not the prerogative of philosophers. All professional scientists studying the human brain must, in one way or another, decide for themselves this question. Dominant in the modern scientific worldview is the reductionist approach, according to which mental states can, in principle, be reduced to brain states. In this article, I review some of the data obtained in recent years that shed light on the nature of the relationship between the content of mental processes and brain activity. These data, concerning the encoding of sensory and semantic linguistic information, as well as more complex information related to the content of abstract concepts, show that the brain activity accompanying the extraction of meaning has a widely distributed probabilistic nature, which does not correspond to the nature of mental content, which is usually certain and has a holistic character. Thus, based on the currently available empirical data, reductionism can hardly be regarded as a viable option, and the only option within the framework of materialism is the recognition of mentality as an emergent entity.

Keywords: content of consciousness, brain activity, reductionism, semantics, fMRI, multivariate pattern analysis