МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
имени М.В. ЛОМОНОСОВА

ФИЛОСОФСКИЙ ФАКУЛЬТЕТ

На правах рукописи

Хасанов Рустам Юрьевич

Коннекционистский подход в современных когнитивных исследованиях

5.7.6. Философия науки и техники

ДИССЕРТАЦИЯ

на соискание учёной степени
кандидата философских наук

Научный руководитель:
доктор философских наук
Алексеев Андрей Юрьевич

МОСКВА – 2023

Оглавление

Введение

Актуальность темы исследования

В настоящее время методы когнитивной науки позволяют по-новому представить активность сетей нейронов в живом мозге и определить разнообразные функции, которые выполняются биологическими сетями нейронов. Однако современная системная нейробиология сталкивается с проблемой многообразия гипотетических объяснений функционирования ансамблей нейронов, исследования которых развиваются параллельно с исследованиями в области психологии и разработками ИИ. Затруднения экспериментальной нейробиологии состоят в том, что эксперименты, разрабатываемые для уменьшения сложности той проблемы, которую пытаются решать авторы эксперимента, только увеличивают набор экспериментальных данных и не уменьшают сложность рассматриваемых проблем. Одним из решений данных затруднений является коннекционистская программа когнитивных исследований, которая опирается на некоторые предположения относительно нейронных сетевых взаимодействий. Коннекционизм кладёт в основу функционирования мозга набор правил взаимодействия внутри сети простых элементов, которые объясняют каким образом происходит кодирование информации, репрезентация среды внутри сети и психологическая интерпретация этой информации. Как метод, такая программа предлагает возможность разложения новых данных из области нейронауки согласно принципам организации таких сетей для удовлетворительного понимания работы мозга и организации познавательных процессов[1]. Одной из базовых вычислительных моделей таких исследований являются искусственные нейронные сети, в прессе упоминаемые в контексте развития искусственного интеллекта основанного на данных. Нейронные сети служат переходным мостом межу исследованиями в области физики, нейробиологии и

[1] F. Rosenblatt The perceptron: A probabilistic model for information storage and organization in the brain Cornell Aeronautical Laboratory Psychological Review 1958. Vol. 65. №. 6.

искусственного интеллекта. Исследования свойств нейронных сетей в первой половине прошлого века стали предтечей программы коннекционизма. Исследователи ИИ на основе нейронных сетей стремятся смоделировать и репродуцировать мыслительные операции с помощью современных компьютерных технологий, когнитивная психология на основе нейронных сетей стремится описать процессы решения задач и объяснить особенности ассоциативного мышления, теория познания исследует познавательные способности и познавательную деятельность человека, возникшие на основе коннекционистской программы когнитивных исследований, онтология изучает мировоззренческие следствия такой программы, сообщающие нечто о природе нашего мышления. В когнитивной науке особое место занимает исследование репрезентативной эквивалентности искусственных и биологических нейронных сетей. Из этих сравнений вытекают общая логика организации сетей, функционирование сетей и репрезентация данных в сетях, что является важной вехой в развитии понимания устройства мозга и мышления человека. В России активно разрабатываются биологически инспирированные когнитивные архитектуры[2], а также исследуются сетевые структуры психики, тогда как сама методологическая программа коннекционизма не проблематизируется и её ограничения детально не исследуются. В самом общем виде коннекционизм считает представимым на основе вычислительных устройств, подобных мозговым структурам, воспроизвести психические свойства в искусственных системах из связанных между собой простых элементов. Современные нейросетевые подходы показывают как непосредственно из сетевых нейронных взаимодействий возникают когнитивные функции. Сегодня коннекционистская метафора стала стандартным исходным предположением во многих областях когнитивных исследований. Ввиду быстрого развития современных

[2] Biologically Inspired Cognitive Architectures (BICA) for Young Scientists: Proceedings of the First International Early Research Career Enhancement School (FIERCES 2016) // Advances in Intelligent Systems and Computing, 2016 Vol. 449

нейросетевых подходов возникла необходимость философского теоретико-методологического анализа, систематизации и понятийного оформления теоретических и практических наработок этой исследовательской программы, методологическую роль которой предстоит прояснить в настоящей работе. Также необходимо критически проанализировать смысл и актуальность выводов коннекционистской программы относительно познавательных процессов человека. В настоящей работе осуществляется исследование методологической роли коннекционистской парадигмы для современных когнитивных исследований.

Степень разработанности темы исследования

К философским исследованиям, повлиявшими на разработку коннекционистских моделей, можно отнести исследования аналитических философов в области логицизма начала XX-го века. Б. Рассел, а также А. Уайтхед разрабатывали логические выражения (expressions) для объяснения любого вида математических выражений. У. Питтс, математик и биолог, был лично знаком с Б. Расселом, он создал первую формальную нейронную сеть, положив в основу булевы операторы. Исследования в области психологии, предвосхитившие коннекционистские модели в когнитивной науке, — это разработки школы ассоцианизма. Дэвид Хартли, один из основоположников ассоцианизма, применил учение об ассоциации идей Дж. Локка к физическим процессам (вибрациям) мозга в своём труде «Размышления о человеке, его строении, его долге и упованиях». Понятия ума связаны между собой с помощью ассоциаций, говорил Д. Хартли. Его современник Д. Юм также предложил гипотезу о психических ассоциациях, выделив целый класс философских проблем, связанных с ассоциативным мышлением. И. Сеченов также может быть назван продолжателем идей ассоцианизма, уже в XIX веке он рассуждает о возможности тождественного рассмотрения психической ассоциации и физической связи элементов в мозге. И. Сеченов предлагает ассоциацию в качестве элемента, связующего психические явления и их

физиологическую основу в мозге. И. Павлов в начале XX-го века доказывает, как возможна ассоциация на физиологическом уровне. Исследования И. Павлова предложили механизм условного рефлекса, основанный на одновременном возбуждении коркового центра индифферентного раздражителя и коркового центра безусловного рефлекса, что приводит к установлению связи двух раздражителей. Динамический стереотип в рамках павловского учения возможно представить в виде сети ассоциаций, возникших внутри ассоциативных зон коры.

Ранние разработки по формализации активности нейронной сети были предложены У. Мак-Каллоком и У. Питтсом в 1943-м году и продолжены в работах Ф. Розенблатта в 1960-х. Но, несомненно, всеобщее признание коннекционистские модели получили после публикации в 1986-м году Д. Румельхарта и Дж. Макклелланда с соавторами, где обосновывалось предложение о параллельной распределённой обработке информации, как самостоятельной парадигме, объясняющей экспериментальные данные, известные в лингвистике и психологии. Смыслообразующей программой здесь стала биологически правдоподобная архитектура вычислительного устройства, а также алгоритмы обучения и целевые функции, как замена классическому программированию. К моменту написания работы по моделям PDP (Parallel Distributed Processing Model) в когнитивной психологии уже были известны случаи, для которых не могла применяться классическая вычислительная метафора. В качестве примера часто используют «эффект превосходства слов», который свидетельствует о том, что в знакомом контексте объекты опознаются быстрее и точнее, чем в незнакомом. В частности, люди могут обнаружить буквы в слове быстрее, чем отдельные буквы в бессмысленной последовательности букв. Для объяснения этого феномена была предложена нейросетевая модель, в её основе простые элементы, связанные между собой, одни из которых отвечают за опознавание букв, другие за части слов и третьи за целые слова. Так можно построить систему, которая может объяснить контекстное

распознавание. Эта система показывала, как можно догадаться, какое слово написано, даже если в нем есть ошибка (и часто эта ошибка не заметна). На основе нейронных сетей было также предложено правдоподобное объяснение классической ошибки сверхрегуляризации, которую осмысляли с помощью формальных правил, применяемых человеком к новым данным. «Модель интерактивной активации» продемонстрировала, как ошибка сверхрегуляризации для неправильных глаголов прошедшего времени в английском языке может возникать в нейронной сети без формулирования правил в явном виде. Эти убедительные свидетельства работоспособности коннекционистской парадигмы заинтересовали когнитивистов. Начиная с 1990-х годов коннекционистские разработки активно используются в когнитивной науке. Разработка моделей PDP позволила смоделировать появление структурированных и абстрактных представлений в многослойных иерархических сетях. Рекуррентные сети помогли понять различные аспекты изучения языка, такие как категоризация, контекстное различение слова и выделение морфем. Сверточные нейронные сети помогли в понимании обработки зрительной информации мозгом и объяснили, как обрабатывается перцептивная информация. Обучение с подкреплением применительно к нейронным сетям дало возможность объяснять действия, желания и мотивацию у когнитивных агентов. Однако выделение нейросетевых методов в отдельную отрасль машинного обучения привело к биологически неправдоподобным алгоритмам обучения и упрощённому пониманию нейронов и архитектур нейронных сетей. Ф. Крик в 1989 году осветил проблему понимания вычислительных свойств мозга посредством искусственных нейронных сетей. С его точки зрения, прямое сравнение обучения искусственных и естественных нейронных сетей неудачно и не отвечает реальным данным из области нейронауки. Критику коннекционистской метафоры продолжают исследователи распределенного хранения данных в сетях нейронов С. Малбург, М. Макклоски, Р. Роджер. Они отмечают, что новое обучение сетей приводит к катастрофе

суперпозиции (superposition catastrophe), в результате которой происходит забывание предыдущего обучения. В реальном мозге это приводило бы к потере долгосрочной памяти. Интересным аспектом коннекционистской проблематики является проблема временной и пространственной инвариантности, обусловленная особенностями механизмов естественного интеллекта, а именно врождённой способностью определения единиц объектов, которой нет у нейронных сетей. Также нейронные сети не формируют правила в явном виде и не могут прямым образом переносить знания на новый тип данных, эта проблема является основной линией критики коннекционизма сегодня. Обучение нейронной сети не позволяет объединить прямым образом элементы относительно их общего смысла, то есть сформулировать правило в явном виде. То же относится к обобщению смыслов и идей, эти вопросы изучены Г. Маркусом[3,4], Дж. Фодором,[5] М. Зайденбергом[6].

Полемика вокруг коннекционистской программы сегодня разворачивается во многом из-за новых способов решения этих старых проблем. Аргументы касательно обратного распространения ошибки обходятся коннекционистами за счёт биологически правдоподобных правил обучения сетей: выравнивание обратной связи (equilibrium propagation), равновесное распространение (Feedback Alignment) и прогностическое кодирование (например, работы Я. Ястрофа,[7] Л. Перье[8]), кроме того, исследователи, убеждённые в работоспособности алгоритма обратного распространения ошибки, ищут его признаки в мозге[9]. Решению проблемы

[3] Marcus G. (a). Deep learning: A critical aPraisal. 2018. ArXiv Preprint ArXiv:1801.00631
[4] Marcus G. (b). Innateness, AlphaZero, and artificial intelligence. 2018. ArXiv Preprint ArXiv:1801.05667
[5] Fodor J. The Mind Doesn't Work This Way; The Scope and Limits of Computational Psychology, MIT Press. 2000
[6] Seidenberg M. Sublexical structures in visual word recognition: Access units or orthographic redundancy? In M. Coltheart (Ed.), Attention and performance XII: The psychology of reading 1987. P. 245 – 263.
[7] Jastorff J., Kourtzi Z., Giese M.A. Learning to discriminate complex movements: biological versus artificial trajectories // J Vis. 2006. Vol. 6, № 8. P. 791–804.
[8] Perrinet L.U. An Adaptive Homeostatic Algorithm for the Unsupervised Learning of Visual Features // Vision (Basel). 2019. Vol. 3. № 3. P.47.
[9] Lillicrap, T. P., Santoro, A., Marris, L., Akerman, C. J., Hinton, G. Backpropagation and the brain. Nature reviews. Neuroscience. 2020. Vol. 216. № 6. P. 335–346.

забывания в настоящее время посвящено довольно много работ. К ним относится, например, работа М. Мермиллорд[10] о системах дополнительного обучения или работа Дж. Боверса[11] о локальном хранении данных в нейронных сетях. Проблема инвариантности также находит свои решения при условии использования архитектур особого вида[12]. Сегодня происходит бурное развитие нейросетевого подхода и машинного обучения формальных нейронных сетей. Модели предсказательного кодирования[13], вероятностные модели языковых способностей человека[14], модели глубокого обучения для нейронауки[15] и даже модели сознательных состояний[16] строятся на основе коннекционистского подхода. Сегодня уже существуют некоторые доказательства репрезентативной эквивалентности между нейронными сетями и мозгом[17,18], на самом деле в настоящее время публикуются сотни работ предлагающие общие свойства моделей «глубоких нейронных сетей» и когнитивных функций реализуемых в мозге. В то же самое время разработчики ИИ, вдохновлённые новыми открытиями в области нейробиологии, активно переосмысляют эмпирический материал, изобретая и тестируя новые искусственные сети[19]. Нейросетевые методы породили целую волну новых исследований в области ИИ и возобновили дискуссии о роли таких моделей в функционировании естественного интеллекта. Эта бурно развивающаяся область философских исследований в русскоязычной

[10] Mermillod M., Bugaiska A., BONIN P. The stability-plasticity dilemma: investigating the continuum from catastrophic forgetting to age-limited learning effects // Frontiers in Psychology. 2013. Vol. 4. P. 504.
[11] Bowers J. Parallel Distributed Processing Theory in the Age of Deep Networks // Trends in Cognitive Sciences. 2017. Vol. 21. №.12. P. 950–961.
[12] Zorzi M., Testolin A., Stoianov P. Modeling language and cognition with deep unsupervised learning: a tutorial overview. Front Psychol. 2013. №. 4. P. 515.
[13] Clark A. Whatever next? Predictive brains, situated agents, and the future of cognitive science // Behav Brain Science. 2013. Vol. 36, № 3. P. 181–204.
[14] Bengio Y., Ducharme R., Vincent P. A Neural Probabilistic Language Model // Advances in Neural Information Processing Systems 13 / ed. Leen T.K. MIT Press. 2001. P. 932–938.
[15] Richards B.A. et al. A deep learning framework for neuroscience: 11 // Nature Neuroscience. Nature Publishing Group. 2019. Vol. 22. № 11. P. 1761–1770.
[16] Bengio Y. The Consciousness Prior // arXiv:1709.08568 [cs, stat]. 2019.
[17] Gauthier I., Tarr M. J. Visual object recognition: Do we (finally) know more now than we did? Annual review of vision science. 2016. Vol. 2. P. 377-396.
[18] Yamins D. L., DiCarlo J. J. Using goal-driven deep learning models to understand sensory cortex. Nature neuroscience. 2016. Vol. 19. №3. P. 356.
[19] Hassabis, D., Kumaran, D., Summerfield, C., & Botvinick, M. Neuroscience-inspired artificial intelligence. Neuron 2017. № 95(2), 245-258.

философской литературе обсуждена недостаточно подробно.

Современными отечественными учёными и философами коннекционизм изучался в рамках разработок в области неклассической эпистемологии В. А. Лекторским[20], в рамках функциональных моделей в естественном языке Т. В. Черниговской[21], в рамках нейрокомпьютерных разработок А. В. Савельевым[22], в рамках междисциплинарных исследований нейрофилософии В. Г. Кузнецовым[23], для осмысления соотношения психических феноменов с их нейробиологическим субстратом Д. В. Ивановым[24], а также в отношении проблематики философии искусственного интеллекта А. Ю. Алексеевым, прояснившим вклад С.Н. Корсакова в развитие коннекционистской парадигмы и Петруниным Ю. Ю в рамках исследований истории искусственного интеллекта. Также стоит отметить работы Е.А. Янковской в гносеологии[25], И.Ф. Михайлова[26] в области социальной философии, В.Д. Арутюняна[27] и И.Г. Овчинниковой[28] в области лингвистики. Коннекционистская парадигма встречается в большом количестве диссертационных исследований (например, Н. Ю. Клюева[29] Е.М.

---

[20] Лекторский В. А. Реализм, анти-реализм, конструктивизм и конструктивный реализм в современной эпистемологии и науке» ИНТЕЛРОС [Электронный ресурс]. URL: http://www.intelros.ru/intelros/reiting/reyting_09/material_sofiy/6141-realizm-anti-realizm-konstruktivizm-i-konstruktivnyj-realizm-v-sovremennoj-yepistemologii-i-nauke.html (дата обращения: 10.06.2022).

[21] Черниговская, Т. В. «Language acquisition device» : Где оно? [Текст]/ Т. В. Черниговская // Детская речь как предмет лингвистического исследования. Матер. Междунар. науч. конф. (Санкт-Петербург, 31 мая-2 июня 2004 г.). - СПб.: Наука, 2004. - С. 280-281.

[22] Савельев А.В. Научная школа «Психофизиология и нейрокомпьютинг сенсорных систем» ИМПБ РАН / В.Я. Сергин, Е.В. Лосева, А.В. Савельев / Выпуск под ред. В.Я. Сергина, Е.В. Лосевой, А.В. Савельева // Нейрокомпьютеры: разработка, применение. 2015. №11. С. 5-7.

[23] Кузнецов В. «Аристотелевская теория категорий и прототипический подход» Вестник Московского университета. Серия 7. Философия, no. 1, 2018, pp. 32-44.

[24] Иванов Д.В. Радикальный энактивизм и проблема субъективности // Вопросы философии. 2016. № 11.

[25] Янковская Е. А. Гетерархический принцип устройства познавательного опыта. Иваново 2009.

[26] Михайлов И.Ф. Человек. Сознание. Сети. -М.: ИФ РАН, 2015. 196 с.

[27] Арутюнян В. Структура ментальных репрезентаций: извлечение текста из памяти, нейронная сеть и искусственный интеллект // Вестник Пермского университета. Российская и зарубежная филология. 2013. №4 240 с.

[28] Овчинникова И.Г. О коннекционистской интепретации речевой деятельности // Вопросы психолингвистики. 2006. №4. с. 37-47.

[29] Клюева Н. Ю. Отношение теоретических концепций и компьютерных моделей в исследованиях искусственного интеллекта. ФГБОУ ВО «Московский государственный университет имени М.В. Ломоносова», 2008.

Панина,[30] А.Г. Сонин,[31] М.А. Сущин,[32] В.А. Герович,[33] А.Ю. Алексеев[34], Барышников П. Н.[35], Савельев А. В.[36]), но как самостоятельный объект философских исследований не проблематизируется. В настоящем исследовании также нельзя охватить весь спектр существующей литературы, связанной с нейросетевыми методами, который очень велик из-за проникновения машинного обучения (своеобразного технологического аналога методологического понятия «коннекционизм») во многие сферы жизни, но попытаемся выделить наиболее демонстративные примеры.

Объектом настоящего исследования является коннекционистский подход к изучению проблем познания в когнитивной науке.

Предметом исследования является философско-методологическая роль вычислительных моделей на основе распределённой и параллельной обработки данных для современной когнитивной науки.

Цель исследования. На основе комплекса современных междисциплинарных исследований системной нейробиологии, психофизиологии, нейрофизиологии, философии и методологии искусственного интеллекта выявить и критически исследовать методологическую роль коннекционистского подхода в становлении и развитии проблематики современной когнитивной науки.

Для достижения поставленной цели решаются следующие **задачи**:

1. Раскрыть роль и место коннекционизма в философской методологии современной когнитивной науки.
2. Выявить роль коннекционизма в методологии нейробиологии и

[30] Панина Е.М. Когнитивная наука как комплекс междисциплинарных исследований. Москва, 2001.
[31] Сонин А.Г. Моделирование механизмов понимания поликодовых текстов. Москва, 2006.
[32] Сущин М.А. Концепция ситуативного познания в когнитивной науке: критический анализ. Москва, 2014.
[33] Герович В.А. Динамика исследовательских программ в области искусственного интеллекта. Москва, 1991.
[34] Алексеев А.Ю. Философия искусственного интеллекта: концептуальный статус комплексного теста Тьюринга. ФГБОУ ВО «Московский государственный университет имени М.В. Ломоносова», 2016.
[35] Барышников П. Н. Методологические возможности и границы вычислительных моделей сознания. Москва, 2018.
[36] Савельев А. В. Философско-методологические основания нейрокомпьютинга. Москва, 2016.

сформировать набор принципиальных характеристик нейросетевых моделей.

3. Выявить роль коннекционизма в методологии современной психологии и определить преимущества и недостатки моделей машинного обучения.

4. Изучить перспективы применения информационного подхода к сознанию для решения проблемы описания психических явлений на базе коннекционистской методологии.

Теоретико-методологическая база

Методы диссертационного исследования – это метод аргументации, включающий идентификацию и проверку аргументов, метод исторической реконструкции и концептуальной инженерии. При освещении темы истории возникновения коннекционистского подхода и самой когнитивной науки, а также при рассмотрении классических моделей PDP Д. Румельхарта, Макклелланда Дж., Элмана Дж., Маркуса Г., Я. Ликуна и др. опора происходит на метод исторической реконструкции научных и философских теорий. Для выявления оснований коннекционизма, а также гипотез «распределённого» и «параллельного» способа обработки информации главным выступал классический метод аргументации, позволивший выделить основные теоретические проблемы и ограничения этих подходов. Во второй главе был выдвинут ряд аргументов в поддержку информационного подхода для объяснения природы высокоуровневых когнитивных функций в рамках коннекционистского подхода в когнитивной науке, а также сформулирована гипотеза о возможности интеграции данных системной нейробиологии в рамках информационного подхода в когнитивной науке. Исторический подход, позволяющий проследить принципиальные и исходные положения коннекционизма в сравнении с символизмом, затрагивает статью А. Лавлейс (1843 г.) и библиографическую

находку последних лет – статью С.Н. Корсакова (1832 г.)[37].

Положения, выносимые на защиту

1. В современной когнитивной науке коннекционизм как методология параллельной распределённой вычислительной реализации когнитивных функций, двойственно дополняет символизм как методологию последовательной лингвистически заданной вычислительной реализации этих функций.

2. Коннекционизм является достаточным логико-эпистемологическим условием решения проблемы представления мозговой активности: он выражает принципы обработки информации в мозге посредством настройки весовых коэффициентов нейронных связей, не выражая онтологических притязаний на морфологическое подобие и физический изофункционализм искусственных и естественных нейронных систем.

3. Коннекционизм является достаточным логико-эпистемологическим условием решения проблемы представления познавательной деятельности человека: он объясняет механизм параллельной работы ассоциативного мышления и способность одномоментного восприятия различных чувственных модальностей сознания.

4. Коннекционизм является необходимым логико-эпистемологическим условием психофизиологического взаимодействия: психические феномены одномоментны и нелокальны с их нейросетевыми коррелятами.

Научная новизна исследования

1. Предлагаются новые критерии сходства с биообъектом мозгом

[37] Алексеев А.Ю. Протонейрокомпьютер Корсакова // Нейрокомпьютер: разработка, применение. №7. 2013. С. 6-17.

искусственных нейронных сетей с помощью уточнения и систематизации основных поверхностных и структурных сходств, а также найденных функциональных примитивов на уровне реализации алгоритмов естественных и искусственных нейронных сетей в виде канонических вычислительных операций «вычисления производной по времени» и «нормализации».

2. Впервые верифицирована гипотеза включённости операций «вычисления производной по времени» и «нормализации» в сигнатуру канонических операций описания динамики естественных нейронных сетей, так и в сигнатуру описания способа функционирования формальных нейронных сетей, которое имеет значение для прояснения подобия свойств между естественными и искусственными нейронными сетями.

3. Впервые аргументирована необходимость коннекционизма для методологической координации психических феноменов и нейробиологических явлений в рамках информационного подхода к сознанию и вычислительной теории разума.

4. Показана роль технической науки (информатики и теории связи) в становлении нейронной теории в прогнозирующем кодировании.

Теоретическая и практическая значимость исследования

Теоретическая значимость исследования формируется его вкладом в современное философское осмысление роли и значения программы моделирования психических феноменов на основе нейросетевых моделей в вычислительной теории разума. Выявлены тенденции в расширении области машинного обучения на основе революционных архитектур коннекцинистского типа, применительно к анализу данных современной нейробиологии. Обоснованы гносеологические предпосылки появления и развития коннекционистской парадигмы в рамках работы по созданию изофункциональных систем естественного интеллекта. Аргументы в пользу канонических вычислительных операций, производимых и мозгом и

нейрокомпьютером, могут быть использованы при разработке теоретических моделей в области системной нейробиологии.

Материалы диссертационного исследования могут использоваться в курсах лекций по теории познания, философии науки, а также в спецкурсах «Философия искусственного интеллекта», «Философия информатики», «Философские проблемы когнитивной науки».

Апробация работы

Основные результаты исследования нашли отражение в выступлениях на конференциях:

1. Хасанов Р.Ю. Коннекционистские интерпретации социального мышления // Международная междисциплинарная конференция «Искусственный интеллект в новой коммуникативной реальности», Москва, Россия, 18-19 июня 2020
2. Хасанов Р.Ю. Нейротехнологии в фармакологии // Международная молодёжная междисциплинарная конференция «Философия искусственного интеллекта», Россия, 30 мая - 20 июня 2019
3. Хасанов Р.Ю. Проблема сознания в нейробиологии // Международная молодёжная междисциплинарная конференция «Философия искусственного интеллекта», Россия, 30 мая - 20 июня 2019
4. Хасанов Р.Ю. Параметры качеств субъективной реальности и их естественнонаучные способы изучения // XXVI Международная научная конференция студентов, аспирантов и молодых учёных "Ломоносов-2019", МГУ имени М.В. Ломоносова, Россия, 8-12 апреля 2019
5. Хасанов Р.Ю. Словарь этических терминов в коннекционистских компьютерных системах // Международная научно-практическая конференция «Искусственный интеллект: этические проблемы цифрового общества», Белгородский государственный технологический университет имени В.Г. Шухова, г. Белгород, Россия., Белгород, Россия, 11-12 октября 2018

6. Хасанов Р.Ю. Дефиниции интеллекта в коннекционизме // Международная молодёжная конференция "Философия искусственного интеллекта - 2018", Российский государственный университет нефти и газа (НИУ) имени И.М. Губкина, Россия, 12-13 апреля 2018

7. Хасанов Р.Ю. Коннекционистский подход в общем искусственном интеллекте // Международная Конференция «Актуальные проблемы гуманитарных и социальных исследований» Институт философии и права СО РАН, Россия, 6 октября 2020

Публикации в изданиях, отвечающих требованиям п. 2.3 Положения о присуждении ученых степеней в Московском государственном университете имени М.В. Ломоносова:

А) Публикации в рецензируемых изданиях, индексируемых в международных базах Web of Science, Scopus, RSCI:

1. Brak Ivan V., Filimonova Elena, Zakhariya Oleg, Khasanov Rustam, Stepanyan Ivan. Transcranial current stimulation as a tool of neuromodulation of cognitive functions in parkinson's disease[38] // Frontiers in neuroscience. — 2022. — Vol. 16 (Web of Science, Scopus; JCR – отсутствует, JCI – 0,87, CiteScore – 6,6, пятилетний импакт-фактор РИНЦ – отсутствует).

2. Соколов И.С., Татаринцев М.К., Хасанов Р.Ю., Азиева А.М., Макаренко Е.Ю., Бурцев М.С. Устойчивость спонтанной электрической активности нейронных сетей in vitro[39] // Вестник Российского государственного

[38] Авторский вклад соискателя в данной работе определяется следующими задачами, выполненными им в рамках исследования, результаты которого были представлены в итоговом тексте, подготовленном к публикации: Хасанов Р.Ю. написал главу о гипотетических механизмах, объясняющих положительные эффекты от транскраниальной стимуляции мозга. Результаты отражены в седьмом параграфе первой главы диссертации, где показаны методологические различия современной вычислительной нейронауки и коннекционистского подхода при разработке моделей когнитивных функций. В статье рассматриваются модели, объясняющие болезнь Паркинсона, которые соединяют вычислительную и экспериментальную нейронауку. Показательно их отличие от коннекционистских моделей когнитивных функций, тем что они учитывают влияние физиологических эффектов на выполнение когнитивной функции.

[39] Авторский вклад соискателя в данной работе определяется следующими задачами, выполненными им в рамках исследования, результаты которого были представлены в итоговом тексте, подготовленном к публикации: Хасанов Р.Ю. участвовал в обеспечении экспериментальной части исследования по обучению нейрональной культуры; планировал постановку экспериментов; производил посадки культур нейронов, выделенных из гиппокампа новорожденных мышей линии C57BL/6, на 60-канальных мультиэлектродных матрицах 60StimMEA200/30-ITO (Multichannel Systems, Германия); записывал спонтанную пачечную

медицинского университета. — 2016. — № 2. — С. 45–49 (Web of Science, Scopus; JCR – 0,04, CiteScore – 0,5, пятилетний импакт-фактор РИНЦ – 0,424).

Б) Публикации в журналах, включенных в Список рецензируемых научных изданий по философским наукам, утвержденный решением Ученого совета МГУ имени М.В. Ломоносова:

3. Хасанов Р. Ю. Время как общий параметр качеств субъективной реальности // Искусственные общества. – 2019. – Т. 14. – Выпуск 2 (пятилетний импакт-фактор РИНЦ: 0,377). DOI – 10.18254/S207751800005656-0.

4. Хасанов Р. Ю. Аналитическое мышление как проблема коннекционистского подхода в когнитивной науке // Искусственные общества. – 2020. – Т. 15. – Выпуск 3 (пятилетний импакт-фактор РИНЦ: 0,377).
DOI – 10.18254/S207751800011011-1.

5. Хасанов Р. Ю. Как сегодня понимают интеллект // Искусственные общества. – 2021. – Т. 16. – Выпуск 3 (пятилетний импакт-фактор РИНЦ: 0,377). DOI – 10.18254/S207751800016769-4.

Структура исследования

Диссертация состоит из введения, двух глав, заключения и списка литературы.

активность культур; производил эксперимент по стимуляции культур; а также принимал участие в обсуждении анализа активности нейрональной культуры.

Глава 1. Теоретико-познавательные и методологические проблемы коннекционизма

1.1 Вычислительные модели познавательных процессов в когнитивной науке

Когнитивная наука, возникшая во второй половине XX-го века, стала новой вехой в истории исследования человеческих интеллектуальных способностей. Объектом исследования когнитивной науки становятся мозг и разум, которые формируют специфику когнитивных исследований. В процессе разработки альтернативных теорий, направленных на объяснение типичного человеческого поведения, например, бихевиоризма, возникли трудности в описании. Классическими примерами подобного рода трудностей становятся психологические эксперименты, для которых невозможно полное исключение влияния субъекта. Субъект действия, обладающий феноменальным опытом, также может быть эмпирически исследован. В частности, в работе Миллера «Волшебное число семь»[40], было показано, что объяснение феномена запоминания человеком набора объектов невозможно без привлечения самого способа запоминания, которым активно пользуется человек. Прямой критике бихевиоризма посвящена работа Н. Хомского «Три модели языка»[41], в которой тот показал, что способность человека к усвоению языка не может быть сведена к узким рамкам условно-рефлекторной схемы. В это же время в области исследования ИИ программа «Логик теоретик» А. Ньюэлла и Г. Саймона[42], программа, которая может рассуждать, привлекла внимание психологов, ведь стало ясно, что с

[40] George A. Miller. The Magical Number Seven, Plus or Minus Two. // The Psychological Review. 1956. Vol. 63. P. 81—97.

[41] N. Chomsky, "Three models for the description of language," in IRE Transactions on Information Theory. 1956. Vol. 2. №. 3. P. 113-124.

[42] Newell A., Simon H. Computer Science as Empirical Inquiry: Symbols and Search // Communications of the Associations
for Computing Machinery. 1975. Vol. 19. № 3. P. 113–126.

помощью процесса вычисления возможно моделирование рассудочной деятельности.

Мышление в связи с разработками в области исследований ИИ, стало возможно трактовать как аналог вычислительного процесса, а мозг стал описываться как аналог цифрового компьютера, что было предвосхищено такими исследователями как А. Тьюринг[43], Д. Хебб и Н. Винер[44]. Опора на интуитивно полу-очевидную метафору обработки сигналов из внешней среды мозгом, ассоциированная с развитием вычислительных моделей в компьютеростроении, стала концептуальным мостом, связывающим нейробиологические и психологические исследования. Н. Блок охарактеризовал вычислительную метафору как идеологию когнитивной науки[45]. В рамках такого подхода гипотеза о вычислительной природе нашего разума предложила понятие «когнитивная функция» для таких свойств человеческого мозга, которые рассматриваются как способность мыслить, что в новой перспективе объединяло явления, связанные с субъективными (приватными) знаниями от первого лица, и объективными физическими (публичными) процессами, наблюдаемыми в мозге.

Особое место компьютера, относительно других разработанных людьми механизмов, было отмечено А. Тьюрингом, Дж. фон Нейманом и Н. Винером. Полезно разобрать по частям работу вычислительного устройства. Логические элементы компьютера выполняют логическую функцию (операцию). Классическими алгебраическими операциями, реализуемыми в компьютере, будут сложение, вычитание, умножение и деление. Реализация этих операций происходит благодаря пропорциональному изменению двух измеряемых физических величин. Маркером осуществляемой операции

---

[43] Тьюринг А. Могут ли машины мыслить? // Информационное общество / Сост. А. Лактионова. М.: ООО «Издательство АСТ», 2004. С. 221–284.

[44] Винер Н. Кибернетика, или Управление и связь в животном и машине. М.: Советское радио, 1958

[45] «Я имею в виду идеологию… предполагаемую значительным направлением работы в области когнитивной психологии и искусственного интеллекта, а также в некоторой степени лингвистики и философии». The Logic Theory Machine -- A Complex Information Processing System. Allen Newell & Herbert A. Simon - 1956 - IRE Transactions on Information Theory 2 (3):61--79.

обычно служит электрической ток, или напряжение, или, зачастую, электрический импульс (для цифровой машины). Приведём пример: реализация умножения с помощью физических эффектов может осуществляться как «прямое умножение», для которого можно создать электромеханическую систему, в которой активная мощность равна произведению мгновенных значений напряжения и силы тока «P=UI». Суммирование и вычитание реализуют в аналоговой схеме с помощью, например, схемы симметричных трансформаторов. В общем случае для компьютера, производящего аналоговое вычисление[46], используются физические процессы, интерпретируя подходящее изменение физической величины как арифметическую операцию. На языке основных логических функций в аналоговой машине происходит выполнение сложной математической задачи, это позволяет нам говорить, что для машины любая задача должна быть представима в терминах элементарных логических операций этой машины. Представление задачи в таком виде – есть задача программирования, которое ориентировано относительно конкретной архитектуры вычислительной машины. Обычно вычислительное устройство воспринимает программу как последовательный набор команд, которые переводят машину из начального в конечное состояние, попутно преобразуя входные данные в выходные.

И машины коннекционистского типа могут преобразовывать входные данные в выходные, но делают это они неклассическим способом. Коннекционистские машины состоят из набора сообщающихся (связанных, как следует из названия подхода) элементов, способных передавать выходные значения друг другу. Правильным образом передавая сигнал между такими элементами, можно преобразовать любой набор входных данных в любой набор выходных данных. Говоря нестрого, такие модели

[46] Цифровая машина качественно не отличается от аналоговой на аппаратном уровне, в механической схеме цифровой машины, для разделения состояний используется пороговое переключение, например, поворот диска на 180 градусов, или появление или отключение тока в транзисторе, в то время как в аналоговой механической схеме используются промежуточные значения логического элемента.

аппроксимируют любую непрерывную функцию многих переменных с любой точностью. Такие машины могут быть Тьюринг-полными, то есть обладать вычислительной мощностью машины Тьюринга, и быть её полной заменой в ситуации составления моделей когнитивных функций, например, рекуррентные сети, использующие рациональные числа для весовых коэффициентов. Более того, существуют работы, показывающие как такие сети с иррациональным значением весов могут превосходить Тьюринг-полные машины, решая задачу остановки[47].

Развитие теории вычислений на абстрактном математическом уровне привело к появлению теоретической информатики, потенциал которой использовался для объяснения поведения объектов из области психологии, нейробиологии. Информатика исследует абстрактные свойства функций на предмет их вычислимости, для невычислимой функции невозможно найти способ последовательной реализации в терминах арифметических операций, то есть нельзя подобрать алгоритм, реализующий её пошаговое вычисление. Если имеется алгоритм для нахождения любого значения аргумента функции, то мы говорим, что функция вычислима. Алгоритм – это набор последовательных операций для нахождения определённого значения функции. Эффективный алгоритм — это такой алгоритм, который находит значение функции за минимальное количество операций. Большое количество алгоритмов может быть применено к одной и той же функции. Кроме того, большое количество устройств ввода-вывода данных с самой разнообразной физической реализацией могут выполнить один и тот же алгоритм. Мозг возможно представить в виде такого устройства ввода-вывода, можно сказать, что правдоподобной выглядит гипотеза о том, что мозг репрезентирует данные о мире в нейронных структурах. Поэтому возможно рассмотреть мозг как компьютер и выделить его физиологический уровень, алгоритмический уровень и вычислительный уровень обработки

[47] Balcázar J. Computational Power of Neural Networks: A Kolmogorov Complexity Characterization // IEEE Transactions on Information Theory. 1997. Vol. 43. №. 4. P. 1175–1183.

информации[48]. Согласно вычислительной теории разума, все когнитивные функции могут быть вычислены с помощью эффективных алгоритмов. В соответствии с тезисом Чёрча-Тьюринга[49], каждая такая функция может быть вычислена на машине Тьюринга. **Если гипотеза о вычислимости разума верна, мозг является аналогом машины Тьюринга.**

Достаточно много веских аргументов было приведено против этой гипотезы. Отметим, что возможность проверки этой гипотезы, сама по себе, предполагает позитивное движение науки о разуме и мозге в целом. Тезис о возможности построения алгоритма для любой интеллектуальной задачи приводит к идее искусственного интеллекта. Такая мыслящая машина предсказывалась Г. Лейбницем, Ч. Беббиджем, А. Лавлейс, самим А. Тьюрингом и Дж. фон Нейманом. Основной линией критики, относительно мыслящей машины, выступает мысленный эксперимент Дж. Сёрля – «Аргумент китайской комнаты», в котором проблема содержания сообщения представляется наглядным образом. Этот эксперимент демонстрирует возможность симуляции человеческого поведения компьютером без понимания выполняемых команд, только на основе правил перевода. Критика машинного интеллекта Дж. Сёрлем[50] также затрагивает проблему отсутствия у машин сознания и невозможность построения сильного ИИ. Также известны работы Х.Дрейфуса[51], Р. Пенроуза[52], Д. Лукаса[53] по проблематике сильного ИИ. Они утверждают, что классической вычислительной парадигмы недостаточно для возникновения сильного ИИ, то есть такого ИИ, который мог бы стать рациональным агентом. По утверждению Дж. Сёрля, аргумент против сильного ИИ также применим и к

[48] Marr D., Poggio T. From Understanding Computation to Understanding Neural Circuitry. Neurosciences Research Program Bulletin. 1979. Vol. 15. №.3. P. 470-488

[49] Church, Alonzo. An Unsolvable Problem of Elementary Number Theory // American Journal of Mathematics: journal. 1936. Vol. 58, №. 58. P. 345-363

[50] Searle J. Is the Brain's Mind a Computer Program? Scientific American T. Vol. 262. №. 1. P. 26–31.

[51] Дрейфус Х. Чего не могут вычислительные машины? М: Прогресс, 1978. 336 с.

[52] Пенроуз Р. Новый ум короля. М.: Едиториал УРСС. 2003. 339 с.

[53] Lucas J. Minds, Machines and Gödel, Philosophy. 1961. Vol. 36 (XXXVI). P. 112–127.

коннекционистской вычислительной модели, так как стандартная модель вычислений аналогична коннекционистской по сути (что демонстрируется на примере неклассического варианта эксперимента с китайской комнатой – китайского спортзала). Однако лучше других демонстрирует ограниченность функционализма следующий мысленный эксперимент – мельница Лейбница. Философ предлагает нам увеличить машину, производящую мысль, до таких размеров, что можно наблюдать за работой её частей изнутри так, как можно наблюдать работу мельничных жерновов, зайдя внутрь мельницы. При её осмотре не обнаружиться ничего в такой машине, кроме частей, толкающих одна другую, и не найдётся мысль (будь то воление или восприятие). Это замечание Лейбница показывает ограничения наших средств и методов познания чувственного опыта посредством моделирования когнитивных функций, построения интеллектуальных автоматов и, даже, так как эксперимент применим и к нервной ткани, визуальных исследований мозга. Таким образом, определение ментальных состояний через их каузальные отношения не раскрывает свойств психических явлений.

Стандартная модель вычислений используется как аналогия символьной обработки данных в когнитивной науке, она узнаваема в архитектуре Дж. фон Неймана. Эта математически не формализованная архитектура представлялась как абсолютная для любой машины до появления альтернативы в виде коннекционистской архитектуры. Эта модель произвела компьютерную революцию и, по заключению Патриции и Пола Чёрчленд, архитектура Фон Неймана имеет такую же важность, как и механика Ньютона, или электромагнетизм Дж. Максвелла.

Состоит она из следующих элементов:

1) «Памяти», в которой хранятся произвольные последовательности значений символов;

2) «Процессора», в котором находятся органы машины, выполняющие арифметические операции с элементами памяти;

3) «Управляющего устройства», которое указывает команды для

процессора и устройств ввода и вывода.

Классическая архитектура обладает ключевыми структурными элементами, которые связывают с символьной манипуляцией, то есть операции в виде правил выполняются над переменными. Эта метафора применялась для выражения мысленных операций как операций над символами. Например, высокоуровневые языки программирования Pascal, Fortran, C++ и др. имеют такие абстракции в виде смысловых конструкций для описания операций, или структуры данных, которые сформулированы в виде коротких команд вместо длинных последовательностей машинного кода. В лингвистике известна модель Д. Фодора[54] – теория языка и модель мышления в виде ментальных репрезентаций, воплощённых в особых нейрофизиологических структурах мозга. Также, классические вычислительные модели используются в психологии, например, разум как программное обеспечение Р. Келлога[55].

Классическая вычислительная парадигма была популярна до середины 80-х годов. Коннекционисты, в то же самое время, разрабатывали модели субсимвольной обработки информации. Они хотели обойти формальные ограничения, которые естественно вытекают из символьных представлений данных (подробнее рассмотрен у Т. В. Черниговской[56]), и разработать модели биологически правдоподобные и достаточно богатые для описания множества атрибутов перцептивного опыта. Важно, что терминологическая дихотомия между символьным и субсимвольным подходом нужна только для формального различения тех и других методов в то время, как в реальном компьютеростроении они с успехом применяются вместе. Коннекционизм может быть дополнением к классической вычислительной теории разума

[54] Fodor J. The Mind Doesn’t Work This Way; The Scope and Limits of Computational Psychology, MIT Press. 2000.
[55] Kellogg, R.T.: Fundamentals of cognitive psychology, 2nd edn. SAGE, Thousand Oaks. 2012
[56] Черниговская Т. В. Язык, мозг и компьютерная метафора //Человек. – 2007. – Т. 2. – С. 63-75.

если он рассматривается как микроструктура познавательной деятельности[57], или быть заменой символьного способа обработки информации если мы полагаем процессы обработки информации мозгом на аппаратном уровне как параллельные и распределённые[58]. **Применительно к когнитивной науке, коннекционизм может быть исследован как теория переработки информации мозгом на основе нейроподобных вычислительных систем, основное отличительное свойство которых – параллельная и распределённая обработка данных**.

В общем виде параллелизм — это воспроизведение аппаратной структуры в виде нескольких подобных структур. Каждая подобная структура решает часть задачи, что повышает производительность. Параллелизм подразумевает одновременность, а распределённость подразумевает разделение в пространстве. Для коннекционизма характерно и то, и другое в то время, как для последовательных вычислений возможно либо одно, либо другое, но не оба вместе. Коннекционистские модели являются нейроподобными в том смысле, что информация, которая хранится в вычислительном устройстве коннекционистского типа, **заключена в силе соединений между элементами этого устройства**. Эти модели экстраполируют результаты сложных вычислений неклассического вида на познавательные процессы, которые реализуются в мозге. Есть согласие в том, что познавательный процесс, рассматриваемый на временной шкале в границе нескольких секунд, имеет явно последовательный характер.

Для нас очевидно, что множество явлений происходит друг за другом, раскат грома следует за молнией, а зелёный сигнал светофора сменяет красный и жёлтый сигналы. Однако уже в 1890-м году Уильям Джеймс предложил специальный термин для психических феноменов, которые

[57]McClelland J. L., Rumelhart D. E., Hinton, G. E. The aPeal of parallel distributed processing. In A. M. Collins & E. E. Smith (Eds.), Readings in cognitive science: A perspective from psychology and artificial intelligence 1988. P. 10-11.

[58] McClelland, J. L., & Rogers, T. T. The Parallel Distributed Processing Approach to Semantic Cognition. Nature Reviews Neuroscience. 2003. Vol. 4. №. 4. P. 310–322.

представляются нами совместно и не разделяются во времени на отдельные звенья или цепи явлений, – «поток сознания». Такая «река» или «поток» существуют в более коротком временном промежутке, не более 300 мс. Этот поток сознания или «текущее настоящее» подробно изучен психологами и часто исследуется в работах по изучению рабочей памяти и внимания.

Итак, возможно ли будет выстроить в ряд или цепь микроявления нашего поведения и нашей психической жизни так, как это возможно сделать для представлений секундного размера? Для любой классической программы такое разложение действий на атомарные команды вполне возможно, но последователи коннекционизма утверждают, что в микроскопических масштабах огромное количество процессов в мозге, выражающих в совокупности единый поведенческий акт, происходят одновременно и распределено. Разные качества объектов, возникающие в сознании человека разделёнными, наблюдаются одновременно и параллельно друг другу, например, цвет, движение, форма и глубина визуального образа воспринимаются одновременно и отдельно друг от друга. Они присутствуют каждый на своём месте. Важно подчеркнуть, что вопрос о длительности самого времени «текущего настоящего» – это отдельный очень интересный вопрос, его ограничения – это лимит рабочей памяти. Он определяет количество доступных к одновременному обозрению параметров перцептивных стимулов и автоматических целенаправленных актов поведения,[59] но для микрособытий, происходящих близко ко времени «текущего настоящего», насколько бы короткое явление нам не удавалось рассмотреть, каждое из таких явлений будет содержать больше одного квале, то есть больше одного качества для наблюдения.

Следовательно, понятие когнитивной функции в рамках коннекционистского подхода не меняет своего первоначального значения, меняется лишь представление о том, как происходит вычисление

[59] Halford G., Wilson W., Phillips S. Processing capacity defined by relational complexity: implications for comparative, developmental, and cognitive psychology. Behavior Brain Science 1998. Vol. 2. P. 803-31.

когнитивной функции в архитектуре, представленной из связанных между собой функциональных элементов. Коннекционизм успешно применяется для объяснения микрокогнитивных актов, для которых предлагается параллельная обработка. Распараллеливание операций даёт возможность ускорить вычисление, но такого рода структура операций будет отражаться и на важных особенностях психологических интерпретаций поведения и перцепции.

1.1.1 Вычисление и шифрование

Важным преимуществом коннекционизма перед классической вычислительной метафорой следует признать его адаптацию к иным формам интерпретации преобразования данных. В исследовании Алексеева А. Ю.[60] обращается внимание на недостаточность традиционных принципов вычислимости, применяемых в компьютеростроении и слабость компьютерного инструментария для исследования социокультурных явлений. На помощь в решении проблем моделирования социума Алексеев А. Ю. предлагает машину Корсакова, работу которой автор предлагает интерпретировать с точки зрения коннекционистской парадигмы. Подробный разбор работы машины Корсакова приведён в параграфе 1.3. Для прояснения идеи автора предлагается различить два понятия, шифрование и кодирование. При кодировании предполагается существование таблицы преобразования данных источника в данные, понятные получателю. Универсальный код определяет точное однозначное преобразование данных любого вида с помощью этого кода, понятного каждому у кого есть таблица преобразований. Например, в мысленном эксперименте Дж. Сёрля «Китайская комната», таблицей для преобразования является китайско-английский словарь. При этом предполагается, что существует взаимно

[60] Алексеев А. Ю. Машина Корсакова (1832 г.) как прототип мультиагентного суперкомпьютерного автомата // Искусственные общества. 2019. Т. 14. Выпуск 1.

однозначное соответствие между символами текста китайского и английского языка, что и обеспечивает возможность перевода без понимания.

Но шифрование имеет цель преобразования данных с помощью алгоритма шифрования с целью сокрытия этих данных от лиц, не имеющих ключа шифра (примером алгоритма шифрования может стать популярный шифр замены букв алфавита – Шифр Цезаря) Шифр — это набор правил работы с данными и только, никакой таблицы соотнесения значения не требуется. Это важное преимущество шифрования перед кодированием, ведь в случае работы машины Корсакова можно сказать, что симптомы зашифрованы, а не закодированы в виде табличных значений. Как это работает? В случае с машиной Корсакова предполагается принцип подобия Юма, сходные естественные идеи отражают своё сходство при запечатлении их на перфокарте машины Корсакова. Дырокол образует на перфоркарте узор, который совпадает с другими узорами, проделанными на других перфокартах для других идей. Результаты изучения идей на машине Корсакова отражают сходство идей, но не их смежность во времени. Работа по изучению идей происходит за счёт распределённого и параллельного сравнения суммы признаков различных идей, что совпадает с принципами работы коннекционистской машины. Однако кодирования идей не происходит, алгоритмы прокалывания служат шифром для каждой идеи, отражённой на поверхности перфокарты. Шифрование без кодирования может происходить схожим образом и в мозге, где непосредственное прикосновение физических стимулов отпечатывается (запечатывается) в нейронной схеме, которая обладает алгоритмами шифрования этих стимулов. Поэтому коннекционизм может сохранять актуальность и в случае теорий обработки сигналов в мозге невычислительным образом.

1.2 История развития нейросетевых моделей обработки сигналов мозгом

В коннекционистском подходе используется метод обратного конструирования «по образу и подобию» нервной системы. Такие модели имитируют отдельные свойства живых нейронов, так, например, «связь» выполняет функцию аксона, а прочность «связи» выполняет функцию синапсов. Исследования нейронных сетей – достаточно старая область исследований, первая нобелевская премия, которую присудили К. Гольджи[61] и С. Рамон-и-Кахалю[62] в 1906-м году была вручена в знак признания трудов по изучению нервной системы. Описанные Сантьяго структуры напоминали сети из клеток с длинными ветвистыми отростками, которые были сложным образом переплетены между собой. Математики и учёные-физиологи долгое время разрабатывали теории для описания взаимодействий внутри сети нейронов, уже было известно об их электрической проводимости и сигнальных функциях. В 1944-м году У. Маккалок и У. Питтс предложили способ организации вычислений посредством нейронной сети, оснащённой пороговыми элементами и весами, такие сети уже могли производить вычисление. Эта модель послужила началом новых исследований в области моделирования работы нервной ткани. Несмотря на то, что она упрощала отдельные функциональные блоки, она показала способность сети простых элементов к решению задач, представленных для мозга.

В настоящее время основными упрощениями, которые обычно приводят к биологической неправдоподобности нейросети коннекционистского типа, критикуемой многими биологами, служит (1) отказ от импульсной активности в угоду детерминированным связям между нейронами и (2) использование алгоритма обратного распространения ошибки, нарушающего правило Д. Хебба – локального взаимодействия между нейронами. Кроме

[61] The neuron doctrine — theory and facts. Нобелевская лекция К. Гольджи (1906)
[62] Bentivoglio M. (1998). Life and discoveries of Santiago Ramón y Cajal. Сайт Нобелевского комитета

того, известно множество искусственных нейронных сетей, архитектура которых неправдоподобна с точки зрения биологии, что также привлекает внимание. В 1986 году была опубликована важная работа в области исследований коннекционизма: «Параллельная и распределенная обработка: исследования когнитивной микроструктуры»[63]. Модели PDP (The Parallel and Distributed Processing), которые были освещены в этой работе, являются предшественниками сетей глубокого обучения распространённых в современном направлении исследований ИИ. Эти модели ориентировались на прояснение вопроса о хранении и переработке информации в мозге и высказывали некоторые новые предложения о познавательном процессе. А именно, что познание опосредованно несимволическими вычислениями и, что знания закодированы в распределённом виде. Модели на основе систем распределённого вычисления оказались более устойчивы к мелким ошибкам. При реализации сетевых принципов в архитектуре информация хранится не в конкретной ячейке памяти, а распределена в большом количестве нейронов, что приводит к устойчивости системы при нарушении отдельных путей передачи сигнала. Кроме того, распределённая система также имеет преимущество в скорости обработки сенсорных сигналов. Реальный мозг, относительно современных компьютеров, медленная система (частота импульсной активности нейрона до 200 Гц), для ответной реакции пальцем на визуальный стимул понадобится в среднем 400 мс, что соответствует примерно 4-40 импульсам, проведённым одним нейроном. Распределённая обработка сигнала может решить эту проблему за счёт увеличения количества функциональных элементов, которые одновременно перерабатывают сигнал. В настоящее время модели глубокого обучения с использованием нейронных сетей, отличаются именно высокой скоростью обработки квази-натуралистических сенсорных сигналов, таких как пиксели

[63] Rumelhart D. E. Parallel distributed processing: explorations in the microstructure of cognition / David E. Rumelhart, James L. McClelland, and the PDP Research Group. / D. E. Rumelhart, Cambridge, Mass: MIT Press. 1986.

изображений или слуховые спектрограммы, за счёт разложения сложной задачи на подгруппы параллельно решаемых простых задач. Модели PDP достигли выдающихся высот по сравнению с классическими психологическими теориями в области лингвистики, психологии и нейробиологии. Но главной неудачей этих моделей также стала проблема объяснения познавательных способностей последовательного и дедуктивного вывода, то, что принято называть аналитическая естественная дедукция. Кроме того, несмотря на работоспособность этой теории, остаются трудные проблемы — это проблема понимания и проблема сознания, классические проблемы когнитивной науки.

Полновластными преемниками PDP моделей считаются сети глубокого обучения, или многослойные персептроны. Они занимают центральное место в исследованиях машинного обучения. Сети глубокого обучения претендуют на объяснение функционирования человеческого языка, ассоциативного мышления и контроля над двигательной активностью. Главными задачами, которые ещё предстоит решить глубоким сетям, это функции, дающиеся животным с развитой нервной системой: восприятие и контроль поведения, долгосрочное прогнозирование, рассуждение, планирование, общение. Несмотря на множество поклонников и уверенность в том, что многослойные персептроны могут анализировать данные без вовлечения в работу абстрактных знаний, они не являются самостоятельными автоматическими системами и во многом зависят от человека. Кроме того, основные достижения глубокого обучения составляют гибридные системы (альфаГо, гугл-поиск), нейросетевые подходы расширили возможности интеллектуальных систем, но также зависимы от классического программирования и написания машинного кода. Основные наработки ранних коннекционистских моделей из 80-х – 90-х годов активно используются в сетях глубокого обучения – это оптимизация целевой функции и алгоритм обратного распространения. Когнитивные архитектуры на основе многослойных персептронов сегодня представлены в работах С.

Гроссберга[64], Р. Сана[65], Д. Ноэлля[66], С.А. Шумского и других авторов. Потенциальные ограничения сетей глубокого обучения сегодня до конца не исследованы, есть мнение о недостаточности сетей глубокого обучения для воспроизведения всех интеллектуальных навыков человека. И есть мнение, что мозг в своей основе — это нейронная сеть, поэтому нам нужно исследовать новые архитектуры, продолжая работу в русле коннекционизма для объяснения автоматической и сознательной обработки информации в мозге.

1.3 Проблема описания познавательных процессов с помощью моделей коннекционистского типа

Понимание особенностей работы мозга для современной проблематики когнитивной науки играет центральное место в объяснении реализации когнитивных функций, наряду с моделированием нейронных сетей и исследованиями поведения человека. Эти труды важны для разработки фундаментальной теории соотнесения ментальных феноменов с функциональными взаимоотношениями нейронов. Для достижения понимания основ взаимодействия реальных нейронных сетей, предлагается создание вычислительных систем с нейроподобной архитектурой для сравнения свойств информационного процессинга мозга и компьютера. В основе такого сравнения лежит свойство изофункционализма систем. Тезис об изофункционализме предполагает, что сознание тождественно своим функциям и не тождественно своему материальному носителю, такие феномены, как восприятие, субъективность, квалитативный характер опыта, функциональны и реализуются в соответствии со своими когнитивными функциями. Гипотетически представимо существование реалистичной

[64] Grossberg S. Competitive learning: From interactive activation to adaptive resonance. Cognitive Science. 1987. Vol. 11. P. 23-63.
[65] Sun, R. The CLARION cognitive architecture: Extending cognitive modeling to social simulation. In: Ron Sun (Ed.), Cognition and Multi-Agent Interaction. Cambridge University Press: New York. 2004.
[66] O'Reilly R.C. Biologically Plausible Error-Driven Learning Using Local Activation Differences: The Generalized Recirculation Algorithm // Neural Computation. 1996. Vol. 8. P. 895–938.

функциональной организации машины, выполняющей вычисление когнитивных функций неотличимым от человека образом. Для такой системы наличие подобия в обработке информации и внутренней репрезентации этой информации также будет означать подобие репрезентации информации о мире и форме, которую компьютерная система может использовать для решения сложных задач. Интуиция функционалистов предполагает, что поведение такой машины не будет отличаться от человеческого. В то же время критики функционализма утверждают, что такая машина в виду отличной материальной основы (физикалисты) или же в виду отсутствия у неё особой «субстанции души» (антифизикалисты) не сможет воспроизвести такие способности. Такая критика, например, со стороны биологического натурализма Дж. Сёрля, предполагает некоторые свойства мозга, которые не смогут быть воспроизведены в машине. Для начала стоит провести сравнительный анализ физических свойств таких вычислительных систем с физическими свойствами мозга.

Идея воплощения естественных интеллектуальных функций в искусственной машине в литературе появляется довольно рано, например, в раннееврейском фольклоре есть образ голема, в Аргонавтике Аполлония Родосского присутствует сюжет о Талосе, бронзовом страже, созданным Гефестом. В новой истории примерами оживших механизмов могут стать «чудовище Франкенштейна» Мери Шелли (1818) или «Россумские универсальные роботы» Карела Чапека (1920). Мифу об искусственном интеллекте в 1950-е – 1970-е в виду развития вычислительной техники отводилась большая роль. Разрушение табу на сравнение машины и человека, в виду особого места во вселенной последнего, позволил себе сделать Н. Винер. «Отец кибернетики» в своём популярном труде «Бог и голем»[67] размышляет о соответствии «творца» и его творения и очеловечивает

[67] Wiener N. God and Golem, Inc: A Comment on Certain Points where Cybernetics Impinges on Religion. The M.I.T. paperback series. M.I.T. Press, 1966.

машину, сравнивая синапсы и логические элементы компьютера. У. Р. Эшби, один из последователей идей кибернетики, разрабатывая гомеостат, проводит аналогию между обучением мозга и адаптацией машины к условиям среды[68]. Но на современном этапе до сих пор остаётся открытым вопрос, где в мозге происходит элементарного вида вычисление между нейронами как элементами сети, или между синапсами этих нейронов и происходит ли вычисление вообще.

Благодаря кибернетической идее обратной связи Винера и разработкам по общей теории машин Дж. фон Неймана, миф об искусственном разуме стал центральной идеей, вдохновлявшей умы писателей второй половины XX-го века: У. Ф. Гибсон[69] (Нейроман), В. Виндж[70] (Истинные имена), Т. Лири (Хаос и киберкультура), М. Малц[71] (Психокибернетика), Э. М. Хилтон[72] (The social implications of mechanization, automation, and cybernation in agriculture), Д. Барлоу[73] (Декларация независимости киберпространства), Т. Мей (Шифромикон) и др., философов: Д. Харроэй[74] (Манифест киборгов), К. М. Сейр[75] (Кибернетика и философия сознания), М. Клайнс (Киборги и космос), и др. В рамках этой парадигмы было принято наделять машины человеческими чертами, и наоборот, например, религиозный культ Р. Хаббарта «Дианетика», работа Т. Лири «Infro-психология», «Психокибернетика» М. Малца используют свойства компьютера как метафору для описания работы мозга и разума. И хотя для нашего исследования эти работы могут быть привлечены только в качестве примера

[68] Эшби У. Р. Конструкция мозга. Происхождение адаптивного поведения М.: ИЛ, 1962. 397 с
[69] Гибсон, Уильям. Нейромант: Фантаст.роман / Пер. с англ. Е. Летова, М. Пчелинцева. — М.: Аст; СПб.: Terra Fantastica, 2000. 317 с.
[70] Истинные имена. True Names. Повесть, 1981 год. Язык написания: английский. Перевод на русский: А. Новоселов (Истинные имена). 2015. 2 изд.
[71] Maxwell M. Psycho-Cybernetics. Simon and Schuster, 1960.
[72] Hilton. A. M. The Social Implications of Mechanization, Automation and Cybernation in Agriculture // Front Cover, 1967
[73]Шарков Ф. И. Общение в Сети и зарождение сетевой киберкультуры. 2013, с. 98-108.
[74] Haraway D. Manifesto for Cyborgs: Science, Technology, and Socialist Feminism in the 1980s // Socialist Review. 1985. Vol. 80. P. 65-108
[75] Consciousness: A Philosophic Study of Minds and Machines. Random House, 1969, 273 pages. Hard cover: Peter Smith, 1972

субкультуры «поклонников ЭВМ», такого своеобразного «киберкульта», последуем за этой метафорой и сравним физические свойства нейронов со свойствами элементов логических схем машины. С какими машинами в таком случае может сравниваться нервная система?

Первым «механическим мозгом» в новое время является разностная машина Ч. Беббиджа – автомат аналогового типа по нахождению функций, которые являются многочленами n-го ряда (собрана в 1854 году) и аналитическая машина Ч. Беббиджа – автомат аналогового типа для процедуры сложения, вычитания, деления, умножения (принципы работы были разработаны в 1834 году[76]). Первый в мире цифровой механический компьютер был построен в 1938 году немецким инженером К. Цузе. В устройстве К. Цузе десятичная система была заменена двоичной для возможности замены десятичных ступенчатых колёс на двузначные контактные пластины реле, они стали функциональными элементами такой машины.

Машины аналогового и цифрового типа качественно не различаются. В аналоговой машине числа соответствуют физической величине, пропорциональной логическому элементу. Чаще всего имеется в виду сила тока, напряжение, угол поворота шестерни и т. д. Для цифровой машины характерно наличие цифровых маркеров, которым соответствует значение числа, записанного последовательно в виде набора двоичных символов 0 и 1. Для каждой десятичной цифры используют маркеры набором из восьми элементов. Обычно физическая величина, выражающая наличие маркера — это напряжение или сила тока. Элементарными единицами для такого рода машин стали транзисторы, реле, диоды, ферромагнитные сердечники. Итак, в вычислительных машинах цифрового типа все числа состоят из строк маркеров, которые появляются либо одновременно, либо последовательно в вычислительном устройстве машины. Очевидно, что в настоящее время для

[76] Петренко А. К., Петренко О. Л. Машина Беббиджа и возникновение программирования // Историко-математические исследования. 1979. Т. 24. С. 340.

машины цифрового типа уместно говорить только о транзисторах. Эти элементы содержатся в вычислительных органах любого цифрового компьютера. Габариты транзисторов ограничены в пределах до 5 нм квантовыми скачками электронов (туннельный эффект). Современные транзисторы имеют величину 20 нм ($2*10^{-8}$ м), они плотно упакованы в процессорах и организованы в виде вычислительных органов для выполнения логических операций. Время, затрачиваемое на одно вычисление, приближается к скорости света, но для компьютера удобно говорить о тактовой частоте процессора, она отражает количество тактов вычисления в секунду. Тактовый сигнал — это элементарное переключение состояния в транзисторе, для процессора **Intel Core i5** это $2,8*10^9$ - $3,3*10^9$ Гц. Цифровые компьютеры состоят из органов памяти и активных вычислительных органов. Активными вычислительными органами являются электрические схемы, исполняющие логические операции, такие как: поиск совпадений, комбинирование сигналов и др., то есть логические схемы арифметических действий, и органы, которые производят регенерацию цифрового сигнала - регениторы. **Такие системы, очевидно, не воспроизводят структуру нашего мозга.**

Нам достаточно хорошо известно, что основные функции нервной ткани реализуются в сетях нейронов. Мозг представляет такую систему, его можно назвать огромной сложнейшей нервной сетью, описываемой в виде коннектома. Нам также известно, что некоторые функции сильно зависят от времени предъявления сигнала, например, реакция на двигающейся стимул должна быть исполнена вовремя, иначе все движения не имеют смысла. В действительности, очень многие реакции и цепочки поведения имеют смысл только если они будут произведены с необходимой точностью как в пространстве, так и во времени. Эта особенность показывает принципиальную невозможность воспроизведения некоторых когнитивных функций в автоматическом режиме, но только в интерактивном. Приведённые выше устройства могут рассчитать для нас любую

последовательность действий, но им понадобится для этого время, которого зачастую не хватает. Распараллеливание производимых операций может решить эту проблему, мгновенный расчёт может произвести операцию по нескольким признакам сразу, сегодня этот факт очевиден как когнитивным учёным, так и разработчикам ИИ. Именно высокий параллелизм и распределённость обработки информации отличают классические CPU[77] от GPU[78]. И мозг и коннекционистские вычислительные устройства обрабатывают данные распределено и параллельно. Рассмотрим устройство простейшего коннекционистского устройства и покажем, насколько важны алгоритмы обучения для сравнения работы мозга и компьютера.

Первой машиной «для улучшения вычислительных способностей человека», машиной коннекционистского типа в аппаратной части, является прямолинейный гомеоскоп Семена Корсакова, разработанный им в 1832-м году[79]. Машина разработана С. Корсаковым для ускорения и упрощения диагностики симптомов заболеваний, а также для составления рецепта лекарства на основе суммы нескольких симптомов заболевания. Принцип работы устройства такого типа реализует параллельную и распределённую операцию поиска совпадений нескольких признаков по заданной таблице. Корсаков использовал перфокарты, для определения заболевания по конечному набору симптомов. Каждому заболеванию соответствовал уникальный набор симптомов, эти симптомы можно было зашифровать на перфокарте в виде уникального узора проколов (отверстий в картоне). Предположим, что наличие симптома кодируется «1» (прокол), а отсутствие симптома «0» (отсутствие прокола), тогда из набора 0 и 1 будет возможно представить все типы заболеваний с нетождественными симптомами.

---

[77] Центральный процессор (CPU) электронная схема, которая хорошо подходит для математических вычислений основных арифметических и логических операций, всех операций вода-вывода, указанных в инструкциях программы

[78] Графический процессор (GPU) электронная схема, которая хорошо подходит для математических вычислений матриц / векторов, используемых при обучении глубоких нейронных сетей, позволяет до 100 раз увеличить скорость работы нейросети

[79] Алексеев А. Ю. Машина Корсакова (1832 г.) как прототип мультиагентного суперкомпьютерного автомата // Искусственные общества. 2019. Т. 14. Выпуск 1.

Основная идея идеоскопа заключается в том, что врачу больше не нужно вспоминать какому заболеванию соответствует представленное множество симптомов. После того, как врач записывал шифр, он мог найти совпадающий набор симптомов с помощью машины Корсакова. Она состоит из таблицы набора болезней, представленных шифрами из симптомов.

Таблица заданных признаков	Перфокарта с шифром симптомов - 1110
0111	не подходит
1011	не подходит
1110	подходит

Табл. 1 Результат работы гомеоскопа Корсакова. Идея определена как формальное понятие, представленное битовой строкой.

Отличие работы гомеоскопа от классического вычислительного устройства в том, что сверка идёт сразу по всем признакам в таблице находится место, где совпадают все проколы и когда они совпадут заболевание распазнается. Близкие по набору признаков заболевания могут быть близкими по своей природе, например, вирусная и бактериальная инфекции могут различаться по симптомам значительно больше, чем две похожие бактериальные инфекции. Подобие признаков отражено в подобии узоров, что может отдалённо напоминать нам ассоциативное мышление (принцип подобия Д. Юма). Однако работа машины Корсакова опирается на подготовительный этап выделения признаков человеком, поэтому «некое подобие узоров» привносит человек. По идее С. Корсакова, его машина будет облегчать врачу поиск ответа, в некоторой степени, вынося процесс вычисления из его ума на деревянную поверхность гомеоскопа. Так процесс поиска совпадающего набора симптомов может происходить и в нашем мозге, но машина, как микроскоп или телескоп, усиливает нашу способность,

позволяя работать со сложенными из множества признаков идеями, что не могут удерживаться вниманием в сознании человека. Такой «коннекционистской метафоры» недостаточно для объяснения производства речи и творческого мышления, далеко отстоящих от простой автоматической системы поиска по набору признаков. Разговор с машиной Корсакова возможен, если машине будет предоставлена библиотека из всех возможных кодов ответов на вопросы. На эти закодированные вопросы машина Корсакова будет искать ответы в бесконечной таблице соответствий, сопоставляя код вопроса с кодом ответа. Интересно, существует ли идеальный ответ для любого вопроса, который можно найти в этой таблице? На самом деле число таких кодов ответов будет астрономически большим. Главная проблема всех методов, которые анализируют сходства — это возникновение комбинаторного взрыва при анализе данных.

Возможна и другая трактовка работы гомеоскопа как устройства шифрования[80]. В любом случае гомеоскоп не претендует на выполнение операции сходства или воспроизводство деятельности мышления, он лишь упрощает ментальную манипуляцию с помощью метода разложения сложных идей на конечное множество существенных признаков. Операция сходства становится доступна для анализа с помощью машины опосредованно при рассмотрении фрагментов пересечения признаков. Нужно отметить, что медицинские исследования сегодня тоже используют фрагментарные сходства для фармакологического поиска новых видов химических соединений, действия которых исследуются относительно известных данных о фрагментах таких веществ. Гомеоскоп Корсакова не может воспроизводить в автоматическом режиме поиск по сходствам признаков у различных идей, и, кроме того, сам гомеоскоп не настраивает таблицу признаков, поэтому всё ещё остаётся важным «экспертное мнение» – ручная настройка гомеоскопа человеком, составляющим саму таблицу из

[80] Алексеев А. Ю. Машина Корсакова (1832 г.) как прототип мультиагентного суперкомпьютерного автомата // Искусственные общества. 2019. Т. 14. Выпуск 1.

признаков. То есть, машина работает по сформулированным заранее правилам, это отличает коннекционистскую машину Корсакова от сетей нейронов, которые проходят предварительный этап обучения, настраивая веса в автоматическом режиме. Они формулируют правила в неявном виде внутри самой сети в весах соединений, что удобно, так как не требует от человека программировать правила действий сети. С другой стороны это является основной сложностью при работе с нейронными сетями, которые после настройки весов не выводят правила в явном виде и не обладают механизмом интроспекции для сообщения той закономерности, которая использовалась для объяснения результатов работы сети нейронов.

Более интересной работой Корсакова следует признать идеоскоп, основанный на результатах предыдущих размышлений автора. Это приспособление совершенствует гомеоскоп, предлагая кроме поиска совпадений, ещё и сравнивать между собой сложные идеи, то есть производит операцию сходства в явном виде. Действительно, некоторые зашифрованные на множестве симптомов болезни могут быть похожи или отличны, по их сходству можно проверять идеи самого разного толка. Идеоскоп мгновенно выдаёт, исходя из специальной таблицы и заранее определённого предмета, следующие результаты:

1) все соответствия, которые есть у сравниваемых идей при их соприкосновении;

2) всё то, что находится в заданной идее, но отсутствует в той идее, с которой её сравнивают, в сей момент;

3) всё то, что отсутствует в заданной идее, но есть в той идее, с которой её сравнивают;

4) всё то, чего нет ни у одной, ни у другой идеи, но есть у других идей из той же таблицы.

И так теперь выставленные на идеоскопе значения **1110** мгновенно дают результат

Значения таблицы	1)	2)	3)	4)
0111	0110	1000	0001	0000
1011	1010	0100	0001	0000
1110	1111	0000	0000	0001

Табл.2 Результат работы идеоскопа Корсакова. Каждая идея представлена битовой строкой фиксированной длины. В столбце 1) производится операция сходства в виде простой конъюнкции, как пересечение множества признаков A∩B. В столбце 2) производится операция логической разности A\B. В столбце 3) производится операция логической разности B\A. В столбце 4) производится операция отрицания нестрогой дизъюнкции $\neg(A \cup B)$.

Для идеоскопа недоступна операция симметричной разности двух формальных определений, или задача XOR. Таблица из совпадающих и не совпадающих значений получена автоматически в результате действия механической силы, двигающей машину. Сравнение сложных идей – задача трудоёмкая, поэтому использование идеоскопа может быть оправдано. **Идеоскоп не обучается, настройка признаков может происходить только в ручном режиме, нужно заранее расставить все возможные варианты признаков в правильном порядке.** Можно подумать, что идеоскоп работает как настроенная трехслойная нейронная сеть, которая производит одну операцию сравнения двух идей, но, к сожалению, такое сравнение слишком условно, идеоскоп может быть коннекционистской машиной, но не нейросетью. Сама операция вычисления в гомеоскопе и идеоскопе вызывает вопросы. Именно в отсутствии удобного алгоритма обучения гомеоскопа кроется его принципиальное несовершенство. Для мозга задача поиска удобных алгоритмов настройки синаптических связей также является актуальной проблемой. По-видимому, именно обучение с помощью такой сложной настройки связей приводит к разнообразным сложным эффектам на уровне поведения организма в целом.

Отметим также преимущество глубоких нейронных сетей перед

гомеоскопом, а именно их многослойность, которая также накладывает ограничения на работу гомеоскопа. На примере машины Корсакова покажем основное преимущество современных глубоких нейронных сетей – возможность декомпозиции задачи на подгруппу малых задач. Действительно, подобные идеи содержат одинаковые наборы признаков, которые несколько раз должны быть продублированы в таблице. Они кодируют каждую сложную идею в виде кодовой последовательности проколов. А в сети нейронов задача вычисления часто раскладывается на множество элементарных операций вычисления, которые перекрываются между собой во время решения сложной задачи. Пример: граф вычисления функции $f(x, y, z) = (x + y)*z^2$ можно построить с использованием простых арифметических операций «+» и можно вычислить квадрат z с переиспользованием одного элемента два раза. Например, на Рисунке 2, а можно возвести z в квадрат умножив его дважды. Для машины Корсакова процедура переиспользования невозможна. Граф вычислений, построенный на её основе, будет состоять из нейронов входного слоя – исследуемой перфокарты; скрытого слоя нейронов равного количеству анализируемых признаков; слоя выходных нейронов. Такая трёхслойная сеть дублирует вычисления множество раз. Она будет уступать современным нейронным сетям глубокого обучения в скорости. В них за счёт увеличения количества слоёв, возможно усложнение связей между элементами, что обеспечивает переиспользование одной операции в другой операции. В нейронных сетях, как графах вычислений, в качестве элементарных функций в узлах графа используются функции, которые строятся на нейронах. Разберём ниже каких из них достаточно чтобы отобразить любую функцию. Но можно уточнить заранее, что скалярного произведения векторов и функции активации в виде логического сигмоида достаточно для того, чтобы выразить любую функцию.

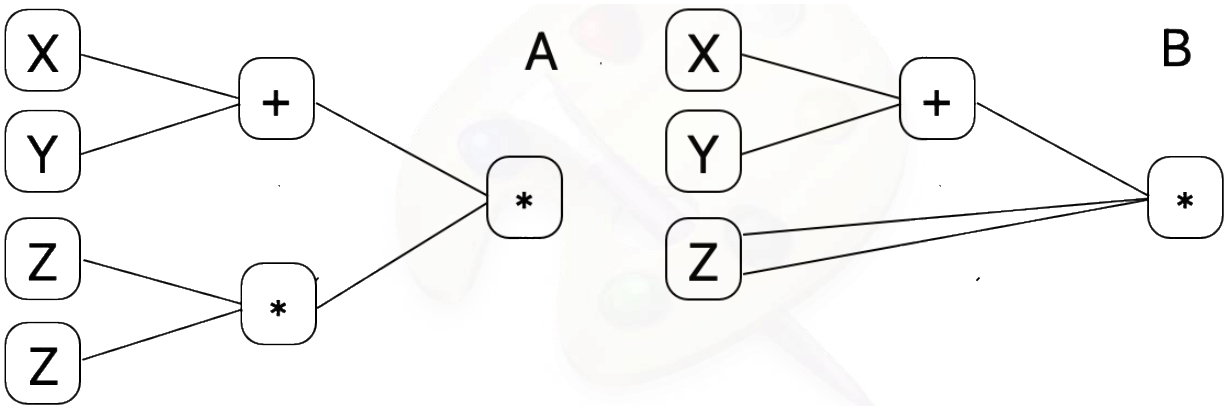

рис. 2 Графы вычислений для функции $f(x, y, z) = (x + y)*z^2$

a) с использованием операций «+» и «×»

b) в виде дерева с использованием операций «+» и «×»

И так, сети глубокого обучения переиспользуют некоторые элементы за счёт изменения графа связей внутри сети, что приводит к более быстрому поиску ответа. Примером такого устройства может быть **Cerebras Wafer Scale Engine** современный компьютер, воспроизводящий отдельные элементы коннекционистской машины. Этот экспериментальный многоядерный процессор гибридного типа $1,2 *10^{12}$ транзисторов, создан специально для задач машинного обучения и задач с применением искусственного интеллекта. Размеры его кристалла феноменальны 21,5*21,5*0,5 см, что примерно в 57 раз больше самого большого GPU (21,1 млрд транзисторов, 815 мм²). Процессор содержит $4*10^5$ ядер для параллельных вычислений и поддерживает стандартные фреймфорки, такие как PyTorch и TensorFlow. Система распределяет 400 000 ядер и оперативную память в соответствии со слоями программируемой нейронной сети. Задача оптимизирована таким образом, что все слои вычисляют одновременно. Любая нейросеть, предложенная этому процессору, преобразуется с помощью компилятора графов (Cerebras Graph Compiler, CGC). Компилятор выбирает уникальный маршрут вычислений, сопоставляя графы сети с вычислительным массивом. Огромные размеры этого процессора позволяют разместить все слои нейросети для параллельной работы, что экспоненциально увеличивает производительность. Таким образом,

революционный дизайн архитектуры вычислительных машин позволил оптимизировать сложного вида задачи ИИ, а также вдохновил системных нейробиологов на поиск структур мозга подобного вида.

Несмотря на то, что кажется вполне правдоподобным подобие обработки информации в сетях нейронов и в сетях глубокого обучения, не понятно на каком уровне мозг реализует подобного рода сети. Элементарным вопросом будет такой: служит ли биологический нейрон функциональным аналогом нейрона в сети глубокого обучения? Или принципы сетевой организации можно выделить уже на уровне генных сетей, или сетей регуляции синапса, или на уровне целого мозга? Многие нейробиологи, в том числе наши соотечественники О. Сварник[81] и Т. Черниговская, согласны, что нейрон способен на произведение регуляции, на много порядков опережающей представленную в искусственном нейроне, поэтому об изофункционализме говорить невозможно. Существует работа, предлагающая объяснение активности одного кортикального пирамидального нейрона на основе свёрточной нейронной сети из семи слоёв[82].

1.4 Принципы функционирования и общая организация нервных клеток в сравнении с искусственными вычислительными органами

В предыдущем параграфе рассмотрена простая машину коннекционистского типа, а также указали на возможность сравнения машины и мозга с точки зрения цели (или функции) выполняемых ими вычислений (вычислительный уровень по Д. Марру). Однако относительно изофункциональности этих двух систем остаётся много вопросов. В нейробиологии способ объяснения работы нейрона также связан с его функцией — порождением и передачей нервных импульсов. Ф. Крик

[81] Сварник О. Активность мозга: специализация нейрона и дифференциация опыта / Российская акад. наук, Ин-т психологии. - Москва: Ин-т психологии РАН, 2016. - 188 с.
[82] Beniaguev D., Segev I., London M. Single cortical neurons as deep artificial neural networks // Neuron. 2021. № 17. P. 2727-2739.

предлагает: *«Хотя нейрон и потребляет энергию, его главная функция — получать сигналы и посылать их вовне, то есть — обрабатывать информацию»*[83].

Упрощённое описание функции нейрона часто описывается как передача нервного импульса, однако порождение этого импульса представляет собой сложный многоступенчатый процесс, регулируемый с помощью физического и химического окружения. Нервный импульс условно постоянен, он представляет собой ответ на множество различного рода (внешних по отношению к нейрону) стимулов. Он также единообразен для всей нервной системы. Нейробиологи показывают как важна химическая среда вокруг нейрона и как условно значение одного потенциала действия, и говорят о частоте импульсной активности нейрона в ответ на предъявляемый стимул. Возможно предположить, что на первый взгляд наша нервная система выглядит как цифровой компьютер коннекционистского типа, так как состоит из нейронных сетей, которые общаются с помощью единообразных воспроизводимых в разных условиях нервных импульсов. Проверим эту гипотезу:

1) В мозге присутствует около 10^{11} глиальных клеток, некоторые из которых (астроциты) имеют отростки и участвуют в передаче нервного импульса[84]. Они обладают способностью обмениваться сигналами между собой, кальциевые волны хорошо различимы внутри тел астроцитов. Глиальные клетки взаимодействуют с нейронами, эти многочисленные вспомогательные элементы дополнительным образом регулируют передачу сигнала от клетки к клетке, например, нормальный химический синапс глутаматэргического нейрона функционирует совместно с глиальной клеткой. Существует явление миелиновой пластичности, особый вид взаимодействия оболочки миелина с аксоном нейрона, который приводит к

[83] Crick F. Astonishing Hypothesis: The Scientific Search for the Soul / F. Crick, Simon and Schuster, 1995. 340 с.
[84] Стасенко С. В., Гордлеева С. Ю., Семьянов А. В., Дитятев А. Э., Казанцев В. Б. Модель астроцитарной координации тормозного и возбуждающего входов интернейрона // Вестник ННГУ. 2014.

изменениям проводимости вдоль тела нейрона, что может также рассматриваться как вычислительная операция, так как на этом уровне возможна регуляция проведения импульса от тела к аксону[85].

2) Живые нейроны разнородны, не всегда возможно отделить аксон от дендритов. Нейрон состоит из тела и нескольких отростков, которые способствуют образованию контактов с телами и отростками клеток, далеко расположенных в пространственном отношении. Однако современные исследования показывают разнообразие соединений нейронов в ЦНС[86].

(3) Передача сигнала идёт не от дендритов к аксону, она распространяется по всей мембране нейрона, как круги на воде[87].

4) Также, многие нейроны демонстрируют спонтанную активность, они временами генерируют потенциал действия, а иногда и задают ритм для нейронов из ближайшего окружения, такая активность была описана для одиночных клеток in vitro, она показывает, что причиной возникновения потенциала может быть внутреннее состояние нейрона. In vivo для клеток гиппокампальной области CA1 человека.

5) Для проведения сигнала важна синхронизация импульса на пресинаптическом окончании с его постсинаптическим окончанием, синхронная активность резко меняет характер импульса, к тому же показывает, как нейрон взаимодействует со средой, предвосхищая будущие химические взаимодействия, циклически изменяя проводимость. Импульсные нейронные сети могут решить эту проблему.

6) Швырков В. Б. в 1983 году показывает, как нейрон привыкает к своей химической среде и реагирует импульсом только на изменение этой

---

[85] Bengtsson, S. et al. Extensive piano practicing has regionally specific effects on white matter development // Nature Neuroscience. 2005. Vol.8. P. 1148–1150.

[86] Существует особого рода электрические синапсы, которые выполнены в виде щелевых контактов между нейронами, такие контакты позволят передать электрический импульс от одного нейрона к другому, без синапсов.

[87] Casale A., McCormick D. Active action potential propagation but not initiation in thalamic interneuron dendrites // Journal of Neuroscience. 2011. Vol. 31. №.18. P. 289-302.

химической среды[88], эта способность называется адаптацией и вызывает много вопросов у биологов и физиков. Особое внимание к этой способности живых систем уделяет Шредингер Э. [89], рассуждая об адаптации, он рассматривает её с позиций Ламарка Ж. Б. и Дарвина Ч. и показывает, как важно индивидуальное развитие используемого качества (свойства, привычки, приёма, поведения) для реализации генетических потенций. Индивидуальное приспособление и адаптация играют громадное значение в жизни клетки. Нейрон способен адаптироваться, вспомним только привыкание к лекарству (drug sensitization) за счёт уменьшения экспрессии генов рецепторов, которые аффинны к этому веществу. Адаптация на уровне нейронной сети хорошо моделируется коннекционистским подходом, но адаптации на уровне нейрона может моделироваться ИНС фрагментарно на основе изменения силы связи.

7) Проведение электрического импульса по мембране — это не единственная сигнальная активность нейронов. Даже при отсутствии потенциала действия, они продолжают обмениваться химическими сигнальными молекулами. Факт того, что нейрон обладает всеми основными свойствами обычной клетки часто забывается, например, во время роста и установления соединений с соседями нейрон может обмениваться сигналами. Такие сигналы связаны с электрической проводимостью, но не обусловлены ей, они опосредуются тысячами различных химических веществ в межклеточной среде. Действие множества из них носит рецепторный характер, но существуют и те, что меняют активность нейронов, взаимодействуя с мембраной или влияя на внутриклеточный энергетический обмен. Химические взаимодействия нейронов множественны и расширяют возможности для сигнальной активности, например, дендритные

[88] Швырков В.Б. Системная детерминация активности нейронов в поведении // Успехи физиологических наук. 1983. Т.14. № 1.
[89] Шредингер Э. Анатомия разума: об интеллекте, религии, и будущем [перевод с немецкого] М: Родина, 2020. 208 с.

вычисления, как вид активности, влияющий на обучение в нейронах[90]. Исследователи показали возможность формирования декларативной памяти при условии ГАМК-эргического ингибирования проведения сигнала с помощью мусцимола. Суть обучения сводится к тому, что долгосрочное потенцирование возможно даже при условии отсутствия потенциала действия. И это явление с одной стороны подтверждает гипотезу о фундаментальном значении нейрональной пластичности при обучении (соответственно показывает ведущую роль соединений при образовании памятных следов), с другой стороны заключает в себе новый способ корректировки синапсов, который реализуется неклассически.

8) Нейрогенез, как комплекс процессов роста нейронов во время эмбриогенеза и во взрослом мозге, приводит к появлению новых связей и новых единиц обработки информации. Этот процесс динамический, наибольшая активность роста нейронов проявляется в пренатальный период, когда происходит процесс дифференциации нервных клеток на подтипы нейронов. Однако мозг продолжает своё развитие и в постнатальный период. Размер мозга новорождённого в первый год жизни увеличивается в 2 раза, мозжечок в первые годы жизни увеличивается в три раза (развитие моторики связанно с ростом клеток и синаптогенезом). Мозг младенца содержит в несколько раз больше синапсов, чем мозг взрослого человека, сокращение числа синапсов связано с развитием навыков и оптимизацией проведения сигнала[91]. Эти колоссальные изменения показывают действительное значение процессов роста в первые годы жизни. Во «взрослом мозге» нейроны продолжают образовываться, устанавливая соединения со своими соседями. Нейроны области гиппокампа активно делятся (около 1400 новых нейронов в сутки[92]), эти новые клетки встраиваются в виде новых элементов

---

[90] Rossato J. I. et al. Silent Learning // Current Biology. 2018. № 21. P. 3508-3515.

[91] The Postnatal Development of the Human Cerebral Cortex. Vol. 1. The Cortex of the Newborn // Journal of Anatomy. 1939. № Pt 4 (73). P. 674–674.

[92] Amrein I. Isler K. LiP H.P. Comparing adult hiPocampal neurogenesis in mammalian species and orders: influence of chronological age and life history stage. Eur. J. Neuroscience. 2011. Vol. 34. P. 978-987.

в нейронные сети мозга, изменяя граф связей. Активный нейрогенез в гиппокампе связан с формированием новых воспоминаний. Модели нейронных сетей не используют преимущества нейрогенеза, возможность смоделировать на современном этапе развития когнитивных архитектур такую активность представляется смутно[93].

Таким образом, машины, имитирующие работу нервной системы, зависят от изначальных гипотетических предпосылок того, как понимается функционирование нервной клетки. Способ идеализации структур уровня реализации (физиологического уровня нервной системы) сильно влияет на функциональные примитивы (элементарные процессы, не поддающиеся дальнейшей детализации) сложной системы. Математическое описание свойств биологических нейронов очень часто акцентирует внимание на проведении нервного импульса, отсюда и такая математическая абстракция. Такой нейрон имеет входные веса, передаточную функцию и результат вычислений на выходе представленную на Рисунке 2. Аналитически можно представить работу нейрона в виде выражения:

$u_k = \sum_{j=1}^{m} w_{kj} x_j$, где x_1, $x_2 \ldots x_i$ – входные сигналы; w_{k1}, $w_{k2} \ldots w_{km}$ – синаптические веса нейрона k; u_k – линейная комбинация входных воздействий;

$y_k = \varphi(u_k + b_k)$, где b_k – порог; φ – функция активации (передаточная функция), y_k –выходной сигнал нейрона.

[93] Kinouchi Y., Mackin K. J. A Basic Architecture of an Autonomous Adaptive System With Conscious-Like Function for a Humanoid Robot // Frontiers in Robotics and AI. 2018. (5). P. 30.

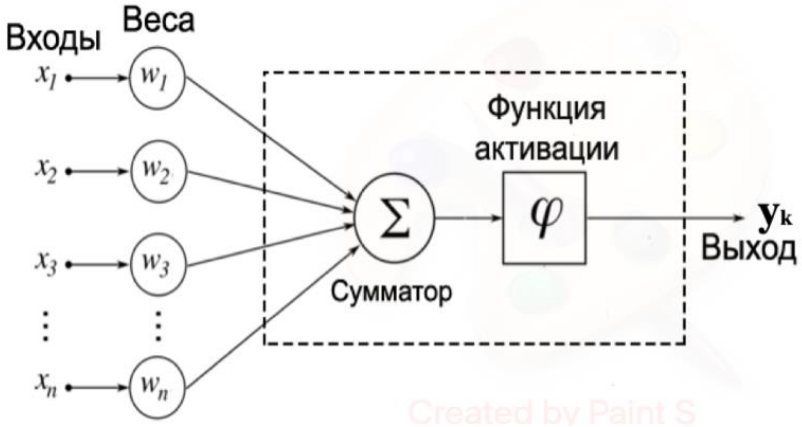

Рис. 1 Схема искусственного нейрона

Критика этой ограниченной модели была представлена выше, она сводится к наличию у нейрона дополнительных свойств, которые обусловлены его формой, поведенческой специализацией, связью с глиальными клетками, внутренней активностью (экспрессией генов), индивидуальной вариабельностью роста аксона, дендритов и дендритных шипиков. Однако модель, в которой нейрон получает множественные сигналы через синапсы, связывает их в своих дендритах и отправляет потенциал через свой аксон, доминирует в современной когнитивной науке, это – «модель связывающего нейрона».

Преобразование сложной системы входных данных в простое решение (потенциал действия) — основное теоретическое положение, которое воспроизводится в моделях деятельности нейрона. Представимо рассматривать работу нейрона в качестве элементарной когнитивной функции, смысл которой в снижении сложности, или категоризации входных данных. Разные нейроны выполняют категоризации разного вида и элементы в искусственном нейроне соответствуют основным элементам биологического нейрона. В этой ограниченной схеме отростки нейрона организованы в виде связей, аналогично синапсам биологического нейрона, каждая из таких связей обладает собственным весом w. Для каждого входного сигнала у нейрона есть собственный вес. Значения входа и веса перемножаются, а затем складываются. Сумматор выполняет арифметическую операцию сложения всех весов в виде их линейной

комбинации. Для выравнивания полученного числа применяют активационную функцию, которая укладывает полученные после сложения значения в область значений от [-1; +1] или [0;1]. Такое значение становится выходным сигналом нейрона. Веса во время одного вычисления не меняются. Вычисление таких математических функций на современном этапе происходит с помощью CPU или GPU. Для оценки подобия функций естественных нейронов и их искусственных аналогов разберём поверхностные сходства, такие как линейные размеры и скорость функционирования нейронов и вычислительных органов машины, предполагая, что форма естественного нейрона едина с его функцией.

1.5 Сравнение физических свойств элементов компьютера в сравнении с элементами мозга (изоморфизм на уровне реализации)

Количественный параметр

Астрономически большое количество нейронов в головном мозге человека часто упоминают на популярных лекциях. Процессор **Cerebras Wafer Scale Engine** состоит из $1{,}2 * 10^{12}$ транзисторов, в то время, как в мозге $0{,}86 * 10^{11}$ нейронов, значит ли это, что машины уже опережают нас в вычислительных возможностях? Это число нейронов не учитывает глиальные клетки. Как было рассмотрено выше, решение о проведении сигнала не может быть описано простой функцией. Решение этой проблемы видится в нейроморфных системах ИИ.

Нейроморфная инженерия разрабатывает симуляторы нейронных сетей из спайковых моделей кремниевых нейронов для реализации функции управления временем и частотой появления спайков[94]. В этих моделях кремниевый нейрон не будет соответствовать одному вычислительному органу процессора, он представляет интегральную схему из тысячи транзисторов и мемристоров (воспроизводит явление долговременной потенциации). Схемы такого нейрона довольно различны, поскольку в мозге

[94] Indiveri, G., Linares-Barranco, B., et al. Neuromorphic Silicon Neuron Circuits // Frontiers in Neuroscience. Vol 5.

существует много типов нейронов. Реализация таких нейроморфных систем, в первую очередь, нужна для асинхронных маломощных параллельных вычислений следующего поколения, которые, вероятно, смогут преодолеть разрыв в вычислительной мощности мозга и компьютера. Предположим, нам удалось собрать схему из 10^{12} транзисторов, тогда она реализует только 10^9 импульсных нейронов, что сравнимо с мозгом какаду или капуцина, но это число все ещё далеко от человеческого мозга.

Есть предположение, рассматривающее в качестве основного вычислительного элемента мозга синапс. В среднем на одном нейроне находится 10^4 – 10^5 синапсов, тогда в мозге из $86*10^9$ нейронов количество синапсов составит $0{,}86*10^{15}$ - $0{,}86*10^{16}$, по другим оценкам число синапсов[95] $0{,}15*10^{15}$, в данном случае разница в 60 раз не существенна. Мозг, по-прежнему, опережает все известные процессоры по количеству синапсов (**Cerebras Wafer Scale Engine** из 1,2 $*10^{12}$ транзисторов отстаёт на три порядка от наших приблизительных оценок).

Линейные характеристики нейронов и транзисторов

Рассмотрим пространственные характеристики нервных клеток. Нейроны в мозге имеют микроскопические размеры: тело (сома) нейрона от 1 до 20 микрометров. Оно имеет отростки – дендриты, длиной до 10 мм, и один самый длинный отросток – аксон, длиной до 10 м и диаметром 1-20 мкм. Объем, занимаемый отдельным нейроном, невелик, его величину можно оценить, разделив объем всего мозга на общее количество нейронов. При объёме мозга 1600 $см^3$ и $8*10^9$ клеток, объем нейрона составит $2*10^{-8}$ $см^3$, при 1000 $см^3$ и 10^{10} клеток объем нейрона составит 10^{-7} $см^3$. Диапазон нейрона от $2*10^{-8}$ $см^3$ до 10^{-7} $см^3$ равный от $2*10^{-14}$ $м^3$ до 10^{-13} $м^3$ показывает, что различие размеров с типовым транзистором объёмом 18000 $нм^3$ (30нм*30нм*20нм), или $1{,}8*10^{-23}м^3$ отличается на 10^8 – 10^9 порядков, для сравнения Солнце больше Земли в 10^9 раз.

[95] Pakkenberg B., Pelvig D., Marner L., Bundgaard M., Gundersen H., et al. Aging and the human neocortex. Expirience Gerontology. 2003 Vol. 38. №. 1-2. P. 95.

Если сравнивать искусственные элементы вычислительных машин с синапсами нейрона, получим следующие результаты: размеры синапса варьируют, в среднем ширина составляет 20-40 нм, а диаметр 1000-2000 нм, значит объем маленького синапса равен

$h*\pi*R^2=20*3.14159265358979323846264338327950288419 7*500^2=1,5*10^7$ $нм^3$. Эти результаты отличаются от транзистора ($1,8*10^4$ $нм^3$) на 3 порядка, что говорит о том, что **объёмные и линейные размеры функциональных элементов машины намного меньше, чем у мозговых структур**.

Энергопотребление

Рассмотрим энергетические характеристики нервных клеток. Известен популярный факт, что мозг потребляет столько энергии, сколько одна люминесцентная лампочка. Поскольку полагают, что энергия вызванного в нейроне импульса не производит никакой дополнительной механической работы, то вся энергия рассеивается в виде тепла. Рассеяние энергии в головном мозге составляет примерно 10 Вт, что действительно соответствует мощности лампочки[96]. Для каждого нейрона рассеяние энергии составит примерно $1*10^{-10}$ Вт. Для процессора GPU **Cerebras Wafer Scale Engine** рассеяние энергии составляет $1,5*10^4$ Вт, что соответствует $1,5*10^{-8}$ Вт на один транзистор и это **больше энергопотребления нейрона в 100 раз.** Ещё 60 лет назад разрыв в энергопотреблении между искусственными и естественными элементами составлял порядка 10^8 – 10^9 Вт, и теперь сокращён в 10^7 раз.

Скорость проведения сигнала

Стоит заметить, что временные реакции оценить легче всего, именно их первыми можно обнаружить если произвести эксперимент с нервной тканью. Рефрактерный период – это период длительностью $1,5*10^{-2}$ с, следующий сразу за потенциалом действия, во время которого проведение сигнала нейроном затруднено. В первые 2 мс после прохождения потенциала

[96] Отметим, мозг потребляет в 16 раз больше энергии чем мышечная ткань всего организма человека.

действия для нейрона невозможно произвести следующий потенциал действия, пока идёт восстановление заряда на мембране. Самая высокая частота активности нейрона составляет 200 Гц, например, гамма ритмы достигают 150 Гц. В то время, как тактовая частота процессора достигает 1-4 гГц, **что на 4 порядка больше скорости самых быстрых нейронов.**

Выводы, которые можно сделать из приведённых выше сравнений элементов машин и нейронов, такие:

1. Искусственные логические элементы сильно отличаются по своим физическим свойствам: линейным, энергетическим и временным. Активные элементы машин (транзисторы) меньше нейронов на 8 – 9 порядков и меньше синапсов на 3 порядка; превосходят нейроны в скорости на 4 порядка; энергопотребление одного транзистора в сто раз больше, чем у одного нейрона. Эти показатели указывают на отсутствие изоморфизма искусственных систем и мозга на физическом уровне.

2. Несмотря на то, что элементы компьютера мельче и быстрее, то есть количество операций в секунду, которые они выполняют, больше, но их производительность ниже. Для **Cerebras Wafer Scale Engine** этот показатель составляет $3{,}3*10^{15}$ флоп/с, против $2*10^{17}$ флоп/с для мозга[97] (10^{10} нейронов*10^4 синапсов*200 импульсов в секунду). Значит производительность мозга опережает современные искусственные системы в сто раз.

3. Очевидна разница тех функций, которые реализуются в мозге и в машине. Поэтому важной частью исследования становится сама *логика вычислительной системы и структура данных,* представленных в этой системе. Логика и алгоритмы обработки информации в современных транзисторных схемах только приближаются к тем, что реализованы в мозге. Это действительно и в обратном случае, поэтому становится важно, что вычисление какой-либо функции при изучении мозга, как вычислительной

[97] A Nick Bostrom and Anders Sandberg, “Whole Brain Emulation: A Roadmap,” 2008, P.130

системы, невозможно без изучения внутренней структуры данных в мозге и самих процедур вычислительных операций.

1.6 Поиск изофункциональных примитивов в искусственных нейронных сетях на алгоритмическом уровне обработки сигнала в мозге

В предыдущем параграфе показано отсутствие изоморфизма на физическом уровне биологических вычислительных элементов и искусственных органов машин. Показаны принципиальные ограничения современных вычислительных архитектур, реализующих команды последовательным образом на транзисторах, а также продемонстрированы отличия вычислительной мощности биологических и механических устройств. Показаны принципиальные свойства нейрона, которые воспроизводятся и не воспроизводятся в искусственных моделях.

Функции мозга можно моделировать на разных уровнях абстракции. Нейробиологи описывают одиночные нейроны и динамику активности этих нейронов с большой степенью детализации. Модели вычислительной нейронауки концентрируются на особенностях функционирования отдельных клеток и ищут математические функции со схожим распределением величин для построения модели наблюдаемых свойств. Машинное обучение на основе нейросети описывает обработку информации в мозге с помощью алгоритмов, которые ссылаются на биологические компоненты, такие как отдельные нейроны, только косвенно, воспроизводя самые простые свойства проведения сигнала между клетками. Если на уровне отдельных нейронов нет изофункциональных примитивов, возможно, коннекционистские модели воспроизводят изофункциональные примитивы на алгоритмическом уровне системы?

Полностью игнорировать уровень реализации нельзя, но для описания сложной системы обязательным условием является возможность абстракции от деталей локальной физической структуры к переходу на уровень работы системы в целом. В качестве примера можно привести самолёт, который хоть

и не хлопает крыльями, но также способен к полёту. Исследование принципов аэродинамики скорее приведёт нас к летающим машинам, чем бездумное копирование птичьего полёта[98].

Основные компоненты ИНС, которые позволяют им делать свою работу, это: (1) целевая функция, которая описывает ту цель, к которой стремятся при обучении нейросети; (2) алгоритм обучения, который сводится к набору правил корректировки весовых коэффициентов во время обучения; (3) архитектура сети, которая будет определять, как элементы связываются между собой и, следовательно, те операции, которые будут выполняться группами таких связанных элементов. Результаты обучения нейронных сетей — это тонко настроенные весовые коэффициенты.

Рассмотрим каким образом выполняются вычисления в нейронных сетях коннекционистского типа. Переход от уровня реализации к алгоритмическому уровню является вопросом функциональной декомпозиции, которая определяет внутреннюю (физическую) структуру примитивных операций и построения на их основе алгоритмов. Архитектура нейронных сетей также будет свойством для появления операций определённого вида. Для данного исследования интересны те алгоритмы и архитектуры, которые принципиально могут быть полезны для объяснения работы мозга и разума. Ниже будут рассмотрены три принципиально важных разработки в нейронных сетях, которые биологически правдоподобны и происходят из работ нейроучёных. Это обучение с подкреплением – объясняет действия агентов в среде, ценностные характеристики, желания и мотивацию. Это свёрточные нейронные сети – объясняют перцепцию. Это рекуррентные нейронные сети – объясняют память о предыдущем шаге, внимание и обработку последовательных данных.

[98] Михайлов И. Методологический выбор между субстанциализмом и функционализмом. Человек вчера и сегодня: междисциплинарные исследования. Вып. 6. М.: ИФ РАН, 2012.

1.6.1 Первые моделей искусственных нейронных сетей и их алгоритмы

Искусственные нейронные сети, по мнению Ф. Розенблатта, напрямую связаны с моделированием физических структур и нейродинамических свойств мозга. Именно как модель мозга задумывались первые успешно работающие нейронные сети – персептроны. Работа персептрона помогает понять основу устройства сети нейронов. Машина Марк-1, собранная Ф. Розенблаттом[99], в 1960-м году состояла из трёх слоёв одинаковых связанных элементов. Работа персептрона, согласно задумке, воспроизводит работу зрительного анализатора. Здесь приводится описание машины из оригинальной работы. S-слой, или слой входа, или слой чувствительных элементов (Марк-1 содержал квадрат 20х20 элементов, реагирующих на свет). Элементы реагируют согласно закону «всё-или-ничего», то есть на следующий слой подаётся значение 0 или 1. Импульсы из S-слоя переносятся в область проекции - слой А-1. Трансляция сигнала происходит от конкретного S элемента к конкретному А-1 элементу, однако количество соединений центральных S элементов с А-1 элементами больше, таким образом Ф. Розенблатт предлагает воспроизвести физиологическую особенность уменьшения плотности чувствительных клеток на периферии сетчатки глаза. Слой А-2 устанавливает случайные связи с предыдущим слоем А-1, а также, по задумке, может содержать прямые связи с S-слоем элементов. Слой А-2 передаёт импульсы на слой R, выходной слой, который состоит минимум из двух элементов, и, в таком простейшем случае, отвечает на вопрос о наличии в области чувствительных элементов образа, принадлежащего к искомой категории. Элементы R содержат рекуррентные связи возбуждения или торможения со слоем А-2.

[99] Rosenblatt F. The perceptron: A probabilistic model for information storage and organization in the brain Cornell Aeronautical Laboratory Psychological Review 1958. Vol. 65.№. 6.

Работающий персептрон, представленный выше — это бинарный классификатор, который может решить на основании входных данных, принадлежит представленный объект некоторому искомому классу или нет. Марк-1 будет формировать понятия сходства на основе совпадающих областей стимулов, а не на основе сходства контуров. Ранние работы по построению персептронов учитывали специфику конкретных задач распознавания образов и справлялись с задачей классификации, но обучение таких ассоциативных параллельных машин было трудоёмким. Персептроны вырабатывают решения (определяют, соответствует ли событие данному образу), суммируя опытные данные, полученные из множества экспериментов[100]. Автоматическая настройка персептрона (автоматическое программирование) сходно с обучением, поэтому этот тип машин иногда называют обучающимися машинами. Классическое обучение персептрона — это обучение с подкреплением. Алгоритм состоит из нескольких шагов. (1) На первом шаге синаптическим весам w_{1}, w_{2}, ..., w_{N} присваивают случайные начальные значения. (2) Затем, на втором шаге предъявляется входной образ X и получается некоторое значение на выходе. Если выходное значение правильное, то алгоритм завершён. Иначе к шагу (3). На третьем шаге применяется правило корректировки весов и получаются новые значения весов. (4) На четвёртом шаге новые значения весов подставляются в первый шаг. Правила корректировки весов могут быть различными. Ф. Розенблатт описывает несколько систем подкрепления для обучения своей машины: с учителем и в автоматическом режиме. В случае обучения с учителем, система управления подкрепления — это человек-оператор, он определяет как машина справилась с задачей. В случае автоматического обучения система подкрепления — это составная часть программы персептрона. Правила подкрепления, которые использует человек или машина, могут быть самые разные. Простейший случай такой: если синапс

[100] Minsky M., Papert S. Perseptrons M.I.T. Press. Cambridge. Mass. 1969.

участвовал в проведении сигнала, то увеличить его на значение сигнала подкрепления X (увеличить силу связи), если нет, то оставить таким же. Персептрон может обучиться в автоматическом режиме в случае системы подкрепления с коррекцией ошибки. Тогда машина в случае ошибки будет уменьшать силу синапсов, участвовавших в проведении сигнала, на значение сигнала подкрепления X, а в случае правильного ответа не будет менять силу синапсов, участвовавших в проведении сигнала. Для Ф. Розенблатта автоматическая настройка коэффициентов может происходить без мотивации и без активного действия персептрона. Он, аргументируя свою точку зрения, говорит, что павловское обучение не требует мотивации и не зависит от неё, делая поправку на то, что порой в эксперименте с условным подкреплением невозможно различить внимание и мотивацию.

В настоящее время продолжение исследований обучения с подкреплением приводят к парадигме, которая рассматривает обучение с подкреплением как достаточное условие для любой познавательной деятельности[101]. Но очень часто животные проявляют активные действия по изучению среды обитания. Их обучение и познание опосредуется действиями и движениями, например, оперантное обуславливание. Для такого рода поведения систем подкрепления Ф. Розенблатта недостаточно.

В случае собаки Павлова, обучение с подкреплением реализуется в ассоциативных зонах коры, которые обладают пластичностью для формирования памяти и закрепления условного стимула. Если рассматривать этот тип обучения как результат простого повторения условного стимула без мотивации и без активного действия со стороны собаки, то персептрон может быть моделью такого запоминания. Такой тип обучения хорошо воспроизводится алгоритмами подкрепления Ф. Розенблатта. Как предполагает Ф. Розенблатт, обучение без мотивации в автоматическом

[101] Silver D. et all. Reward is enough // Artificial Intelligence. 2021. (299). С. 103535.

режиме (повторение стимула) в ряде случаев для животных вполне возможно. В случае оперантного обуславливания необходимо активное действие животного и мотивация для того, чтобы возникло обучение. Рассмотрим биологически правдоподобное мотивированное обучение машин в следующем параграфе.

1.6.2 Современные модели искусственных нейронных сетей в когнитивной науке и их алгоритмы

В случае оперантного обуславливания требуется активное участие животного. Лабораторная мышь или крыса, случайно нажимая на педаль в клетке, получает еду, она воспроизводит это действие и снова получает еду, мышь понимает последствия своего поведения и устанавливает ассоциацию между педалью и едой – это оперантное обуславливание. Алгоритм обучения (и собаки, и мыши) включает вознаграждение в виде еды. Поведение животного модифицируется при получении вознаграждения – это обучение. Но оперантное обуславливание животного нацелено на получение награды в результате целой серии действий. Алгоритм состоит в (1) изучении ситуации, (2) построении предсказания действия. Если ошибки нет (3) то использовать предсказание в действии. (4) Иначе корректировка предсказания и возвращение к шагу (2). Ключевым элементом здесь является способность к активному перемещению и совершению действий, которые производит животное. Система вознаграждения в мозге находится в самом сердце процессов, связанных с центрами регуляции движений (это базальные ядра и кора головного мозга). Такого рода процессы, связанные с желанием и мотивацией, некоторые исследователи выносят на передний план, предлагается, например, термин «протосамость». Протосамость определяется как минимальный набор целенаправленных поведенческих действий, которые уже можно называть психическими процессами. Мозг — это изначально орган для регуляции моторных команд, поэтому очень важно

рассмотреть, как когнитивная деятельность обеспечивает целенаправленное движение.

Предположим, что нам удалость проникнуть в центр реакций, рассчитывающих направления и результаты наших действий, тогда внешнее воздействие на этот центр будет разрушать цепочку целенаправленного поведения животного или создавать другую цепочку действий. Такие эксперименты по электрической стимуляции области перегородки мозга были проведены Д. Олдсом и П. Милнером[102]. В случае такого вторжения крыса или человек попадает в цикл стереотипных поведенческих актов, которые производят работу, направленную на получение электрического стимула в область близкую к гиппокампу или перегородке мозга. И животное, и человек отдают предпочтение электростимуляции в сравнении с едой, водой или сексом. Вместо подкрепления, определённого биологическим перцептивным входом, обуславливающим разрешение биологической потребности в цепи системы вознаграждения, появляется более сильный и более быстрый способ удовлетворения потребности. Организм ориентирует своё поведение относительно этого нового импульса, что приводит к вырожденному случаю стереотипного поведения. Такие эффекты подкрепления электрическим током показывают роль мотивации в обучении и организации моторных действий.

В эксперименте с электростимуляцией подкорковых структур мозга человека можно узнать о переживании «от первого лица», то есть о психологической интерпретации формирующейся потребности. Человек сообщает, что испытывает приятные ощущения в случае нажатия кнопки, подающей ток, или раздражение в случае длительного отсутствия стимуляции. Естественные пути иннервации подкорковых центров, связанных с порождением у нас эмоций и мотиваций, также изучают с

[102] Olds J., Milner P. "Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain". Journal of Comparative and Physiological Psychology. 1954. Vol. 47 (6). P. 419–427.

помощью методов электростимуляции. Эти и другие открытия в организации обучения с помощью подкрепления, по мнению Ф. Розенблатта, являются важными открытиями для построения теории работы мозга.

Обратим внимание на важную роль дофамина в работе систем вознаграждения[103]. Его роль не сводится только к синхронизации корковых сетей во время получения награды. Дофаминовые нейроны активируются в ответ на предсказуемые и непредсказуемые стимулы вознаграждения, кроме того их активность зависит от понимания человеком структуры задачи. Теория прогнозирования ошибки (reward prediction error) предлагает способ объяснения работы дофамина в мозге для положительного и отрицательного подкрепления. Мозг строит предсказание относительно будущего и если оно ошибочно – происходит выброс дофамина, пропорциональный силе ошибки. Импульсная активность дофаминовых нейронов кодирует разницу между фактическим и ожидаемым результатом действия[104]. Эта теория подкрепляется открытием эффекта блокировки[105], который сводится к тому, что уже обученное животное не реагирует на второй (новый) условный стимул, который подаётся вместе с уже выученным (первым) условным стимулом. Первый стимул, которому обучили, блокирует обучение второму стимулу. Объясняется это тем, что второй стимул не обладает эффектом новизны. Он находится в поле зрения животного вместе с первым и это значит, что ошибки прогнозирования при его появлении не происходит. На психологическом уровне при достаточном внимании также может быть испытано чувство удивления или страха, которое оказывает влияние на пластические процессы в мозге во время обучения. Чем более ошибочно предсказание появления стимула, тем сильнее ответ дофаминовых нейронов. Эффект блокировки косвенно указывает на отсутствие в схеме условного рефлекса Павлова алгоритма обучения Ф. Розенблатта. Так как для условного

[103] Haber S.N., The place of dopamine in the cortico-basal ganglia circuit. Neuroscience. 2014. № 282. P. 248-257.
[104]Schultz W. Dopamine reward prediction error coding. Dialogues Clin Neuroscience. 2016. Vol. 18(1). P. 23-32.
[105]Kamin L. Selective association and conditioning. In Fundamental Issues in Associative Learning (Mackintosh, N.J. and Honig, F.W.K., eds).1969. P. 42–64

рефлекса также необходимо наличие новизны сигнала и внимания животного, то простой автоматической реакции для мозга недостаточно для модификации поведения в результате обучения.

Теория кодирования временной разницы поддерживает и развивает представления об активности дофаминовых нейронов и объясняет большее количество данных. Сигнал дофаминовых нейронов в такой модели также программируется алгоритмом обучения и управляется прогнозами о будущих возможных наградах, он называется алгоритм временной разницы. Алгоритм предсказывает значение ошибки прогнозирования, затем собирает фактические значения, суммирует их и сравнивает с начальной оценкой. Значение обновления (update value) соотносится со средним значением частоты импульсной активности всех дофаминовых нейронов. Такой алгоритм критикуется в виду того, что ответ для положительного и отрицательного подкрепления усредняется и функционально не различается[106].

В новой работе[107] команда DeepMind предлагает собственную модель алгоритма обучения с подкреплением на основе параллельного и распределённого кодирования временной разницы (distributed reinforcement learning). Вместо усреднённого значения возможных прогнозов предлагается параллельное вычисление прогноза для каждой задачи с фиксированной ценой вознаграждения. Авторы утверждают, что такая модель хорошо предсказывает асимметрию ответов дофаминовых нейронов на положительные и отрицательные ошибки прогнозирования, а также может объяснить несколько параллельно сосуществующих целенаправленных действий. Это и предполагает коннекционистская программа параллельной обработки на алгоритмическом уровне.

---

[106] Niv Y., Duff M.O., Dayan P. Dopamine, uncertainty and TD learning. Behavioral Brain Function. 2005. Vol. 1. P. 6.

[107] Wang, J. X., Kurth-Nelson, Z., Kumaran, D., Tirumala, D., Soyer, H., Leibo, J. Z., … Botvinick, M. Prefrontal cortex as a meta-reinforcement learning system. Nature Neuroscience. 2018. Vol. 21(6). P. 860–868.

Для объяснения любви к рискованному поведению также предлагаются алгоритмы с подкреплением[108]. Предположим, что нейросеть и крыса выбирают путь в лабиринте, где спрятано лакомство. Проведём серию экспериментов по обучению: после обучения нейросеть, основанная на обучении с подкреплением, будет воспроизводить одну и ту же последовательность действий, направленных на поиск лакомства, в то время как крыса будет в редких случаях ошибаться, проявляя «болезненное любопытство». Человек также показывает интерес к проверке плохих, проверенных ранее гипотез, а также осознанно нарушает ранее выученные правила. Эта особенность позволяет предположить, что обучение на организменном уровне обуславливается не одним алгоритмом, но связанно с высшими когнитивными функциями. Повышенная реакция дофаминовых нейронов на маловероятные стимулы, связанные с азартом, любопытством или прокрастинацией также предлагаются в моделях обучения с подкреплением[109][110]. Объясняются и некоторые познавательные способности, которые есть у человека и животных, но нет у искусственного интеллекта, например, метапознание. Метапознание – это знание о своём собственном знании, или способность эффективно использовать свой прошлый опыт применительно к новым задачам. Метапознавательные когнитивные способности включают в себя регуляцию собственной мотивации. Например, спросив себя зачем мне это нужно и поняв зачем это нужно, можно намного быстрее приобрести навык. Абстрактное знание о поставленной задаче влияет на ход выполнения этой задачи, ускоряя обучение. Например, обезьяны способны понять обратную задачу без предобучения. Если впервой задаче требуется смотреть направо для получения награды, а затем в следующей задаче требуется смотреть налево для получения награды, то для понимания

[108]Olds J., Milner P. "Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain". Journal of Comparative and Physiological Psychology. 1954. Vol. 47 (6). P. 419–427.
[109] Savinov N. et all. Episodic Curiosity through Reachability // arXiv:1810.02274 [cs, stat]. 2019.
[110] Deepak Path Pulkit Agrawal, Alexei A. Efros and Trevor Darrell. Curiosity-driven Exploration by Self-supervised Prediction. NTSL. 2017.

структуры второй задачи не требуется предобучение. Для объяснения таких способностей исследователи ИИ предлагают усложнить парадигму и выходят за рамки одного алгоритма. Например, в статье[111] разработчица компании DeepMind Джейн Ванг с коллегами предлагает рассматривать префронтальную кору в качестве центра реализации метапознанавательных способностей в совокупности с системой вознаграждения. В приведённой работе авторы рассуждают о способности машин обучаться обучению и предлагают собственную разработку когнитивного агента на основе LSTM сети (Длинная цепь элементов краткосрочной памяти (Long short-term memory; LSTM)), который симулирует метапознавательную деятельность человека. Модель реализует гипотезу, согласно которой дофаминовые нейроны способны кодировать значения стимулов посредством процесса абстрактного вывода без привязки к реальному вознаграждению. Сеть изображает работу префронтальной коры в виде рекуррентной сети, связанной с подкорковыми центрами (дорсальным стриатумом и медиодорсальным таламусом), образовывая кортико-стриальную петлю. Такие алгоритмы настройки сети нейронов в многослойных нейронных сетях могут объяснить появление сложного поведения в рамках парадигмы обучения с подкреплением.

В то же время обучение животных связано с хорошо понятным набором ситуаций, в которых они ориентируются. Для организмов с развитой нервной системой необходимо ориентироваться в совершенно особенном мире, где существует только одна попытка для осознания правила взаимодействия со средой, без возможности эволюционной адаптации. Эти условия хорошо отображает *Умвельт* - то есть набор особенностей перцепции и символического отражения физической среды обитания организма. Этот перцептивный мир параллельно, то есть одновременно, представляет набор возможных объектов и вещей для выбора и действия. Все они выбираются также исходя из соответствия возможности целевого взаимодействия с ними.

[111] Wang, J. X., Kurth-Nelson, Z., Kumaran, D., Tirumala, D., Soyer, H., Leibo, J. Z., … Botvinick, M. Prefrontal cortex as a meta-reinforcement learning system. Nature Neuroscience. 2018. Vol. 21(6). P. 860–868.

В виду того, что целевые функции у организмов разные, предметы и вещи этого перцептуального мира также имеют свои уникальные характеристики, необходимые для жизнедеятельности конкретного животного[112], что является трудным местом для когнитивных биологов при разработке когнитивных задач для оценки интеллектуальных способностей.

Обучение с подкреплением предложено в качестве универсального механизма управления процессами выработки новых реакций у животных. Алгоритмы обучения с подкреплением могут объяснить все формы обучения и поведения животного. Такое предложение в рамках коннекционизма предлагает фундамент для всякой интеллектуальной деятельности, как приобретённой в ходе эволюции способности организовывать действия в среде определённым образом на основе проб и ошибок. Живой или искусственный агент, согласно этой гипотезе, исследует среду обитания, реализует свои познавательные функции с помощью алгоритмов обучения с подкреплением и ему таких алгоритмов достаточно[113]. Все без исключения когнитивные акты, включая особенности перцепции физической среды обитания, зависящие от биологической определённости организма (формы тела, экологической ниши и способов реализации его биологических функций), могут быть описаны в рамках этой парадигмы. Они биологически правдоподобны и могут объяснять любого рода действия животных. Однако достаточно ли нам иметь в арсенале только алгоритмы обучения с подкреплением? В ряде случаев архитектура и связанность нейронов имеют решающее значение для той или иной функции мозга, кроме того, перцепция необходимо связана с организацией перцептивных органов, будь то ухо или глаз. Как уже было отмечено, скорость, с которой происходит обучение, – это лимитирующий фактор эволюции и если для обучения недостаточно времени или попыток, то оно не сможет быть отобрано естественным отбором. В

[112] Schaffner J. et al. Neural codes in early sensory areas maximize fitness //bioRxiv. – 2021.
[113] Silver D. et all. Reward is enough // Artificial Intelligence. 2021. (299). C. 103535.

настоящее время не предложено эффективных алгоритмов обучения с подкреплением для искусственных агентов, не уступающих человеку в обучении при малом числе попыток.

1.6.3 Современные модели искусственных нейронных сетей в когнитивной науке и их архитектура

Увеличение слоёв и развитие биологически правдоподобных алгоритмов обучения привело к появлению большого количества новых моделей ИНС. Если мозг человека обгоняет одну машину по количеству функциональных органов, то относительно нескольких, параллельно соединённых машин, это преимущество пропадает. Потенциально количество искусственных нейронов ничем не ограничено, так как можно программировать сколь угодно большое количество нейронов в сети, самые большие нейронные сети содержат порядка $1{,}6*10^{11}$ нейронов, что в 2 раза больше, чем количество нейронов в мозге человека и число таких нейронов можно сгенерировать ещё больше.

Увеличив число элементов, возникает вопрос об их архитектуре, поэтому также к основным отличиям современных нейронных сетей относят топологию их соединений. Рассматривают полносвязные и неполносвязные сети, сети прямого распространения и рекуррентные. Увеличение количества нейронов приводит к проблеме макромасштабного изоморфизма. Известно нейроны в мозге сгруппированы в функциональные структуры – ядра, а кора представляет многослойную (3 – 5 слоёв) структуру. Коннекционистским моделям редко удаётся правдоподобно воспроизводить нейроанатомические структуры, не говоря уже о репродукции всего мозга. Тем важнее будет разобрать особенную ветвь нейронных сетей – свёрточные нейронные сети (СНС), которые с успехом применяются для распознавания изображений и происходят из исследований органа зрения.

Первые модели СНС были основаны на работе К. Фукусимы, биологически-правдоподобной функциональной системы неокогнитрона,

которая объединяла в себе экспериментальные результаты Д. Хьюбела и Т. Визеля[114] по изучению зрения кошки. Эта модель организовывала несколько слоёв нейронов с небольшими рецептивными полями, которые самостоятельно обучались распознавать паттерны на предъявляемом изображении, соответствующие узнаваемой структуре. Согласно иерархической модели Д. Хьюбела и Т. Визеля обработка зрительной информации в визуальной коре происходит в организованной из нескольких слоёв сети прямого распространения, состоящей из простых нейронов латерального коленчатого тела, сложных нейронов V1 и сложных клеток высших отделов (V2, V4, нижневисочная извилина). Сложные нейроны имеют тенденцию выборочно реагировать на особенности воспринимаемого стимула, например, целый треугольник, квадрат, силуэт руки, рецептивными полями таких клеток становятся многочисленные входные отростки нижележащих нейронов. Сегодня иерархическая модель также актуальна несмотря на то, что она не воспроизводит все особенности функционирования клеток зрительной области: систему внимания, управление мышцами глаз и коллатерали с другими областями коры[115].

Особенностью СНС является наличие свёрточного слоя и слоя подвыборки внутри сети. Свёртка в сети необходима для извлечения признаков из изображения, подобных признакам, на которые реагируют сложные нейроны зрительной коры. На изображении 4x4 определим матрицу со значениями яркости пикселей, а также свёрточный фильтр (матрица весовых коэффициентов 2x2), который будет последовательно перемножаться с подматрицей матрицы 4x4. В первой матрице выбирается матрица 2x2 и перемножается со свёрточным фильтром, один элемент за другим, а после результат складывается. Представим свёрточный фильтр: он пробегает по всему изображению, на нем выставлены одни и те же весовые

[114] Hubel D., Wiese T. Brain mechanisms of vision // Scientific American. 1979. Vol. 241. P. 150—162.
[115] de Vries, S.E.J., Lecoq, J., Buice, M. et al. A large-scale standardized physiological survey reveals functional organization of the mouse visual cortex // Nature Neuroscience. 2020. Vol. 23. P. 138–151.

коэффициенты, таким образом, при нахождении похожего паттерна, фильтр будет выдавать примерно одно и то же численное значение умножения и, к тому же, после свёртки адресовать это значение в место на скрытом слое, соответствующее месту на изображении. В 2012-м году сеть глубокого обучения AlexNet победила в ежегодном конкурсе ImageNet с коэффициентом ошибки в 16%. Свёрточные нейронные сети не только хороший инструмент разработки машинного зрения, они воспроизводят внутренние свойства обработки зрительной информации. Свёртка легко находит границы и переходы между фигурами, различающимися по цвету и яркости. Блоки свёртки запоминают пространственное расположение, их легко определить. После операции свёртки можно составить двумерные карты объектов на изображении, что обеспечивает пространственное расположение внутри слоя. После операции свёртки выполняются дополнительные вычислительные операции — это нормализация для уточнения местного ответа, а также max-pooling (операция подвыборки) для уменьшения масштаба на поздних свёрточных слоях. Нормализация производится путём деления активности ответа нейрона в определенном пространственном местоположении на карте признаков на активность нейронов в том же месте на других картах признаков. Пример нормализации имеет особый интерес, так как эта операция применяется для объяснения ранних этапов обработки сигнала в зрительной системе. Здесь тоже делается упор на относительное, а не на абсолютное изменение интенсивности[116].

Предполагается, что нормализация играет ключевую роль в обработке сигнала в зрительной системе живого мозга. Она принимает участие в настройке динамического диапазона ответа нейронов, а также в нейронном кодировании. Функция нормализации состоит в вычислении отношения реакции отдельного нейрона и суммарной активностью пула смежных нейронов, её общее уравнение можно описать так:

[116] Carandini M., Heeger D. J. Normalization as a canonical neural computation. Nature Reviews Neuroscience. 2011. Vol. 13. P. 51–62.

$R = \frac{N}{D+\sigma^2}$, где N – входные стимулы для нейрона, D – нормализованный сигнал, концептуально представляющий нейронную активность «нормализованного пула» и σ^2 – константа.

Для зрительной системы эта формула должна быть модифицирована. Сетчатка работает в широком диапазоне интенсивностей, а адаптация глаза к изменениям входного сигнала достигается:

– изменением диаметра отверстия зрачка;

– перемещением темнового пигмента в слоях сетчатки;

– различной реакцией палочек и колбочек.

Нормализация в сетчатке объясняет изменение реакции палочек и колбочек, которые зависят от мощности светового излучения. Минимальным абсолютным порогом восприятия вспышки света будет поглощение 10 квантов света площадью сетчатки 10 угловых минут в течение 0,1 секунды. Для реакции на минимальный стимул зрительной системе нужны сигналы одновременно от 10 палочек, для суммации сигнала и восприятия вспышки света – это «минимальный» абсолютный порог зрительного восприятия. Кроме того, существует дифференциальный зрительный порог, определяющий относительную разницу в яркостях и адаптацию глаза к уровню освещённости. Дифференциальный порог известен, как эмпирическая психофизиологическая закономерность, которая устанавливает рост значения ощущения зрительного, слухового, осязательного, обонятельного раздражителя в логарифмическом масштабе. Логарифмические шкалы распространены для установления пропорционального увеличения ощущения, например, шкала яркости звёзд, шкала pH (ощущение кислого вызывают атомы водорода), шкала интенсивности звука, шкала частоты звука

$p = k * ln\frac{S}{S_0}$, где S — значение интенсивности раздражителя, S_0 — нижнее граничное значение интенсивности раздражителя.

Модель нормализации в сетчатке объясняет каким образом нейрон будет реагировать в зависимости от окружения, то есть как происходит вычисление относительного значения интенсивности света. На первом этапе вычисляется локальный контраст, на втором нормализуется значение этого локального контраста по отношению к общему контрасту. Локальный контраст связан с дифференциальным зрительным порогом, так как обусловлен различием яркости фона и яркости предмета на этом фоне[117]. Под яркостью фона обычно понимается среднее значение яркости, под яркостью предмета – локальное изменение яркости на общем фоне.

$$\text{Контрастность} = \frac{\text{Яркость предмета} - \text{Яркость фона}}{\text{Яркость фона}}$$

Увеличение интенсивности фона сдвигает кривую ответов вправо по логарифмической оси, этот эффект указывает на зависимость величины ответа отдельного нейрона от интенсивности фонового освещения. Эффект нормализации состоит в регуляции ответа нейрона в соответствии со средней фоновой интенсивностью света. То есть ключевым событием, определяющим ответ нейрона, становится не число квантов света, прилетевших на рецептор, а активность локуса, относительно которого нормируется активность этого нейрона. Сравнение с широкой полосой фона приводит к большим отклонениям от среднего значения активности пула нейронов. С другой стороны – сравнение с маленькой полосой фона приведёт к шуму, сигнал не будет специфичным. После адаптации глаза к яркости фона происходит нормализация локального контраста. Нормализация контраста — это второй этап, он происходит после адаптации глаза к интенсивности освещения (хотя механизм, вероятно, не разделяется на эти два этапа).

Эту эмпирическую закономерность возможно представить в аналитическом виде:

---

[117] Обычно контраст оценивается с помощью двигающихся прямоугольных решёток из черных и белых полос, так в эксперименте воспроизводят отклонение от недавнего значения локальной интенсивности света.

$Rj = \mathrm{Rmax} * \frac{\sum_i w_i C_i}{\sqrt{\sum_k a_k C_k^2 + \sigma}}$, где Rj – ответ нейрона, больше не пропорционален локальному контрасту C_i, но пропорционален общему контрасту – σ. Уравнение в числителе содержит весовые коэффициенты w_i – определяют пространственный профиль поля суммирования нейрона, они могут быть как отрицательными, так и положительными, входы a_k – определяют пространственный профиль пула нейронов, подавляющих активность.

Считается, что нормализация действует не только в сетчатке, но и на нескольких последующих этапах обработки зрительного сигнала. Модель нормализации была впервые разработана для учёта физиологических реакций нейронов в первичной зрительной коре (V1). Кроме того, предложенная модель используется для объяснения данных в слуховой коре, латеральной интратеменной коре и моторной коре.

Однако уравнение, приведённое выше, только описывает эмпирические закономерности, поэтому стоит объяснить механизмы лежащие в основе нормализации. Предполагается, что в основе операции нормализации, предложенной для этой модели, лежит латеральное торможение, в частности, для сетчатки это ингибирующее влияние ганглиозных клеток. Единообразие объяснительных моделей в вычислительной нейронауке и ИНС показывает глубокое родство принципов обработки зрительной информации. Современные СНС имеют несколько стандартных слоёв: операция свёртки, нелинейная нормализации и max-pooling. Архитектура в конкретных случаях может меняться, однако такие машины, которые могут распознавать образы на основе идей, описанных выше, могут решать сложные задачи категоризации объектов. Например, анализ репрезентативного сходства некоторых архитектур СНС показывает, что активность нейронов нижневисочной извилины обезьян репрезентирует информацию схожим с СНС способом. Кроме того, ответы нейронов из этой области возможно

предсказывать по анализу активности таких сетей. Лучше всего СНС воспроизводят активность высших отделов обработки зрительной информации в мозге, эти модели согласуются с разработками моделей Т. Поджио (модель стопок свёрток[118]), Т. Серре[119], Р. Ризенхубера[120].

Можно ли сказать, что СНС воспроизводят активность зрительной коры? Как отмечено выше, они воспроизводят некоторые принципы, которые оказываются достаточными для правильного функционирования таких сетей. Во-первых, это усложнение признаков, на которые реагируют нейроны высших отделов коры (иерархический принцип воспроизводится с помощью свёртки). Во-вторых, размеры рецептивного поля растут каждый раз после применения операции max-pooling, что также соотносится с реальными данными об обработке зрительного сигнала мозгом. В-третьих, модель нормализации, которая воспроизводится на всех основных блоках обработки информации в зрительной системе.

Работа с нейронными сетями всегда ведёт к абстрагированию от точных биологических деталей, результаты СНС нельзя интерпретировать только как модель, так как СНС могут использоваться в реальных задачах распознавания образов. Это не значит, что СНС на вычислительном уровне решают ту же задачу, что и наше зрение. Сознание в таких моделях не присутствует, экспликация и удержание во внимании феноменальной структуры и квалитативных качеств зрительного образа из набора обработанных данных не реализуется. СНС описывают ранние и поздние этапы обработки перцептивного сигнала в мозге, в то же время, несмотря на отсутствие физических сходств на уровне аппаратной реализации, СНС могут быть моделью репрезентации информации в мозге, а также могут быть подобны работе отделов обработки перцептивной информации в мозге. На это

[118] Serre T., Oliva A., Poggio T. A feedforward architecture accounts for rapid categorization // Proceedings of the National Academy of Sciences. 2007. Vol. 104. №. 15. P. 6424–6429.
[119] Serre T. Models of visual categorization. Wiley Interdisciplinary Reviews: Cognitive Science. 2016. Vol. 7. №. 3. P. 197–213.
[120] Cadieu C., Kouh M., Pasupathy A., Conner C., Riesenhuber M., Poggio T.A. A Model of V4 Shape Selectivity and Invariance // Jornal Neurophysiology. 2007. Vol. 98. P. 1733-1750.

указывают феномены иллюзий, на которые попадаются свёрточные нейронные сети, они аналогичны иллюзиям, которые видят животные[121].

Свёртка не единственный инструмент, показывающий, как обработка зрительной информации, которая происходит в первичной зрительной коре, может объясняться на основе работы нейронных сетей. Прогностическое кодирование — это хорошо зарекомендовавший себя инструмент обработки информации с широким спектром приложений, который полезен, например, при сжатии аудио- и видеоданных. Скажем, передаётся изображение пейзажа с голубым небом. Так как большинство пикселей в верхней половине изображения примерно одинаковы, очень неэффективно записывать значение цвета вновь и вновь для каждого пикселя, и, так как значение одного пикселя предсказывает значение его соседа, то эффективным способом является запись вместо атрибутов каждого пикселя разности между прогнозируемым и фактическим значениями для этого пикселя. Этот метод также напоминает свёртку, так как переходы между пикселями анализируются относительно их взаимного положения, а не абсолютных значений яркости. Так, для представления затенённого синего неба нужно только один раз записать значение синего цвета для целой группы. Поэтому основные ресурсы кодирования необходимы только для отслеживания точек «неожиданного» изменения, таких как ребра или другие линии формы объекта. Ранняя визуальная обработка в мозге предполагает использование различий между соседними значениями, например, для идентификации визуальных границ. Мозг может использовать предсказательное кодирование в восприятии, в выводе или даже в действии[122]. Модели предсказательного кодирования фиксируют существенные детали обработки данных зрительной системой в мозге млекопитающих[123]. Например, при обучении на прототипах

[121] Watanabe, E., Kitaoka, A., Sakamoto, K., Yasugi, M. Tanaka, K. Illusory motion reproduced by deep neural networks trained for prediction // Frontal Psychology. 2018. Vol. 9. P. 345.

[122] Huang Y., Rao R. Predictive coding // Wiley Interdiscipte Review Cognitive Science. 2011. Vol. 2. P. 580–593.

[123] Rao R. P. N., Ballard D. H. Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. Nature Neuroscience. 1999. Vol. 2. №1. P. 79–87.

визуального ввода, эти модели самопроизвольно развивают функциональные области для обнаружения границ, ориентации и движений, которые, как известно, существуют в зрительной коре. Эти работы также позволяют предположить, что визуальная архитектура коры головного мозга может развиваться в ответ на статистическое преобладание различных образов.

В этом параграфе указываются функциональные примитивы в виде групп нейронов, выполняющие элементарные вычислительные операции. Рассмотрена нормализация, как вычислительная операция, которая происходит на нескольких этапах обработки информации вентрального зрительного потока. Обсуждается, как операция нормализации и операция свёртки используются в СНС. Возможно, что существуют и другие функциональные примитивы архитектуры, которые роднят ИНС и ЕНС.

Для рассмотрения репрезентативной эквивалентности ИНС и ЕНС рассмотрим гипотезу о канонических вычислительных операциях. Большинство элементарных (канонических) вычислительных операций, предложенных сегодня биологами, реализуются не одним нейроном, а группой нейронов. К таким операциям относятся: сенсорная избирательность[124], декорреляция[125], линейный фильтр, нормализация[126]. Любая из этих функций может быть смоделирована на компьютере или может вычисляться с помощью ИНС. Нормализация, как каноническая вычислительная операция, была рассмотрена подробно выше.

Рассмотрим производную по времени, как элементарную каноническую вычислительную операцию. Червь вида C. elegans, как модельный объект часто используется в вычислительной нейробиологии, биологии развития и нейрофизиологии, так как имеет постоянное число нервных клеток (302

[124] Priebe N., Ferster D. Inhibition, spike threshold, and stimulus selectivity in primary visual cortex. Neuron. 2008. Vol. 57. №4. P. 482-497.
[125] Wiechert M., Judkewitz B., Riecke H., Friedrich R. Mechanisms of pattern decorrelation by recurrent neuronal circuits. Nature Neuroscience. 2010. Vol. 13. №8. P. 1003-1010.
[126] Carandini M., Heeger D. J. Normalization as a canonical neural computation. Nature Reviews Neuroscience. 2011. Vol. 13 №1. P. 51–62.

нейрона), связи между которыми полностью описаны[127]. C. elegans удобный объект для моделей, описывающих адаптивные возможности нервной системы и простейшие схемы контроля за движением организма. В исследовании Й. Ларша и его коллег 2015 года[128] описано поведение червя в градиенте диацетила и приведён механизм движения в направлении возрастания концентрации диацетила. Показана навигация червя в различной концентрации диацетила в диапазоне 10^5 нмоль (в исследовании говориться о 0,115 нмоль до 115 мкмоль). Авторы предлагают алгоритм вычисления *относительного* изменения концентрации запаха, состоящего из двух нейронов AWA – чувствительного нейрона и AIA – интернейрона, связанного с AWA щелевыми контактами. AIA получает информацию об изменении концентрации диацетила. В эксперименте исследуется биохимическое распознавание запаха и адаптация AWA при продолжительном воздействии химического вещества. Авторы с помощью изменения концентрации одоранта в среде наблюдают изменение направления движения животного в сторону увеличения концентрации запаха. Как показывают эксперименты, животное реагирует на обнаруженные изменения концентрации во времени (dC / dt)[129]. Поэтому становится интересным прояснить механизм такого поведения. Реакции AWA на увеличение концентрации оценивали с помощью внутриклеточного Ca^{2+}-имиджинга на микрофлюидной платформе, обеспечивая точную пространственную и временную волну распространения запаха диацетила в полностью жидкой среде. Динамика Ca^{2+} внутри AWA определяется концентрацией одоранта и историей взаимодействия с этим одорантом. Авторы изучили, как чувствительный нейрон адаптируется при увеличении концентрации, привыкает к воздействию диацетила и десенсибилизируется

[127] White J. G., Southgate E., Thomson J. N., Brenner S. The structure of the nervous system of the nematode Caenorhabditis elegans // Philosophical Transactions of the Royal Society A. 1986. Vol. 314. P. 1-340.
[128] Larsch J., Flavell S., Liu Q., Gordus A., Albrecht D., Bargmann C. A Circuit for Gradient Climbing in C. elegans Chemotaxis. Cell Reports. 2015. Vol. 12. №11. P. 1748-1760.
[129] Benhamou S., Bovet P. How animals use their environment: a new look at kinesis. Animal Behavior. 1989. Vol. 38. P. 375–383.

для средних и малых концентраций запаха посредством механизмов, связанных с трансдукцией сигнала внутри клетки. При высоком увеличении концентрации одоранта на несколько секунд (30 с, 115 мкмоль) десенсибилизации AWA почти не наблюдалось, при малых и средних величинах увеличения концентрации (30 с, 1,15 мкмоль) десенсибилизация наблюдалась в качестве адаптивного свойства. Чувствительный нейрон AWA запоминает историю взаимодействия с диацетилом в течении 30 секунд. Интернейроны AIA реагировали на кратковременное усиление запаха, однако полностью десенсибилизировались в течение 5 секунд. Изучение совместной реакции AWA и AIA показало, что реакция нейронов AIA зависит от величины изменения концентрации, при 14% повышении концентрации одоранта не происходит ответной реакции AIA в большинстве случаев, но при 58% и 115% повышении концентрации наблюдается однотипная реакция увеличения Ca^{2+} в течение нескольких секунд. В то время, как для AWA при увеличении концентрации диацетила на 14% AWA почти не реагировал, но при увеличении концентрации на 58% или 115% AWA реагировал на каждое увеличение диацетила, последовательно увеличивая внутриклеточную концентрацию кальция. Реакции AIA на увеличение кратного изменения запаха были стереотипизированы по величине и динамике независимо от абсолютной концентрации диацетила. Таким образом, была продемонстрирована схема AWA-AIA, которая обнаруживает увеличение концентрации диацетила на 58% или 115%. Такая простая система реализует вычисление с помощью двух нейронов, один из которых реагирует на абсолютное, а другой на относительное изменение концентрации. Интернейроны AIA дают нормализованный ответ, отбрасывая информацию об абсолютных уровнях запаха, представленную в AWA. Ответ AIA можно понять, как операцию dC/dt, то есть как вычисление скорости изменения концентрации вещества. В случае увеличения концентрации одоранта поведенческий ответ — это локомоция червя в направлении источника запаха. Червь движется хаотично в случае

постоянной концентрации и в случае увеличения концентрации начинает плыть в сторону пребывающей волны запаха. *В чистом виде описывается алгоритм поиска с памятью о предыдущем шаге.* Последовательное вычисление относительного изменения концентрации может быть использовано для поиска максимальной концентрации запаха. Такое поведение при хемотаксисе у C. elegans для других одорантов в градиентах аммония хлорида или биотина были описаны в работе 1999 года[130]. Анализ скорости поворота показал, что червь движется случайным образом то прямо, то начинает делать повороты. Повороты коррелируют со скоростью изменения концентрации одоранта (dC/dt), но не с абсолютной концентрацией одоранта. Поворот переориентирует животное по градиенту концентрации одоранта.

Эта модель работает, когда червь окружён только одним типом аттрактанта. Однако C. elegans распознает более сотни запахов и дикая среда — это всегда многофакторная среда. Помимо такого движения вдоль роста градиента, червю требуется производить выбор между несколькими аттрактантами, для которых существуют аналогичные системы чувствительных нейронов, с похожей логикой работы. Например, для C. elegans описана система, реагирующая на уменьшение запаха, как предполагают исследователи — это система AWC сенсорного нейрона. AWC связан с AIB и AIY интернейронами которые приводят к противоположной локомоции[131]. Также показаны параллельные механизмы для навигации в растворе NaCl[132]. Конвергенция сигналов происходит на уровне AIA интернейрона, интегрируя положительные и отрицательные химические сигналы AIA модифицирует поведение червя, например, обуславливая

[130] Pierce-Shimomura J.T., Morse T.M., Lockery S.R. The fundamental role of pirouettes in Caenorhabditis elegans chemotaxis. Neuroscience. 1999. Vol. 19. P. 955-997.

[131] Chalasani S.H. et al. Dissecting a circuit for olfactory behavior in Caenorhabditis elegans // Nature. 2007. Vol. 450. P. 63–70.

[132] Iino Y., Yoshida K. Parallel use of two behavioral mechanisms for chemotaxis in Caenorhabditis elegans // Neuroscience. 2009. Vol. 29. №17. P. 5370-5380.

аверсивное обучение[133]. Исследования параллельной интеграции показывают, что эффект интернейрона AIA определяется контекстом схемы, в которой он находится, и в результате AIA может подавлять хемотаксис, инициируемый AWC. Эти результаты подчёркивают важность понимания нейронных сетей как «аккумуляторов» всевозможных ассоциаций элементарных вычислительных процедур в рамках общей истории взаимодействия организма с сигналами из внешней среды.

Следуя за мыслью авторов статьи, можно обнаружить использование производной по времени также и в моделях, применяемых для объяснения свойств зрительной системы. Производная по времени применяется в объяснительных моделях ранней визуальной обработки сигнала в сетчатке. Д. Марр предлагает использовать первую и вторую пространственную производную для определения точек изменения яркости, а производную по времени для первичного анализа движения. Для эффективного обнаружения изменения яркости Марр и Хилдрет предлагают ∇^2*G-фильтр, который наиболее точно отражает экспериментальные данные относительно строения рецептивных полей сетчатки. Фильтр представляет из себя процесс (или устройство) измерения входных сигналов и преобразование их в выходные, подобно тому, как это происходит в рецептивном поле,

1. ∇^2 – это оператор Лапласа, $\frac{d^2}{dx^2} + \frac{d^2}{dy^2}$ дифференциальный оператор, который описывает рецептивное поле с круговой симметрией

2. G – распределение Гаусса, $e^{-\frac{x^2+y^2}{2r\sigma^2}}$ (гладкое, оптимально локализованное)

Фильтр должен быть использован, как экономичная вычислительная процедура для измерения разности яркости двух соседних точек. Д. Марр

[133] Cho C., Brueggemann C., L'Etoile N., Bargmann C. Parallel encoding of sensory history and behavioral preference during Caenorhabditis elegans olfactory learning // Elife. 2016.

рассматривает изменение яркости в виде вычисления значения G-фильтра, аппроксимирующего процедуру взятия второй производной. Для второй производной $\frac{d^2}{dx^2}$ или $\frac{d^2}{dy^2}$ изменения яркости соответствуют нулевым точкам (это удобно, потому что для первой производной в модели G-фильтра придётся искать локальные максимумы). В качестве прототипа такого фильтра предлагают ганглиозные клетки. Эти клетки известны тем, что формируют «off-центры» и «on-центры», организованные по типу центр-периферическое кольцо, таким образом, что могут максимально быстро определять изменение интенсивности света, попадающего в область рецептивного поля. Так как в силу, адаптации рецепторов к постоянной интенсивности ответная реакция клеток затухает, то имеет смысл предположить, что в закодированном сообщении находится именно момент пересечения световым пятном «on-зоны» или «off-зоны», такое пересечение может вызывать ответ в ганглиозной клетке в виде стереотипной активности. Резкое изменение интенсивности света происходит по всей поверхности сетчатки, оно одновременно приводит к формированию сложного узора, на который реагирует глаз. Если такому изменению яркости на рецептивном поле соответствует реальный яркостный переход (соответствующий некоторому объекту в окружающем мире), то, полагает Д. Марр, возможен переход от ориентированных отрезков, образованных из точек пересечения нулевого уровня к необработанному первоначальному эскизу. Такой анализ происходит уже в высших центрах обработки зрительной информации в латеральном коленчатом теле и зрительной коре. Хотя анализируемые на сетчатке данные описываются Д. Марром как точки и пятна, их трудно соотнести с элементами феноменального поля зрения. Это трудно в силу того, что представление этих данных в виде точек уже предполагает некоторую сложную феноменальную структуру (цвет, форма, временная протяжённость) и потому что непонятно какое место точки и пятна должны занять в сознании. С другой стороны, очевидно, что психологическая

интерпретация также как и элементарные изменения на сетчатке будет иметь гомологию черт, распознающихся вместе и отдельно, то есть будет соотносится с образом и временем предъявления. Обратив внимание на адаптивные свойства рецепторов и влияние истории взаимодействия колбочек и палочек со средой можно предположить - *элементарным событием на сетчатке будет именно переход светового пятна из off-центра в on-центр (или наоборот) так как в этом случае появляется сигнал об изменении относительного контраста.*

Похожую схему, Д. Марр предлагает для анализа движения[134]. Его модель основывается на G-фильтре, при анализе движения более чем в любом другом случае важную роль играет время, нужно улавливать только свежую информацию так как старая быстро становится не актуальной. При мгновенном быстром взгляде все кажется статичным пишет Марр, но если есть несколько мгновенных взглядов, то возможно уловить движение, отмечая различия в прошлом и настоящем. Фиксация изменений происходит уже на стадии первичного улавливания сигнала на сетчатке двумя рецепторами, они соединены между собой логическим элементом «И-НЕ», и если световое пятно активирует соседние рецепторы с задержкой во времени, то происходит передача сигнала; такой механизм может обнаружить направление движения. Но более интересным становится элементарное событие, которое распознается первым в нервной системе, это измерение значений производной по времени $\frac{\partial}{\partial t}[\nabla^2 * G * I]$. Известным фактом для зрительного восприятия будет то, что неподвижные объекты исчезают из поля зрения. Иллюзии фиксации глаз показывают, что адаптация глаза к какому-то зрительному стимулу приводит к исчезновению его из поля зрения, эффект пропажи сохраняется при повторном эксперименте. Также

[134] Глаза находятся в постоянном движении, синхронно двигаясь они ненадолго замирают, фиксируясь на интересном объекте, и снова как бы вздрагивают, резко меняя угол обзора. Во время фиксации они почти неподвижны, однако выделяются особенные микродвижения. Саккады и микросаккады необходимы для изменения угла обзора предмета, их функция состоит в получении сравнительной информации, что также показывает, что операция сравнения двух сигналов имеет значение для выделения искомого признака.

объяснение через изменение относительной светочувствительности может получить явление послеобраза. Так несмотря на факт ранней обработки можно заметить, что элементарное изменение сигнала во времени является информативным как в области феноменальных явлений, так и в области элементарных вычислительных операций.

Производная по времени рассматривается не случайно. В ИНС вычисление градиента используется для обновления весов в многослойных сетях. Регуляция весов во время обучения происходит небольшими шагами, приближая сеть к целевой функции. Эффективный метод обучения, состоит из дифференцируемых последовательных операций изменения проводимого сигнала на каждом нейроне. Как небольшое изменение в конкретном весе влияет на производительность всей сети вычисляться, как частная производная погрешности веса. Эффективные способы вычисления производных для всех весов рассматриваются во второй главе. Сначала сигнал по сети проходит от входных нейронов к выходным, и вычисляются все коэффициенты. Получается некое значение целевой функции. Затем для каждого нейрона вычисляется его влияние на полученный результат. Если нейрон имеет ненулевое значение сигнала на выходе, то его производная также будет ненулевой. Вычисление производной происходит относительно каждого веса. Влияние веса определяется активностью входных весов исходной единицы (преактивация) и чувствительностью к активации нейрона. Регуляция веса каждого нейрона направлена на уменьшение ошибки в выходных данных. Метод поиска оптимального значения всех коэффициентов обычно градиентный спуск. Градиентный спуск делает минимальные корректировки весов для уменьшения общей ошибки сети, последовательными шагами приближаясь к локальному минимуму целевой функции вдоль градиента. Несмотря на различные способы построения системы наблюдается одинаковый смысл вычислительных операций, производимых на каждом отдельном синапсе, а контекст-зависимое изменение проводимости, которое приводит к глобальной перестройке

проведения сигнала. Такие элементарные операции поиска производной характерны и для моделей вычислительной нейронауки и для нейросетевых алгоритмов настройки весов, а также находят подтверждения в биологических нейронных схемах. Такие элементарные функциональные примитивы позволяют перейти от уровня реализации к алгоритмическому уровню обработки сигнала в живых сетях нейронов и могут рассматриваться как канонические вычислительные операции.

1.7 Методологическая роль параллельной и распределенной обработки данных в современных когнитивных исследованиях

Стоит разделять вычислительную нейронауку и коннекционистскую вычислительную метафору. Б. Закман, известный нейрофизиолог, лауреат Нобелевской премии говорит: *«Я знаю все данные: типы клеток, свойства их активности, связность клеток, возбудимость дендритов, синаптическую динамику, но я не могу понять это, я вынужден это моделировать».* Эта цитата демонстрирует задачу вычислительной нейронауки – разрешить в виде модели проблему сложных разнообразных взаимодействий, которые возникают между нейронами и внутри нейрона. В рамках этого подхода, например, реализуется моделирование индивидуальных разнообразных типов нейронов (которых насчитывают около 10000), в то время как для коннекционизма предлагается набор изначальных абстрактных свойств нейрона, из которых предлагается вывести все возможные типы нейронной активности. Для моделирования мозговой активности набора этих свойств может быть недостаточно. Однако коннекционистская программа настаивает на возможности объяснения всех аспектов мышления, исходя из свойств, заданных в параллельной и распределенной вычислительной системе. Предложение по моделированию отдельных признаков взаимодействия нейронов, исходя из экспериментальных данных сталкивается с проблемой поиска свойств, объединяющих все виды нейронной активности. Такого рода вычислительные модели выражают взаимоотношения в живых нейронных

сетях. Моделирование взаимоотношений тысяч и сотен тысяч нейронов помогает выделить отдельные признаки, возникающие в активности этих клеток. Такие феноменологические модели стремятся правдоподобно воспроизвести все виды нейронной активности. Такие модели критикуют формальные нейронные сети за их неправдоподобность и нереалистичность. Однако коннекционистский подход в ответ на это возражение может указать на присутствие у коннекционистских моделей тех общих свойств, которые необходимы для моделирования и воспроизведения любого рода нейронной активности. Набор свойств в моделей коннекционистского типа вызывает большое число вопросов, и в отношении этого набора свойств, предлагаемого в качестве необходимых условий для возникновения сетевых взаимодействий, ведутся споры. Вычислительная нейронаука разрабатывает модели, которые предсказывают поведение реальных нейронов. В таких моделях выделяются признаки конкретных нейронных систем. Кроме того, прямого сравнения аппаратной части компьютера коннекционистского типа и мозга как машины здесь нет. Математические модели вычислительного типа возникают из-за большого количества данных о нейронах, взаимодействия меду которыми пытаются формализовать в исследованиях интегративных механизмов нейронной сигнализации. Отсчёт исследований подобного рода принято производить от А. Ходжкина и Э. Хаксли, описавших модель распространение потенциала действия по аксону кальмара, такая модель есть разновидность модели одного свойства живого нейрона. «Многосвойственные» модели представляют большой интерес, такие модели реализуют не одно, а несколько свойств нейрона. Они фиксируют в качестве параметров выброс медиатора, постсинаптический ток, свойства синапсов. Самые известные модели — это модель интегрирующего нейрона и модель «нейрон детектор совпадений». Такие модели имеют большую правдоподобность и основываются на результатах экспериментов нейробиологов.

Выводы

Таким образом предположение о важной роли нейрона в психике человека является общим местом, с которым согласны все, что подтверждено большим количеством наблюдений и экспериментов. Коннекционистская программа предлагает в качестве основы элементов нейронной системы предложить не сами нейроны, а сеть из нескольких нейронов. Свойства такой сети образуют фундамент для любых когнитивных функций. Коннекционистский подход предлагает механические устройства в качестве изофункциональных систем, воспроизводящих элементы психики. Коннекционизм исследует мышление как биологическое явление, которое возможно представить путём построения аналогичных механических устройств, воспроизводящих архитектуру и алгоритмы работы мозга. Таким образом коннекционистский подход предполагает возможным исследование психических свойств на основе физических свойств, которые раскрываются в функциональной организации мозга. История исследования нейронных сетей позволила математикам создать удачные и неудачные модели нейронной активности и позволила задать в виде архитектур коннекционистского типа алгоритмы преобразования данных, которые могут быть работающими моделями воспроизведения реальных механизмов обработки информации мозгом. Однако в настоящее время нет последовательного рационального способа соотнесения признаков, получаемых в результате изучения активности мозга с моделями коннекционистского типа. Нейросетевые модели могут быть применимы для описания различных эффектов восприятия: явление перцептивного гистерезиса, слепота невнимания, контекст-зависимое распознавание образов, но являются ли эти модели необходимыми для описания активности нейронов в мозге человека? Множество особенностей современных нейронных сетей не соотносятся с данными вычислительной нейронауки. А эти особенности являются невычитаемым компонентом мозговой активности

и от их присутствия зависит процесс обработки информации. Стоит отметить и детальнее изучить каким образом возможно применение коннекционистских архитектур для объяснения работы мозга. Эта проблема, по-видимому, обойдётся коннекционизмом, так как коннекционисты предлагают модель мыслительной деятельности человека, а не модель уровня реализации по Д. Марру, как можно было бы изначально подумать, рассматривая коннекционистскую архитектуру для процессов параллельной и распределённой обработки информации мозгом. Также отмечается, что нейросети могут эмулировать любого рода программы классического вида. Возможно, что возникновение когнитивных функций не строго привязано к биологическому субстрату. Тогда воплощение этих когнитивных функций в коннекционистского типа машине возможно, также как возможно воплощение этих функций в мозге, или в классическом компьютере. Вычислительная нейронаука — это подход, моделирующий психические функции посредством воспроизведения морфологических и функциональных систем свойственных мозгу, с помощью формулирования правил сетевых взаимодействий и большого количества особенностей активности этих клеток. Направление исследований вычислительной нейронауки то же что у коннекционистских моделей, однако в виду конкретных нейрофизиологических свойств и процессов, которые моделируются в этой области, вычислительная нейронаука далеко отстоит от коннекционизма. Конкретные нейрофизиологические детали нервной системы различаются для разных видов животных, поэтому всегда есть основания для проведения исследований по выявлению этих деталей. Коннекционизм утверждает, что намного перспективнее искать универсальные принципы для всякого рода мыслительной деятельности, воплощение которых в конкретном мозге может носить случайный характер. Коннекционизм утверждает, что для правильного вида репродукции когнитивных функций необходимо наличие параллельной работы нескольких нейронов, связанных между собой. Описанные выше элементы

строения нейронной сети предлагаются в качестве этого минимального набора свойств при разработке перцептивных моделей в системах ИИ. В первой главе перечисляются основные проблемы, которые могут возникнуть при разработке коннекционистской вычислительной модели и показываются свойства нейронных сетей коннекционистского типа, как набор из нескольких признаков живых нейронов, который оказался работоспособен для реализации некоторых когнитивных функций. В главе демонстрируются нестандартные вычислительные операции (производная по времени и нормализация), которые могут быть настоящими признаками, сближающими искусственные нейросети и естественные нейросети. Основной гипотезой в рамках вычислительной теории разума становится изофункционализм свойств живых нейронов и элементов нейрокомпьютера, который позволяет на основе механических элементов репродуцировать функции, реализуемые в мозге и гипотетически создать машину, которая может обучиться любого вида деятельности свойственной человеку и воспроизвести её. Мышление, сознание и способность к пониманию у такой гипотетической машины будут исследованы в следующей главе.

Глава 2

Обработка информации в коннекционистской системе

О том, что мышление человека представляет собой реальность отличную от телесной, говорили греки и ранние христиане. В новое время Р. Декарт первым высказал предположение, что сложные физиологические реакции могут быть объяснены путём разложения их на конечное число самостоятельных нервных механизмов. Механистическая картина, которую создал Р. Декарт в «Описании человеческого тела», представляла собой гидравлическую систему из множества тонких трубок, переплетённых в сложной схеме[135], но некоторые функции нашего тела не поддавались механистическим описаниям. По мнению Р. Декарта творческий процесс мышления и живая речь не могут быть разобраны на элементарные нервные механизмы и возникновение каждого нового акта мышления и речевого акта не подчиняется логике применимой для механического описания тела. Это может свидетельствовать о том, что акты мышления и речевые акты производятся не на основе механического процесса внутри гидравлического мозга, а на основе движений субстанции иного рода, которая может повлиять на решения о движении тела, через тонкие трубочки нервной ткани. Эта «субстанция души» является носителем ментальных состояний и находится в особом органе – шишковидной железе, там, где она очень сложным образом переплетается с материальными компонентами механического мозга, который приводится ею в движение. Современник Р. Декарта, Томас Гоббс, в то же самое время также рассуждал о природе мышления и пришёл к отличным выводам. Согласно Т. Гоббсу мозг создаёт чувственный опыт на основе действия объектов внешнего мира, которые вызывают в нем последовательности мыслей. Идеалом рассуждения для Т. Гоббса служит

135 Декарт Р. Сочинения в 2 т.-Т. 1.- М.: Мысль, 1989. - 654 с.- (Филос. наследие; Т. 106) С.423-460.

математическое доказательство, поэтому разум также ориентирован относительно законов логики, он производит суждения вычитая и складывая простые идеи, где способы переходов между мыслями обусловлены целью (например, блуждание мысли и стремление — это случайный поиск подходящей мысли в восприятии и памяти). Закономерности движения мысли, выделенные Гоббсом, не отличаются от закономерностей движения телесных вещей и могут порождать друг друга. Если разум подчинён тем же законам, что и тело из которого он сделан, то он не имеет отличной от тела природы, и представляет аппарат по вычислению идей[136]. Дуализм Декарта, имел большой успех, несмотря на проблемы с объяснением взаимодействия души и тела. В то же время проект Гоббса тоже развивался и нашёл своих поклонников. Различение любого рода опыта на элементарные идеи, соотнесённые между собой некоторыми математическими отношениями, — это проект Г. Лейбница по замене рассуждения исчислением[137,138,139]. Немецкий математик работал над проектом по разработке «алфавита для человеческих мыслей» на протяжении всей своей жизни. Его проект во многом предвосхитил современный когнитивный подход, но Г. Лейбница не интересовало соответствие элементов мысли некоторым элементам ткани мозга. Монадология Лейбница не оставила подсказок того механизма как некоторые монады достигают сознания. Однако именно эти первые попытки построения интеллектуальных автоматов стали отправной точкой исследования искусственных логических систем для обработки информации и изучения работы мозга как машины, обрабатывающей информацию.

Коннекционистские модели психической жизни отстоят от других, так как с самого начала понимаются как интерактивные и параллельные, что означает, что психические процессы встроены в физические носители из

136 Гоббс Т. Сочинения в 2 т. Т.1. М. Мысль, 1989.- 622с.- (Филос.насл. Т.107.) С.518-519.
137 Лейбниц Г. В. Порядок есть в природе // Он же. Соч.: в 4 т. М., 1982. Т. 1. С. 234.
138 Ягодинский И. И. Философия Лейбница. Процесс образования системы. СПб. 2007. С. 200.
139 Клюева Н. Ю. Влияние идеи г. Лейбница на развитие компьютерных наук и исследования в области искусственного интеллекта // Вестник Московского университета. Серия 7. Философия. 2017. №4. с. 79-92.

сетей нейронов и во многом определены свойствами архитектуры нейронной сети. Эта дистрибутивная модель была отмечена уже в рассуждениях Декарта. Ассоциативность, возникающая из-за того, что одна и та же складка памяти может соответствовать многим памятным следам в «мозге Декарта», пересекается с современной коннекционистской концепцией памяти[140]. Компьютерные архитектуры моделей PDP проясняют то, как обрабатываются сигналы когнитивным агентом, то каким образом происходит обучение и то каким образом представлены в мозге концепты. Они могут быть прямым продолжателем декартовского подхода по изучению телесного автоматического поведения, в то же время в рамках вычислительной теории разума PDP модели развивают лейбницианский подход к объяснению разума как машины для производства идей. Удивительно, но трудное место, озадачившее Декарта и не рассматривающееся Гоббсом как проблема, также осталось незамеченным когнитивной наукой. Только в середине 80-х годов XX-го века, когда психологические абстрактные когнитивные модели интеллектуальных навыков стали соотносить с данными нейробиологии, вновь возникла концептуальная трудность соотнесения сознания и тела. Состояния сознания и их отношения с мозгом могут быть интерпретированы с точки зрения самых разных философских концепций. Теории тождества сознания и мозга предлагают рассматривать процессы, протекающие в разуме и процессы, протекающие в мозге, как одни и те же процессы. Двухаспектные теории подразумевают дуализм в разной его степени выраженности. Здесь, как и в случае с Р. Декартом выделяются особенные внутренние феномены, которые обладают уникальным набором свойств.

Так как в современной нейробиологии явления субъективной реальности соотносят с динамическими процессами в различных структурах

140 Sutton J. Philosophy and Memory Traces: Descartes to Connectionism. Cambridge University Press, 1998. 372 p.

мозга (см. таблицу 3) [141], а динамические процессы в искусственных нейросетях — это информационные процессы, то в рамках коннекционизма в когнитивной науке явления субъективной реальности понимаются как информационные. Достаточны ли коннекционистские модели (PDP модели) для объяснения всех свойств психических явлений? Что они могут рассказать о динамике вещества в мозге, полагая что это движение организовано в соответствии с вычислением когнитивных функций? Могут ли они разрешить некоторые вопросы в отношении сознания?

	Нейроанатомические	**Клеточные**	**Нейрофизиологические**
Полные Нейронные корреляты сознания	Подкорковые системы возбуждения Ретикулярная формация ствола мозга Парамедиальный таламус	Ганглианарный слой и слой мультиформных клеток, 5 и 6	гамма-активность или гамма-синхронность ERP P3b
Контент-специфические нейронные корреляты сознания	Задняя кора (posterior cortex)	Пирамидальные нейроны субгранулярных слоев Пирамидальные нейроны супрагранулярных слоев	Активированная ЭЭГ Дифференциация и интеграция активности - например, сложность ответов ЭЭГ

Таблица 3. Нейронные корреляты сознания

[141] Koch, C., Massimini, M., Boly, M. et al. Neural correlates of consciousness: progress and problems. Nature Review Neuroscience. 2016. № 3, P. 307–321.

Основной процесс, запускающий когнитивное поведение в моделях PDP и сетях глубокого обучения, — это динамическая корректировка текущего состояния активации сети в соответствии со взвешенной суммой своих входных связей от других нейронов. Когнитивная обработка включает распространение активации по сети нейронов через связи с определёнными весами. Узор из таких элементов репрезентирует информацию в когнитивной системе о предметах мира и соответствует паттернам активации, генерируемым набором нейронов. Как уже было отмечено выше, мозг, понимаемый в качестве вычислительного устройства коннекционистского типа, хранит любого вида ощущения в сети нейронов, связанных между собой через синапсы. Модели PDP предполагают, что такие группы нейронов, распределённые по мозгу, хранят:

- элементы слов
- буквы и фонемы
- элементы визуального образа: цвет, движение, форма, глубина тоже будут хранится в таких группах клеток
- семантические и концептуальные компоненты мысли

Модель интерактивной активации Дж. Макклелланда и Д. Румельхарта, моделирующая работу памяти человека, подразумевает обработку информации в памяти человека в параллельном виде. Такая модель устанавливает двунаправленные связи между элементами различных модулей (категорий), а также ингибирующие связи меду двумя компонентами внутри одной категории (взаимоисключающие друг друга). Наличия таких связей достаточно для того, чтобы сеть правильно определила с какими модулями связан объект. Обработка данных внутри сети происходит в течение нескольких шагов, она развивается с течением времени, изменяя как его выходные, так и входные значения через систему обратных связей. Сеть изменяет свои значения весовых коэффициентов до достижения некоего динамического равновесия. Динамическое равновесие

может быть распределено между несколькими аттракторами и тогда сеть приобретает вид хаотической активности. Это особенно характерно для рекуррентных сетей, для которых состояние, в котором все элементы сети приняли определенные значения активации, заданные их входами, динамически изменяется, причём внутренние изменения могут быть значительными даже при небольших изменениях входных значений[142]. Устойчивые состояния внутри сети, названные аттракторами, могут интерпретироваться как репрезентация системой входных данных. Обработка сигнала также возможна только как интерактивная, Динамика аттракторов в моделях PDP особая большая тема для обсуждения, здесь модели PDP тесно пересекаются с моделями динамических систем, которые могут быть описаны с помощью коннекционистского концептуального понятийного словаря. Как известно из теоретических работ по моделированию активности корковых сетей млекопитающих, реальные сети нейронов также являются хаотичными[143], что сближает коннекционистские модели с работой сетей естественных нейронов. Сам процесс урегулирования активности внутри рекуррентной сети предоставляет возможный механизм для понимания того, как поведение во времени разворачивается в рамках когнитивной обработки сигнала.

2.1 Репрезентация знаний в когнитивной системе коннекционистского типа

Приобретённые знания закодированы в весах сети, а не в памяти, как особой структуре, хранящей данные. Развитие моделей PDP привело к так называемому золотому веку коннекционизма, эти исследования освежили дебаты о нативизме и эмпиризме в философии сознания. Ссылаясь на классические работы Д. Юма и Дж. Локка философы отмечают[144], что

142 Sompolinsky H., Crisanti A., and Sommers H. J. // Physic Review. Lett. 1988. Vol. 61. P. 259.

143 London M., Roth A., Beeren L., Häusser M., Latham P.E. // Sensitivity to perturbations in vivo implies high noise and suggests rate coding in cortex. Nature. Vol. 466. P. 123–127.

144 Silver D. and all. Mastering the game of Go without human knowledge // Nature. 2017. № 7676. Vol. 550. P. 354–359.

благодаря нейронным сетям стало возможно говорить о возникновении определенных типов знаний, которые возникают не из врождённых структур, но только благодаря чувственному опыту. Это вновь возрождает классический эмпиризм под эгидой коннекционизма в противоположность рационализму классических вычислительных моделей.

С точки зрения эмпиризма законы, установленные в результате опыта, состоят из ассоциаций ощущений, которые соединяются в сознании в результате предыдущего опыта. Любой новый возможный опыт может отменить эти законы. Так возникает относительность известных нам эмпирических закономерностей, полученных из опыта, которые могут быть изменены в результате нового опыта. Это приводит к тому, что эмпирическое суждение о связи явлений не обладает всеобщей и необходимой значимостью. Эмпиризм показывает, как ассоциативные связи явлений в сознании оказываются следствием закономерной повторяемости связей этих явлений в опыте. Вера в повторяемость явлений формирует предвосхищения, формирующие наши убеждения, однако они могу быть опровергнуты новым опытом. Опыт, который формируется на основе привычки, не приводит нас к невозможности иного опыта. Таким образом, эмпиризм критикует эпистемологическое обоснование принципов составления истинных высказываний о предметах мира, используя описание психологических механизмов установления связей между предметами. Такое описание на основе привычки, памяти и предвосхищения хорошо объясняется коннекционистскими моделями, однако следует разделять психологические механизмы формирования высказываний (или причинных связей) и эпистемологические инструментальные схемы, которые используются для установления истинности высказываний.

Скепсис эмпиризма возникает при переходе от рассмотрения философских теорий об установлении истинности между языком, мышлением и предметами этого мира к психологическим механизмам

формирования таких связей в разуме. Такое движение мысли не может быть оправдано, так как из психологических особенностей восприятия не следует истинность тех или иных умозаключений. Подобного рода проблема возникает также при рассмотрении естественных категорий и их психологических интерпретаций в прототипическом подходе, где психологический механизм выяснения принадлежности к категории отождествляется с самой естественной категорией[145]. Коннекционистские модели проясняют именно психологические аспекты установления связей, но не дают ответ на то, как нам поступать с такими высказываниями, а значит они применимы как для поклонников эмпиризма, так и сторонников рационализма.

Могут ли коннекционистские модели рассказать нам о невозможном для нас опыте, когда происходит переход к анализу условий возможности опыта? Согласно И. Канту важнейшей задачей нашего когнитивного аппарата является пространственно-временная обработка восприятий и применение понятий к сенсорным ощущениям. Пространство и время – особого рода предвосхищения, в соответствии с которыми устанавливаются любого типа восприятия и любой возможный опыт. Пространство и время – условия возможности опыта. Они представляют собой правила, согласно которым образуются любые знания, и они предвосхищают любого рода знание (условно назовём их врождённым знанием). Одновременность сосуществования различных предметов в одном акте сознания можно пояснить, вспомнив что в акте сознания сосуществуют различные предметы, доступные для манипуляции с ними, также, когда *«мы думаем о чём-то и в то же самое время судим об этом или желаем этого»* (как рассуждал Ф. Брентано) мы говорим об одновременности присутствия различных ментальных феноменов. Одновременная параллельная обработка состоит в том, что мозг может обрабатывать стимулы разного качества одновременно

[145] Кузнецов В. "Аристотелевская теория категорий и прототипический подход" Вестник Московского университета. Серия 7. Философия, no. 1, 2018, pp. 32-44.

и что это соответствует одновременному их присутствию в акте сознания (например цвет, форма, глубина одного предмета). Также уместно говорить, что параллельно происходят различного рода когнитивные процессы, такие как эмоции, восприятие и мышление которые сосуществуют вместе. В то же самое время физические феномены также происходят одновременно в различных точках пространства и объясняются каждый своей причиной (например, видимый цвет и слышимый звук). В современной физике причинная структура в пространстве-времени это лоренцево многообразие. Причинно-следственные отношения между точками такого многообразия объясняют какие события в пространстве-времени могут влиять на другие события. Формирующийся «причинный конус» связывает события через причинные связи и показывает к каким изменениям может привести активность конкретного элемента. Если роль того, что представляется в акте сознания состоит в потенциальной возможности взаимодействия с окружающей средой в рамках одного интенционального акта, то возможно говорить о связи феноменологии от первого лица с причинной структурой протекания физических феноменов, происходящих одновременно и параллельно в разных точках пространства и в разных временных отрезках. Таким образом параллельность и распределённость являются характерным признаком как системы обработки сигналов (мозг), таки и для феноменов окружающего физического мира, поэтому коннекционистская модель формирования психических феноменов может использоваться для разработки механизмов чувственности человека.

2.2 Память в моделях PDP

Одним из следствий функционирования алгоритмов в нейронных сетях является то, что долговременная память в сети не сохраняется в первозданном виде. Копии ментальных событий из прошлого — это не образы, извлечённые из памяти, как проявленные фото с негатива или фото из памяти компьютера. Вместо этого коннекционизм полагает, что следы

памяти представляют собой элемент изменения весов соединений, который они производят при активации паттерна. С этой точки зрения, воспоминание о предыдущем опыте включает несовершенную реконструкцию аспектов паттерна активации, соответствующего первоначальному опыту. Эта гипотеза соотносится с понятием реконсолидации памятных следов. Воспоминание возникает при новом опыте, который затрагивает старые связи, а также изменяет эти связи, так что долговременная память всегда является реконструкцией. Поэтому различия между обработкой нового сигнала, репрезентацией и актуализацией памятных следов, которые чётко размечаются в классическом символьном подходе, стираются для коннекционизма.

2.3 Проблемы несимвольного кодирования информации в искусственных нейронных сетях

Основные проблемы, которые вызывают дискуссии относительно распределённого кодирования информации – это отсутствие биологической правдоподобности распределённых репрезентаций концептов высокого уровня в мозге, проблема несимвольных вычислений и трудность с использованием знаково-символьного выражения применительно к интерпретации последовательного типа рассудочной деятельности. Биологическая правдоподобность распределённого кодирования информации относительно категорий высокого уровня постоянно критикуется. Гипотеза «бабушкиных нейронов», которые специализируются на знакомых словах, объектах и лицах предполагает, что на вершине высокоспецифичного психического различения одного концепта от другого стоят уникальные нейроны, рецептивным полем которых становится этот концепт. С другой стороны, большое количество данных по «бабушкиным нейронам» свидетельствуют о переспециализации нейронов бабушки, но феномен избирательного реагирования нейронов коры на стимул все ещё не объяснён. Часто можно услышать, что нейроны коры редко отвечают

спайковой активностью сразу на несколько стимулов, но очень часто высокая частота спайков возникает при демонстрации одного специфического стимула. Доказательств тому множество[146], и часто для каждого нейрона можно подобрать категорию, на которую будет такая ответная реакция. Также стоит отметить, что «концептуальные» нейроны могут реагировать на разные стимулы, для которых исследователи быстро определяют общую категорию на которую происходит реакция. Нейробиологам очень удобно выявлять и сопоставлять признак и активность нейрона. Более интересными результатами рабами являются те, где вместо селективных нейронов применяется редко-распределённое кодирование в памяти аттрактора сети[147,148]. Такая методология может быть очень наивна в своих выводах, так как сличает ответы нейронов с некоторым «ментальным театром» на языке «народной психологии». Этот язык «народной психологии» критикуется Полом Черчлендом и Патрицией Черчленд, которые последовательно показывают отсутствие здравого смысла в исследованиях сопоставлений ментальных феноменов с некоторыми нейронными активациями безотносительно теории обработки информации мозгом[149]. Представимо, что некоторую информацию, например, такой факт: «кофейный автомат на первом этаже сломался» просто невозможно донести до нейрона. Нейроны ничего не «думают» о кофейном автомате, так как на уровне нейронов не существует никакой концепции кофейного автомата, вместо этого существует сложный набор нейронных репрезентаций, информационных различий внутри ассоциативной зоны коры, но также, как и модели PDP не догадываются о кофейном автомате, так о нем и не догадываются сети биологических нейронов. Если такие нейрональные репрезентации имеют

[146] Fried I., Rutishauser U., Cerf and M. Kreiman G. Single Neuron Studies of the Human Brain: Probing Cognition. The MIT Press: Cambridge Massachusetts and London, England. 2014.
[147] Rolls E.T., Tromans J., Stringer S. Spatial scene representations formed by self-organizing learning in a hippocampal extension of the ventral visual system. Eur. J. Neuroscience. 2008. Vol. 28. P. 2116–2127.
[148] Kesner R., Rolls E. A computational theory of hippocampal function, and tests of the theory: new developments. Neuroscience Biobehavioral Review. 2015. Vol. 48. P. 92-147.
[149] Churchland M. A Neurocomputational Perspective: The Nature of Mind and the Structure of Science. Cambridge. MA: MIT Press. 1989.

вид локального или мелко-распределенного кодирования, то для искусственной нейросети должны быть правдоподобные методы построения такого же рода репрезентации. Строго говоря, сегодня существуют доказательства того, что нейроны гиппокампа реагируют на некоторую информацию очень избирательно, что только подтверждает предположение о локальном хранении информации о событиях далекого прошлого. Кроме того, Ф. Розенблаттом отмечается потенциальное ограничение возможности распределённого хранения информации и это ограничение связанно с катастрофой суперпозиции. С этой точки зрения наложение новых распределённых внутри сети представлений приводит к перемешиванию информации и невозможности восстановления определенного представления. Локального вида репрезентации в сети возможно создавать за счёт особого вида гауссовских нейронов. Исследования различных видов функции активации показывают, что распределённого типа репрезентации не является внутренними свойствами моделей PDP, а возникают при взаимодействии элементов определенного типа. Компоненты местного кодирования существуют для сетей глубокого обучения[150]. Отметим, символьный коннекционизм, или коннекционизм имплементации, обсуждал ещё Д. Марр. Он предлагает использовать нейронные сети только как среду реализации вычислительных операций классического вида. Модели символьного коннекционизма возникают в попытке объяснить композиционность языка и рассудочных операций и упорядоченность событий в долговременной памяти. Композиционность заключается в том, что части сложного выражения определяют его значение согласно набору правил и они сохраняют свою идентичность в разных контекстах. Трудность в определении таких символьных манипуляций в моделях PDP связана с утверждением, что для моделей PDP невозможно контекстно-независимое

[150] Bowers J.S. Grandmother cells and localist representations: a review of current thinking // Languge Cognitive Neuroscience. 2017. Vol. 32. P. 257–273.

представление и обобщение вне тренировочного пространства. Действительно, нейросети плохо усваивают идею причинности, но прекрасно обнаруживают корреляции. Из этого следует, что несимвольные системы кодирования не подразумевают процедуру вывода или обобщения, которая формулируется в виде абстрактного правила, применяемого в новой ситуации и с новыми типами данных. Приведём пример с самой простой формулой, которая может получится. Человеку нужно найти закономерность в последовательности 1+1=2, 1+2=3, 1+3=4, 1+4=5 и т.д. Человек может распознать функцию на нескольких примерах и продолжить мысленно ряд f(x)=x+1 до бесконечности. Формула не зависит от категории предметов, которые можно посчитать, тем не менее это тривиальное обобщение может быть проблематичным для нейросети. Например, научившись отвечать «один» на произнесённое слово «один», и «два» на произнесённое слово «два», нейросеть не уясняет абстрактную последовательность ряда чисел, но просто строит ассоциативный ряд между символом и звуком. У нейросети не появляется способность обобщать за пределами тренировочного пространства, что необходимо для множества элементарных когнитивных задач. Если классическая когнитивная архитектура для объяснения интеллектуальных действий использует символьную парадигму А. Ньюэлла и Г. Саймона, в которой обработка информации заключается в манипуляции символами в явном виде, то для дистрибутивной и распределенной архитектуры трудно представить как возможность процедуры аналитического вывода, если правила, управляющие сочетаниями символов, находятся на уровне связанных групп нейронов и не образуют реальную логическую систему внутри сети. Нейронные сети могут охватывать автоматические процессы, касающиеся чтения слов или образования морфем, но применяя модели PDP к аспектам семантического познания и многим

формам рассуждений, включая причинно-следственные, математические и силлогистические рассуждения, сохраняется скептицизм[151].

Все эти сведения позволяют некоторым авторам предложить, что коннекционизм не может быть использован для объяснения «языка мысли». Коннекционисты провели достаточно большую работу, направленную против критики подобной этой. Для удобства рассуждения разделим процессы, которые хорошо описываются коннекционистской архитектурой, и процессы, которые плохо ей описываются. В когнитивной психологии в настоящее время стала популярна теория дуального процесса мышления. Она может прояснить ограничения коннекционистских архитектур[152]. Согласно этой теории, существуют две системы, «Система 1» – бессознательные, параллельно реализованные процессы автоматического принятия решений, обычно описываемые коннекционистской архитектурой, и «Система 2», подразумевающая сознательные последовательные рассудочные действия гипотетико-дедуктивного типа, описываемые с помощью классической символьной парадигмы и сравнивающиеся со структурой стандартной двоичной логики. В рамках реализации когнитивных способностей последовательного типа «Системой 2» можно выделить речь и письмо, «ментальную арифметику» или любого рода вычисления в уме, мысленные перемещения во времени и процесс принятия решений. Рассмотрим каждый тип последовательной деятельности человека подробнее.

2.3.1 Принятие решений

В рамках малых временных масштабов для человека характерно существование психологического рефрактерного периода. Это такой период времени, который не может быть использован для решения следующей задачи потому, что человек все ещё размышляет над предыдущей. Например,

[151] Timothy T.R. Neural networks as a critical level of description for cognitive neuroscience // Current Opinion in Behavioral Sciences. 2020. Vol. 32. P. 167-173.
[152] Oaksford C., Oaksford M., Chater N. Dual processes, probabilities, and cognitive architecture // Mind Society. 2012. Vol. 1. P. 15–26

в эксперименте с двумя простыми задачами от испытуемого требовалось в первой задаче – нажимать на букву «Т» на клавиатуре, когда он видит зелёную рамку, а во второй – нажать букву «М» на клавиатуре, когда на экране появляется цифра «3» или нажать букву «С», если появляется цифра «4». Во втором случае человеку нужно было делать выбор и люди испытывали затруднение при параллельном их выполнении первой и второй задачи вместе. Решение одной задачи мешало выполнению другой[153]. То есть чем ближе по времени стимул из второй задачи к стимулу из первой задачи (50-150 мс после первого), тем больше времени потребуется для того, чтобы ответить на второй стимул. Это наблюдение указывает на узкое место переработки информации мозгом. Вовлечённый в процесс принятия решений субъект должен ориентировать свои действия во времени относительно одной задачи и не может выполнять другую задачу потому, что его внимание сконцентрировано в одном направлении.

Рефрактерный период указывает нам на невозможность параллельной работы когнитивной системы на уровне принятия решений, сознательно человек не может размышлять над двумя задачами вместе, но переключает внимание между ними. Пример демонстрирует функциональные ограничения мозга и некую последовательность мышления, которая возникает в перспективе «от первого лица» при решении задач. Это трудное место для коннекционистской архитектуры, потому что в основе коннекционистских моделей принятия решений лежит идея выбора из нескольких вариантов на основе предпочтения[154], которое зафиксировано в силе связей между различными альтернативами. В случае простого выбора между А и В механизм, описываемый коннекционистской моделью, представляет процедуру поиска по нескольким атрибутам, которые определяют предпочтение когнитивной системы варианту А или В. Если в

[153] Pashler H. Dual-task interference in simple tasks: Data and theory // Psychological Bulletin. 1994. Vol. 116. №. 2. P. 220–244.

[154] Angel I., Dolores del Castillo M., J., Serrano O. Connectionist Models of Decision Making // Chiang J. S. (ed.) Decision SuPort Systems. IntechOpen. 2010.

задачу добавить ещё одну пару альтернатив, то модель не изменится, но в эксперименте мы видим увеличение времени, требуемого для ответа на вторую задачу.

Коннекционистские модели обладают рядом преимуществ по сравнению с классическими моделями последовательной обработки сигнала. Они более гибкие и подходят для ситуаций с нечёткими альтернативами. Предпочтение в различных нейросетевых моделях может определятся рейтингом, порогом, правилом или эмоцией, что психологически правдоподобно. Например, «пороговая модель поля решений» пытается объяснить эффекты, возникающие при принятии решений, такие как сходство, компромисс или эффект притяжения на основе категорий, сформированных в прошлом[155]. Эта модель хорошо описывает «Систему 1», например, если человек выбирает автомобиль, то он оценивает разные свойства, такие как цена и качество, во многом автоматически согласно набору ранее усвоенных знаний. Однако для преднамеренных решений (для «Системы 2»), которые характеризуются последовательными мысленными представлениями, применение коннекционистской модели проблематично. Несмотря на то, что человек не осведомлён о процессах, лежащих в основе его выбора, он активно участвует в принятии решений и не является эпифеноменом автоматических процессов рассудочной деятельности[156].

Коннекционистская модель, как автоматическая система, определяет лучшего кандидата и человек осознает доминирующий вариант предпочтения (предпочтение, которое даёт наиболее связное мысленное представление). О соотнесении автоматических и преднамеренных процессов принятия решений рассуждает А. Глекнер[157], предлагающий

---

[155] Johnson J.G., Busemeyer J.R. A dynamic, stochastic, computational model of preference reversal phenomena // Psychological Review. 2005. Vol. 112. №. 4. P. 841-861

[156] Simon D., Snow C.J., Read S. The redux of cognitive consistency theories: evidence judgments by constraint satisfaction // Journal of Personality and Social Psychology. 2004. Vol. 86. P. 814–837.

[157] Glöckner A., Betsch T. Modeling option and strategy choices with connectionist networks: Towards an integrative model of automatic and deliberate decision making // Judgment and Decision Making. 2008. Vol. 3. №. 3. P. 215–228.

модель из двух сетей. Первая автоматически репрезентирует первичные варианты поведенческих реакций, а вторая сеть «преднамеренных стратегий» функционирует как вспомогательная система, помогая первичной сети выполнять свою работу, выбирая стратегии, которые помогают реструктурировать первичную сеть. Актуальный мысленный контроль посредством второй сети имеет свои ограничения, что может служить объяснением задержки при решении двух задач одновременно. Стоит подчеркнуть, что стратегия поиска последовательных решений по мнению некоторых авторов[158] не доминирует для повседневных решений, сознательный и преднамеренный контроль с точки зрения коннекционизма сильно ограничен и скорее представляет собой регуляцию автоматических процессов.

2.3.2 Манипулирование образами

Последовательность в ходе мыслительной деятельности также возникает при сознательном манипулировании образами, что иллюстрирует задача с мысленным поворотом. В эксперименте человек видит на рисунке две объёмные фигуры с разного ракурса, фигуры похожи, но не очевидно одинаковые они или не одинаковые. Решить задачу возможно путём создания образов фигур. Когда человек пытается развернуть фигуру в мысленном поле он может заметить, что мысленный поворот происходит поэтапно, как и в случае с реальными фигурами. Складывается впечатление, что человек симулирует поворот в поле ментального образа этой фигуры, что подтверждается исследованиями, показывающим похожие паттерны активности при работе с реальными и воображаемыми фигурами[159]. Удивительно, что в вопросе манипуляции объектами в мысленном поле зрения наблюдается линейное рассуждение. Почему результат вычисления не

[158] Bargh J., Chartrand T. The unbearable automaticity of being// American Psychologist. 1999. Vol. 54. №. 7. P. 462–479.

[159] Ganis G., Thompson W.L., Kosslyn S.M. Brain areas underlying visual mental imagery and visual perception: an fMRI study // Brain Res Cogn Brain Res. 2004. Vol. 20. №. 2. P. 226-241.

появляется сразу, без переходных шагов между образами? В частности, задача поворота фигуры требует последовательно перевести фигуру путём мелких шагов, причём мы задерживаем в актуальном поле ментального представления только один шаг, выводя его из предыдущего шага[160].

Образное мышление трудно изучить ввиду приватного доступа к этим визуальным, слуховым и тактильным квази-ощущаемым феноменам, кроме того, образное мышление плохо вписывается в современную теорию восприятия как процесса переработки информации, поэтому для коннекционизма образное мышление последовательного типа и аналитические выводы посредством образов проблематичны. Классическое психологическое объяснение образного мышления как интериоризации внешних структур опыта, к сожалению, ничего не сообщает о механизмах, лежащих в основе образного мышления. К моделям, проясняющим механизм возникновения образов, можно отнести теорию эмуляции репрезентации Р. Груша[161]. Идея состоит в том, что, помимо взаимодействия с телом и окружающей средой мозг может создавать нервные контуры внутри себя, которые создают образы для обеспечения сложной сенсомоторной связи с изменяющейся средой. Эти образы формируют ожидания, интерпретируют новые сенсорные сигналы в контексте прошлого, а также могут использоваться для синтеза когнитивных функций подобных речи и последовательному рассуждению.

Когда я подсчитываю мелочь какой когнитивный процесс ответственен за такое вычисление? Манипуляции образами неотъемлемая часть «ментальной арифметики» набора методов для устного счета, использующихся для работы с большими числами. Существуют методы устного счета, основанные на ментальных манипуляциях с абаком. Навык представления в уме элементарных операций, производимых на абаке,

[160] Shepard, R.N. Cooper L. Mental Images and their Transformations. MIT Press. 1982.
[161] Grush R. The emulation theory of representation: motor control, imagery, and perception // Behavior Brain Science. 2004. Vol. 27. №. 3. P. 377-442.

улучшает скорость вычислений больших чисел. Стоит отметить, что умножить числа в уме можно и столбиком, если визуализировать «метод столбика». Ментальная арифметика хорошо согласуется с «языком мысли» для которого когнитивная деятельность требует внутренней системы языковых репрезентаций и формальных синтаксических операций, которые могут применяться к этим репрезентациям.

Классическая вычислительная модель психического процесса описывает взаимодействие образов согласно формальным правилам, а семантическое содержание образа согласно его ментальной репрезентации. Однако для коннекционистской архитектуры процесс рассуждения не может быть представлен в качестве внутренней символьной манипуляции, с точки зрения коннекционизма когнитивная система не образует синтаксические структуры в явном виде, вместо этого сеть собирает информацию в процессе обучения, позволяющую ей узнавать составную структуру представленных предложений, и фиксирует информацию в значениях весов формальных нейронов. Нейронная сеть может взаимодействовать с символическими представлениями, такими как естественный язык или элементарная арифметика без необходимости создания символьной репрезентации внутри сети внешних знаковых систем в явном виде.

В рамках коннекционистского подхода логико-дедуктивное рассуждение может рассматриваться не как продукт самой когнитивной системы, которая выполняет внутренние логические манипуляции с символами, а как способность, которая развивается в системе, которая может распознавать образы и удерживать их с помощью внешних по отношению к ней систем символов. Такая модель предложена В. Бехтелем[162], она рассматривает ментальную арифметику, которая может служить примером логико-дедуктивного вывода вместе с системой символов внешних по отношению к этой системе. Коннекционистская система распознает

[162] Bechtel W. Natural deduction in connectionist systems // Synthese. 1994. Vol. 101. P. 433–463.

паттерны этих символов и, как продукт распознавания этих паттернов, генерирует дополнительные символы согласно выученным правилам. Результатом работы системы будут высокоструктурированные последовательности из символов, производимые согласно метаязыковым принципам. Стратегии построения символов — это классические распределённые паттерны активности сети, а рассудочная деятельность ума относительно следующего символа — это попытка предсказывания последовательности из этих символов.

Временные связи не заменяют причинные. Порядковая связанность не означает причинность. Но на этапе выяснения причины во время рассуждения человек использует гипотезу «после, значит по причине». Причинность предполагает круг связанных доводов, аргументов и оснований такого вида, что их достаточно для существования других. Если рассматривать категорию причинности с точки зрения её реализации средствами языка, то можно различить семантическую форму мысли, включающую в себя сами доводы, аргументы и основания, и логическую форму мысли, представляющую силлогизм как метаязыковую абстракцию. Силлогизм как логическая форма тождественен для любого рода содержания логического вывода, он служит ориентиром для построения аналитического вывода, в то время как семантические компоненты силлогизма меняются каждый раз и очень разнообразны по содержанию. Таким образом, нормативная, целенаправленная сознательная деятельность выделяет временные и постоянные компоненты в процессе аналитического вывода.

Анализ постоянных компонентов (или формы мысли) нейросетевыми моделями возможен с помощью категорий внутри сети и с помощью описания такой архитектуры сети, которая неизбежно будет приходить к такой форме рассуждения.

2.3.3 Речь и письмо

Линейность речи и линейность письма также являются примером последовательных мысленных операций. Речь и письмо образуют современную концепцию понятия знака, манипуляции с которыми происходят в мысленном поле. Манипуляции образами также можно отнести к линейному мышлению, как и манипуляции тактильными образами у слепоглухонемых в соответствии с азбукой Лорма. В общем случае сложные когнитивные процессы с привлечением сознания, использующие краткосрочную и долгосрочную память, ретроспективно представляются как последовательное рассуждение. Такое рассуждение в последовательном виде отражает отношения между явлениями условно закрепляя их в символьном виде. Символы отражаются в многочисленных формальных знаковых системах, ориентированных согласно свойствам «Системы 2».

Д. Элман разработал модель PDP для объяснения особенностей категоризации последовательного вида данных. Модель Элмана — это простая рекуррентная сеть с внутренним слоем контекстных нейронов в который копируется значение внутреннего слоя нейронов на каждом цикле обработки данных и сохраняется до следующего цикла. Такая сеть может сформировать категории глаголов и существительных на основании расположения слова в предложении (в английском языке глаголы всегда на втором месте). Как объясняет Элман такие сети служат для объяснения временной структуры внутренней репрезентации данных в коннекционистской архитектуре. Внутренние репрезентации в моделях PDP как правило состоят из шаблонов активации нейронов. Сеть Элмана может решать задачи со сдвигом сигнала во времени, определяя его как один и тот же, например 1 1 1 1 1 0 1 1 1 и 1 1 1 1 1 1 1 0 1 1 можно интерпретировать как сдвиг 0 в последовательности из 1. Благодаря тому, что сеть запоминает предыдущий шаг можно говорить, что в сети есть внутреннее представление о времени, которое характеризует стимул в зависимости от предыдущего

стимула, то есть сеть обладает временным контекстом. Ключевым аспектом обработки последовательности в модели Элмана в целом является способность сохранять информацию о контексте в котором происходят последовательные события. Слова в предложении связаны синтаксически и их значение определяется положением внутри текста. В предложениях: «У него косой подбородок» и «Я иду в сарай за косой» слово «косой» имеет разные значения в зависимости от контекста. Модель Элмана прекрасно различает эти два слова, разводит их по разным категориям, хотя формально они представлены одним и тем же набором знаков. Сеть не различает слова и каждое слово представлено в виде вектора на входном слое нейронов, но после обучения сеть может предсказать следующее слово и определить его категорию. С помощью обучения взаимодействия с такими категориями мы можем обучаться категориям слов даже не имея интенции что-либо запомнить.

Входные значения сети Эльмана представляют собой информацию о данных об окружающей среде в настоящее время, а ее выходы представляют собой предсказание модели о предстоящей информации так, как произнесение слов разворачивается в течение долгого времени. Нейроны первого слоя связаны со слоем скрытых нейронов, которые, в свою очередь, связаны со слоем выходных нейронов. Также сеть имеет дополнительный слой единиц (слой контекстных нейронов) который сохраняет копию шаблона активации скрытого слоя на предыдущем шаге обработки сигнала, что позволяет сети ассоциировать элементы во временном контексте. Эльман продемонстрировал, используя сеть, которая не включает словоформ в качестве репрезентативных элементарных единиц, как возможна категоризация во время обучения относительно границ слов. С каждой новой буквой сеть обучалась предсказывать следующую букву. После обучения частота появления ошибок прогнозирования, как правило, была высокой на границах между словами и низкой для букв внутри слов, поэтому ошибка

прогнозирования могла служить сигналом для категоризации. Эта же схема применима и к речи. Таким же методом на основании только расположения в предложении можно отделить глаголы в английском языке, что отразилось на шаблоне активации внутри сети. Например, все глаголы имели тенденцию вызывать очень похожие шаблоны активации, которые сильно отличались от существительных. Напомним, что сеть не была специально обучена генерировать семантическую или грамматическую информацию, а просто учитывала предыдущее слово в предложении. Тем не менее, внутренние представления, возникающие в результате обучения в этой задаче, по-видимому, захватывают важную грамматическую и семантическую информацию. Чем объясняется это явление? Ответ заключается в том, что такая информация присутствует в статистических характеристиках обучающей среды и полезна для предсказания последующих слов. Обучение сети Элмана предсказанию каждого последующего слова заставляет ее назначать одинаковые веса связи и, следовательно, аналогичные внутренние представления элементам, которые делают аналогичные прогнозы. Таким образом, семантическая и грамматическая структура выученных репрезентаций возникает как следствие необходимости оптимизировать прогнозирование. Однако, как и в случае сегментации слов, структура не представлена прямо или прозрачно – она скрыта в структуре паттернов активации (то есть в силе связей нейронов), генерируемых в ответ на каждое слово, что, в свою очередь, является результатом процесса обучения, чувствительного к статистической структуре языка.

Модель Элмана продемонстрировала все основные особенности обучения грамматике не в виде заучивания правил, а в виде категоризации входных данных на основе подобия их структуры без применения синтаксических правил. Например, в английском языке в прошедшем времени правильные глаголы заканчиваются на -ed, а неправильные приходится заучивать. Англичане в детстве часто делают ошибку, ставя

неправильный глагол в правильную форму с окончанием -ed. Символьные модели объясняют этот факт с помощью построения синтаксических деревьев, которые в символьном виде обрабатывают данные. Сеть Элмана делает такую же ошибку, хотя не обрабатывает текст в символическом виде. Грамматические правила могут быть в неявном в виде (субсимвольно) закодированы нейронной сети и без построения дерева синтаксического анализа, никаких врождённых принципов грамматики не требуется.

2.4 Информационные процессы в искусственных нейронных сетях

Как уже было отмечено выше, процесс вычисления когнитивной функции заключается в аппроксимации правильного вида непрерывной функции, которую пытается найти нейросеть во время обучения. Пространство состояний для репрезентаций многомерное, так как каждому весу соответствует измерение, и континуальное, так как изменение весового коэффициента происходит в поле действительных чисел. В этом пространстве состояний каждой точке соответствует некоторое различимое изменение, которое изменяет паттерн активации. Все потенциальные репрезентации находятся где-то в этом пространстве, что обеспечивает естественную среду для выражения непрерывного сходства между различными репрезентациями. Эти сходства, в свою очередь, обеспечивают основной механизм ассоциации в моделях PDP.

Важно, что репрезентация и обучение в каждой сети производят набор уникальных для сети паттернов активации. Результатом одного и того же входного сигнала для двух разных сетей могут быть совершенно разные результирующие паттерны активации, соответствующие альтернативным интерпретациям неоднозначной фигуры, слова или предложения. Даже при условии одинаковой реакции двух сетей наблюдается варьирование паттерна активации в информационном пространстве каждой сети. Сложная когнитивная структура коннекционистского типа будет содержать элемент

индивидуальности и уникальности в обработке сигналов, обучении и репрезентации, поскольку обучение происходит в вариативной среде сети.

2.5 Философское осмысление понятия информации

Классический путь к интерпретации философского понятия информации сводится к противостоянию двух традиций толкования природы информации: материализма и идеализма. Информацию могут содержать переживания и ощущения субъекта, а, так как информации не свойственна пространственность, то её можно трактовать как мысли, которые также не описываются в пространственных терминах. С другой стороны, материалистическое представление информации полагает применимость этого понятия ко всей (или к части) материального мира, таким чисто количественным измерением и определением информации занимается теория вероятности (Байесовская статистика), теория связи Шеннона, квантовая механика и термодинамика.

Сегодня информация является ключевым понятием в науке, говорят, все что мы знаем о мире – есть информация. Информация используется для обозначения количества данных, кода или текста. Любое знание можно закодировать, составить библиотеку таких данных, содержащую все знания о мире, то есть представить знания в виде последовательности кода. Такое мировоззрение развилось вместе с информационными технологиями. Сегодня эта традиция известна, как дигитальная философия и дигитальная культура.

Конструируя цифровую модель мира, мы рассуждаем о возможности симуляции посредством техники любого аспекта реальности окружающего мира. Такой метафизический образ Вавилонской библиотеки представляет все возможные сочетания букв, собранные в коллекцию. Простым перебором можно получить все возможные тексты с описанием всех возможных событий. Дигитальная философия в своём крайнем проявлении, как монистическая натурфилософия, использует цифровой код в качестве

субстанции. Возможности современного конструирования систем виртуальной реальности позволяют представить информационное пространство, в которое погружается субъект, аналогично тому как субъект погружается в реальный мир. Само информационное пространство концептуально может пониматься как система из дигитальных объектов взаимосвязанных каузальными отношениями на основе сетевых принципов. Из такой структуры конструируется мир, например, в виде коннекционистской модели обратимого универсального клеточного автомата[163]. В другом варианте материалистической трактовки понятия информация, рассматривается отдельный уровень взаимоотношений, возникающий при усложнении химических автокаталитических реакций. Самовоспроизводящиеся рибонуклеотиды в рамках гипотезы мира РНК демонстрируют, что помимо химических и физических взаимодействий для этих молекул существуют информационные взаимоотношения. Различия в последовательности мономеров в этих нерегулярных гетерополимерах определяют форму и каталитическую активность. Аналогия такой кодовой зависимости используется в информационном подходе Д.И. Дубровского[164]. Для когнитивной науки в качестве концептуального моста для перехода от физических состояний мозга к психическим явлениям различного толка предлагается использовать информацию. Функция нейрона в коннекционизме меняется в зависимости от его связей с другими нейронами. Функциональное состояние, которое определяется через каузальные взаимоотношения с окружением в функционализме, определяют функциональную роль, реализуемую конкретно для каждого ментального состояния. Таким образом через различия совокупности каузальных отношений предлагается типологизировать ментальные состояния. Кроме

[163] Fredkin E. An Introduction to Digital Philosophy // International Journal of Theoretical Physics. 2003. Vol. 42 №. 2. P. 189–247.

[164] Дубровский Д.И. Проблема «Сознание и мозг»: Теоретическое решение. М.: «Канон+» РООИ «Реабилитация», 2015. 208 с.

того, эти типы ментальных феноменов будут в некоторых теориях отождествляться с информацией, понимаемой в виде кода.

Информация, как концепт, буквально создана чтобы проявлять двухаспектность, то есть наличие структурных свойств, реализованных на физическом носителе. Информация удобна для описания состояний феноменальных переживаний, которые не привязаны к конкретному физическом процессу, которые не могут быть описаны в физических терминах (не имеют массы, энергии и импульса). Субъект от первого лица сознает различные качества физических феноменов, переживает различного вида опыт. Так как субъект осведомлён о переживаемом феноменально зрительном или слуховом опыте он обладает знанием об этом опыте. Это знание несёт информацию о качествах субъективной реальности и в рамках некоторых теорий предполагается, что эта информация о качествах субъективной реальности тождественна каузальным структурам, обеспечивающим её воплощение в материальном носителе, что приводит нас к возможности «чтения мозга» для «чтения мыслей».

Философским осмыслением данной проблематики активно занимается отечественный философ Д.И. Дубровский. Круг его интересов связан с непосредственной «расшифровкой мозговых кодов» «феноменов субъективной реальности» (СР), для решения этой задачи формируется информационный подход к решению трудной проблемы сознания. Далее реконструируируется его подход к решению трудной проблемы сознания. Как мы знаем из формулировки Д. Чалмерса: *«Неоспоримо, что некоторые организмы являются субъектами опыта. Но остаётся запутанным вопрос о том, каким образом эти системы являются субъектами опыта. Почему, когда наши когнитивные системы начинают обрабатывать информацию посредством зрения и слуха, мы обретаем визуальный или слуховой опыт — переживаем качество насыщенно синего цвета, ощущение ноты «до» первой октавы? Как можно объяснить, почему существует нечто, что мы*

называем «вынашивать мысленный образ» или «испытывать эмоции»? Общепризнано, что опыт возникает на физическом фундаменте, но у нас нет достойного объяснения того, почему именно он появляется и каким образом. Почему физическая переработка полученной информации вообще даёт начало богатой внутренней жизни? С объективной точки зрения это кажется безосновательным, однако это так. И если что-либо и можно назвать проблемой сознания, то именно эту проблему.» Проблема заключается в самом контенте сознания, который не удаётся разобрать на стандартные функционалистские объяснения. Проблема соотнесения явлений феноменального сознания с функциональными мозговыми системами разного толка активно разрабатывается в когнитивной науке начиная с восьмидесятых годов прошлого века. Коннекционистский подход, развивающийся в рамках когнитивной науки, использует информацию в качестве объяснения психических феноменов посредством воспроизведения структурных особенностей мозга и типологизации их в соответствии с ментальным феноменами. Подход Д. И. Дубровского предлагает разрешить трудность в определении информационного процесса и понятия информации применительно к реальным структурам мозга. Для начала автор формулирует тезис о субъективной реальности. Отметим, что понятие СР подробно разработано автором в нескольких статьях. Качества субъективной реальности предлагаются как особый вид явлений, известный человеку из его опыта нахождения в состоянии бодрствования, это знание от первого лица. Д. И. Дубровский перечисляет все известные на сегодняшний день виды субъективных явлений: «*Понятие СР охватывает как отдельные осознаваемые явления и их виды (ощущения, восприятия, чувства, мысли, намерения, желания, волевые усилия и т. д.), так и их целостное персональное образование, объединяемое нашим Я*[165]». Интуитивно не понятно, как в явлениях СР с неопределённой феноменальной структурой,

[165] Дубровский Д. И. Проблема «Сознание и мозг»: Теоретическое решение. М.: «Канон+» РООИ «Реабилитация», 2015. 208 с.

выявить, или оценить количество информации? Какие из объективных свойств отражаемого объекта репрезентируются в форме СР и являются внутренним проявлением СР?

Три тезиса Д.И. Дубровского:

1) Информация необходимо воплощена в своём физическом, материальном носителе.

2) Информация инвариантна по отношению к физическим свойствам своего носителя (т.е. одна и та же информация может быть воплощена в разных по своим физическим свойствам носителях, может кодироваться по-разному.

3) Явление субъективной реальности (скажем, переживаемый мной сейчас зрительный образ экрана компьютера или желание включить его) есть информация об определенном объекте или действии[166]*;*

Д.И. Дубровский предлагает, что информация необходимо воплощена в своём физическом, материальном носителе. Тезис указывает на некоторую последовательность каузальных отношений внутри физического объекта, интерпретируемых как информация. Отделение информации от носителя могло бы породить множество метафизических спекуляций на тему души и первая посылка ограничивает такие интерпретации. Однако остаётся непрояснённым отношение Д. И. Дубровского к множественной реализуемости информации на разного рода носителях. Неясность в определении такой информации сохраняется применительно к субъекту. Трудной проблемой становится семантический аспект «информации», который не может быть понят как техническое определение количества связанных сигналов. Если переместиться в будущее, где уже строго обоснован рациональный выбор нейродинамического процесса на роль

[166] Дубровский Д. И. Проблема «Сознание и мозг»: Теоретическое решение. М.: «Канон+» РООИ «Реабилитация», 2015. 208 с.

эквивалента элементарного информационного явления в мозге, этот выбор должен в первую очередь обосновать семантический аспект информации.

Вторая посылка говорит об инвариантности информации по отношению к её носителю. *«Одна и та же информация может быть воплощена в разных по своим физическим свойствам носителях, может кодироваться по-разному».* Эта посылка в дополнение к предыдущей увеличивает объем понятия «информация». Например, слова «ясу», «шалом», «гомар джоба», «нихао», «гутн так\хой», «ассалам алейкум», «амантрана» и другие содержат в себе некую одинаковую информацию - знак приветствия, и это их все объединяет. Но что их все объединяет кроме общей функции? Идея об инвариантности информации Д. И. Дубровского похожа на идею множественной реализуемости функции Х. Патнема (Х. Патнем отождествляет ментальное состояние с функцией). Любое сознательное состояние также соответствует своей функции[167]. Но Д. И. Дубровский обращает внимание на информационные процессы, а не функциональные каузальные отношения внутри системы. В нервной ткани эти информационные процессы предназначены для эффективного функционирования организма. Однако сами они не редуцируются к их функции, но только ею описываются, их природа изначально определена и место, которое они занимают, обосновывается как отдельный род сущего. Их специфическая функция состоит в том, что они есть представления, суждения и воления, однако они не редуцированы к этой функции.

По аналогии с ДНК предполагается кодовая зависимость и эта специфическая область действия при которой происходит перекодирование распределённых данных в новый вид «активности», специфика этой активности отражает специфику информации в СР. Явления СР не просто

[167] Информация необходимо воплощена в своем физическом носителе, но так как носитель одной и той же информации может быть разным по величине массы, энергии, пространственно-временным свойствам, то сугубо физическое объяснение функционирования информационного процесса, информационного воздействия становится несостоятельным. Д.И. Дубровский Критический анализ теории сознания Пенроуза–Хамероффа. Часть 2 // Философия науки и техники 2017. Т. 22. № 2. С. 89–102

несут в себе некоторое количество информации, они сами и есть информация. В этом смысле, конечно, понятие информация отдаляется от сознания-функции Х. Патнема. Кроме того, здесь появляется возможность описания функции как отображения, которая соотносит явления СР с некоторыми физиологическими состояниями. В работе Л.М. Веккера «Психические процессы» (модель соотнесения образа СР с физическим сигналом) рассматривается возможность отображения в виде информационного кода явлений СР. Указывается принцип изоморфизма –взаимной упорядоченности двух множеств состояний. Элемент множества X однозначно сопоставлен элементу множества Y тогда и только тогда, каждый конкретный элемент Xi сопоставлен с конкретным элементом Yi. Любая функция f выражающая отношения Xi-Xj на множестве X однозначно сопоставлена с функцией F выражающей отношение Yi-Yj на множестве Y. Соответствие отношения пары элементов одного множества с отношением пары элементов другого множества обеспечивает однозначное соответствие элементов данного множества. Такое соотношение интерпретирует функциональные отношения в отличном от Х. Патнема смысле. Функция – отображение, а не каузально связанный набор элементов**.** Взгляд на сознание, как на одну из когнитивных функций, не проясняет, где и как эта функция реализуется в мозге и соответственно каким образом она может быть воспроизведена в машине, кроме того, он не позволяет выделить специфику сознания. В то же время, предложение об особом роде явлений СР может дать точку опоры для дескриптивного анализа феноменологической структуры сознания. Информационные состояния будут элементами СР, соотносящимися с контент-специфическими анатомическими и физиологическими коррелятами сознания.

Интроспекция играет важную роль при изучении сознания и важно понимать, что самоотчёт человека и его осведомлённость относительно нахождения его в сознании необходимы при изучении сознания. Такой

подход предполагает изучение психофизиологических закономерностей процессов восприятия с участием испытуемого. Изучение феноменов отложенного сознания, феноменов угасания сознания, потери доступа к содержимому сознания, спутанности сознания, альтернативных состояниях сознания, соматогнозий, слепоты невнимания, феноменов слепозрения невозможны без отчёта человека, сообщающего нам о своём состоянии от первого лица. Некоторые философы утверждают, что не существует никакого состояния сознания без возможности произвести отчёт и сообщить о нем, как например предлагает Д. Деннет. Некоторые такие самоотчёты хорошо согласуются с представлениями об канонических вычислительных операциях. Например, феномен перцептивного гистерезиса описывает то, как объекты, воспринимаемые при медленно убывающей освещённости, перестают быть видимым, когда освещённость падает ниже критического уровня, однако, когда освещение начинает плавно возрастать, нам необходим более высокий уровень освещения для различения предметов. Такой эффект может быть объяснён нормализацией и контекстным распознаванием[168]. Данные явления сознания здесь прямым образом сличаются с производными по времени. Для наглядности приведём пример «лягушки в кипятке[169]» или «парадокса кучи». Лягушка выпрыгнет из нагретой воды, но, если температура поднимается недостаточно быстро, она не заметит изменения и сварится. Отметим, что аргумент анти-светимости Т. Уильямсона также строится на идее о реакции на изменение, как основе феноменального различения. Для такого различения необходим контраст и минимальный несократимый временной промежуток за который этот контраст реализуется. То есть рассмотрение реакции на изменения— это общее свойство сознания.

---

[168] Костелянец Н. Б., Левкович Ю. И. Зрительное распознавание при предваряющей или запаздывающей настройке наблюдателя на определенный набор изображений // Журн. высшей нервной деятельности. 1982. Т. 32. № 2. 292 с

[169] Приведенный пример носит иллюстративных характер, проясняющий общую интуицию, в реальности лягушки постоянно двигаются, что позволяет им соотносить изменения температуры не только в пассивной, но и в активной форме.

Проблема «светимости» от части решается с помощью введения понятия СР (хотя без объяснения функциональной роли явлений СР нет смысла говорить о сознании с точки зрения когнитивной науки), а проблема разделения феноменов СР на составные элементы решается с помощью информационных состояний как дифференциалов. Субъективные семантические пространства строятся аналогичным образом. Следовательно, можно гипотетически представить сознание для коннекционистской системы как интерактивный режим реализации множества когнитивных функций, представленных и обрабатываемых вместе. Решение проблемы связывания и других свойств феноменальной структуры сознания требует отдельной разработки.

Мозг содержит примерно 10^{14} синаптических связей, возникающих и пропадающих на всем протяжении жизни. Задача нейрона состоит в том, чтобы преобразовывать сигналы (до 10 000 на одну клетку) в виде волнообразного перехода подпорогового потенциала в потенциал действия. Такие сложные нейродинамические трансформации происходят одновременно по всему мозгу. Если каждый синапс производит выброс нейромедиатора примерно 100 раз в секунду (100 Гц), общее количество параллельных операций по обработке информации, выполняемых мозгом, должно составлять примерно $10^2 \times 10^{14}$, или 10^{16} операций в секунду. Нейронаука и компьютерное моделирование нейронных сетей мозга в рамках коннекционистских моделей предполагает подобный вывод. Такой подход, предложенный ещё Дж. фон Нейманом, сводится к умножению отдельных распределённых по мозгу потенциалов действия, возникших в одно и то же время (легко представить как входной вектор), на коэффициенты динамической матрицы из синапсов, это даёт выходной вектор в виде новой волны потенциалов действия. Эта громадная машина по обработке сигналов как только что поступивших, так и возникших внутри мозга, представляет любые знания и навыки, которые приобретает мозг. Нейробиологи говорят

об «активности мозга», допустимо говорить, что любые события в нервной системе составляют такую активность. Активность мозга определяется коннектомом и движением вещества коннектома. ИНС претендуют на репродукцию, а не только на моделирование когнитивных функций, в том числе и когнитивных функций высокого уровня.

Как показал анализ коннекционистского подхода предыдущей главы, такими функциональными примитивами являются синаптические связи, они переключаются и эти переключения трактуются как обработка информации. Сигналы среды поступают в активные центры коры головного мозга и переводятся в различного рода мышечные движения. При таком объяснении «психическая информация», или сознание, не выделяется в отдельного рода процесс, а также не специфицируется относительно выполняемой функции. Эта нестыковка может быть частным случаем философского зомби. Она иллюстрирует провал в объяснении психических феноменов на основе физических процессов. Параллельная обработка сигнала мозгом и его архитектура позволяют предположить распределённую в сети репрезентацию данных в виде весов связей, определяющих силу связи между элементами нейронной сети. Однако для когнитивной науки важна не гомология между хранимой информацией в модельной сети и функциональной сетью мозга, но важно объяснить каким образом и с помощью какого рода процессов мозга самоорганизовывается сложное поведение субъекта, включающее его представление о самом себе. Поэтому сведение ментальных феноменов и физических феноменов методом пошагового соотнесения хоть и признает проблему ментального и физического, но строит функциональную связь двух типов явлений искусственным образом в виде отображения. В предыдущей главе показано как операция производной по времени, и операция нормализации могут быть интерпретированы как функциональные корреляты сознательных состояний,

однако подлинной необходимости в существовании сознания в этом случае нет.

Многие философы (например Д. Чалмерс) предлагают определять сознание как фундаментальный элемент вселенной для объяснения которого невозможно предложить редукционистскую программу. В этом случае действительно нужно ориентироваться на эмпирические исследования сознания. Поиск функциональных коррелятов сознания становится предметом научного поиска. Также как ДНК являет собой новый уровень информационных кодовых взаимодействий, так и субъективная реальность принадлежит к такому новому уровню. Как моделирование когнитивных функций не приводит к появлению сознания, так и частичная репродукция когнитивных функций в ИНС не проявляет явления СР. Кодовой зависимости может быть недостаточно для объяснения всех аспектов природы сознания. Возможно сознание появляется в мозге или репродуцируется в искусственной системе в результате иных процессов.

Третья важная посылка, которая предлагается для прояснения вопроса о природе сознательных переживаний: явление субъективной реальности Y — это информация об определенном объекте или действии X. Всякому такому Y сложно подобрать X в качестве нейродинамического кода. Здесь можно отметить, что сама по себе гипотеза о тождестве ментальных феноменов и их нейродинамических носителей должна быть проверена в эксперименте. Такая научная гипотеза, которая предлагает нам соотнести всякого рода ментальные феномены с их нейродинамическими носителями. Несводимые к проявлениям поведения понятия о чувственных содержаниях (боль, удовольствие, красный), которые возможно верифицировать только с помощью интроспекции и каких-либо активных состояний мозга, являются информационными состояниями мозга. Попробуем объяснить это на уровне бытовых понятий: сами по себе состояниями мозга, соответствующие ментальным феноменам никакие не красные или зелёные. Когда я

представляю красный цвет, то не возникает краснота, но возникает информационное состояние моего мозга, которое я интроспективно воспринимаю как ощущение красного. Однако для такого проекта необходима дескриптивная феноменологическая теория, последовательно переводящая интроспективные наблюдения в понятия языка. Но как возможна такая теория и как её возможно верифицировать и фальсифицировать (проблема интроспекции)? Также стоит отметить, что последовательное сличение ментальных состояний с конкретными процессами в мозге, необходимо произвести на уровне типов, а не индивидуальных явлений. Но ни о каких типах ментальных процессов, которые могли бы быть соотнесены с типами физических процессов в мозге теория нам не сообщает. И если предполагается случайное тождество физического и ментального, то возникает трудность с проверкой такой теории, хотя логическая возможность проверки остаётся. Также стоит отметить, что теория, в которой утверждается, что А не есть Х, но А необходимо связанно с Х, является версией дуализма, а не формой материализма.

Решением проблемы соотнесения ментального и физического на уровне типов озабочен Л.М. Веккер. Он вводит шкалу уровней изоморфизма для прояснения типов ментального и того, что можно было бы точно сопоставить, а что более проблематично сопоставить (для Y и X). Но даже при доказательстве наличия такого строгого соответствия остаётся непонятно, что считать минимальным феноменальным различием или атомом ментальных ассоциаций. То, что следует назвать «простой идеей» Локка или «простым впечатлением» Д. Юма? У. Джеймс пишет, что школа ассоцианизма имеет дело с такими простыми ощущениями, как продуктами различений, доведённых до высшей степени[170]. Общие свойства таких элементарных различений ещё не найдены. Такие противоречия возможно связаны с неясностью в отношении самого понятия сознания и недостатками метода

[170] Джеймс У. Научные основы психологии. СПб.: Санкт-Петербургская электропечатня. 1902. 192 с.

интроспекции. Некоторые авторы указывают на то, что понятие сознания противоречиво.

В.В. Миронов выделяет несколько дихотомий:

1. Имманентное – трансцендентное. Сознание всегда моё, в него никто не может проникнуть, но сознание одновременно и нечто трансцендентное или «сверхличное».

2. Субъективное – объективное. Сознание есть нечто существующее и необратимо протекающее в потоке образов, ассоциаций, воспоминаний. Но при этом в нем могут присутствовать общие ценности, архетипы и т.д.

3. Сознаваемое – неосознаваемое.

4. Свободно – несвободно.

Здесь В.В. Миронов суммирует интуиции относительно природы, происхождения и функционального содержания сознательного опыта. Поразительный вывод В.В. Миронова касается самих свойств которыми должно обладать сознание: – *С точки зрения онтологической позиции более широкой, предлагается говорить о том, что само сознание принципиально представляет такую вот дихотомическую сущность, и оно принципиально само противоречиво.*

Такая палитра мнений может быть объяснена. Каждый «пользователь» сознанием уверен в собственной компетентности относительно явлений субъективного опыта, феноменальных явлений и квалитативных ощущений. Каждый своего рода эксперт в своём теле, но сознание является горизонтом всякого опыта вообще, поэтому определение сознания, и об этом говорит В.В. Миронов, всегда совпадает с границами опыта отдельного человека, а значит формирует мировоззрение этого отдельного человека.

Большая победа современной философии в том, что существуют общие мировоззренческие установки, с которыми соглашаются практически все, когда говорят о сознании. Эти основные дефиниции, с которыми не спорят, и которые принимаются в современной философии как разумные гипотезы

относительно сознательного опыта, есть набор свойств присущих сознанию. Их предлагается объяснять при разработке подходов в системной нейробиологии при моделировании и репродукции сознательных феноменов на базе ИНС.

Сознание обладает следующими характерными свойствами:

1. Квалитативность (Ч. Пирс)
2. Интенциональность (Ф. Брентано, Э.Гуссерль)
3. Субъективность (Р. Декарт)
4. Без пространственного протяжения (Р. Декарт)
5. Имеет внутреннюю природу (Р. Декарт, Э. Гуссерль)
6. Знакомо каждому, сознание прямого доступа (Н. Блок)
7. Безошибочность (Э. Гуссерль)
8. Простота
9. Невыразимость (проблема текучести чувств)
10. Осведомлённость
11. Аттендированность
12. Связность и интегрированность

Этот список основных свойств, с которыми также можно отдельно полемизировать, предложен А.В. Кузнецовым. Эти свойства должны быть объяснены, разобраны, позитивным образом предсказаны, или воспроизведены, той теорией, которая предлагает себя в качестве теоретического решения проблемы сознание-тело. Некоторые из этих свойств следуют из представления когнитивной архитектуры как самостоятельной замкнутой на саму себя системы (субъективность и внутренняя природа), другие свойства пересекаются, или являются двумя сторонами одной медали, например, квалитативность и безошибочность, кажется, повествуют о той стороне феноменальной данности, которая не производится в результате размышлений, а уже предлагается в качестве

готового решения автоматической бессознательной системы обработки сигнала субъекту в виде образа. Другие все ещё остаются загадкой. Способ данности субъекту опыта — это отдельная большая тема феноменологии. В представлении от первого лица наборы этих данностей предстают одновременно и удерживаются в кратковременной памяти. Природа информационного процесса в коннекционистской системе и логика обработки данных позволяют предложить удобное объяснение, так как параллелизм и распределённость которые обеспечивают сосуществование ментальных феноменов вместе присутствует и на уровне обработки информации мозгом.

2.6 Информационное пространство коннекционистской архитектуры

Понятие информации К. Шеннона может служить для решения количественных проблем эффективности связи при наличии шума в канале связи, в том числе связи между внешней средой и нервной системой. В таком виде теория К. Шеннона не проясняет семантический аспект информации, но сосредотачивается на технических аспектах передачи сигнала в коммуникационном канале при наличии шума. К. Шеннон предлагает измерять количество информации с помощью определения энтропии (как вероятности осуществления какого-либо макроскопического состояния). Д. Чалмерс трактует количество информации К. Шеннона как измерение специфики конкретного состояния в пространстве состояний. Самой простой системой, в которой есть вероятность осуществления какого-либо макроскопического состояния, или системой, имеющей «пространство состояний», может быть транзистор, стрелка на железнодорожном пути или бытовой выключатель с двумя стабильными состояниям. Он будет содержать один бит информации, иметь состояние 0 или 1.

Самые разные физические системы (в том числе компьютеры и мозг) можно специфицировать относительно «пространства состояний» этой

системы. Этот подход может быть очень удобен для определения особенностей репрезентации информации коннекционистских моделей. Д. Чалмерс берет спекулятивное понятие информации Г. Бейтсона, для которого информация связана с производством различий. Информация – различие, производящее различие (a difference that makes a difference). Понятие различия Г. Бейтсона – абстрактное понятие, которое охватывает как события нашей психической жизни, так и явления природы, которые наблюдается в опыте, специфицируя предметы и познавательные акты с точки зрения производства различий. Так, например, Г. Бейтсон предлагает модификацию «вещи в себе» И. Канта, как объективного элемента мира, который в перцептивном акте определяется в качестве ограниченной выборки различий, которые упорядочиваются внутри тела как информационные цепи. Таким образом теория различий с одной стороны выделяет агента, который усваивает информацию, с другой стороны в онтологическом плане «внешний мир» не является инородным, или же отдельным, или же «противопоставлением агенту».[171]

Теория Г. Бейтсона была в значительной степени конкретизирована Д. Чалмерсом для работы по созданию информационной теории применительно к любым физическим системам, в том числе и когнитивным агентам. Д. Чалмерс использует возможности, предлагаемые концепцией различия, для установления связи между физическими состояниями и информационными. Понятие информационного пространства Д. Чалмерса создаёт информационные состояния с помощью простого различения двух и более макроскопических состояний. Структура информационного пространства таких состояний может быть усложнена за счёт увеличения количества состояний (различение сразу трех, четырёх и более, до бесконечности

[171] «в целом остается верным что кодирование и передача различий вне тела очень сильно отличается от кодирования и передачи различий внутри тела. Об этом отличии нужно сказать, поскольку оно способно привести нас к ошибкам. Обычно мы думаем о внешнем "физическом мире" как о чем-то отдельном от внутреннего "ментального мира". Я полагаю, что это разделение основывается на контрасте в кодировании и передаче различий внутри и вне тела». Бейтсон Г. Экология разума. Избранные статьи по антропологии, психиатрии и эпистемологии / Пер. с англ. М.: Смысл. 2000. — 476 с.

(континуум состояний, например, все числа между 0 и 1). Континуальные информационные пространства состоят из множества состояний, связанных отношением близости, какие-то из них находятся близко, другие далеко друг от друга, то есть полностью соответствуют топологии континуума. Расширение мерности (2-n, 3-n и так до бесконечности) континуума таких информационных пространств сближает понятие информационного пространства с аналогичной структурой топологических пространств, где информационное состояние определяется как точка. Допускается наличие внутренней структуры у каждого отдельного состояния. Это расширение позволяет выделить внутри информационного пространства подпространство со своими элементами. Например, пространство со структурой из четырёхбитных состояний содержит такие состояния, как 1111, 1110, 1101 и т.д., где каждое состояние образовано четырьмя элементами. Такое пространство можно представить, как пространство, сложенное из четырёх подпространств с двумя состояниями. Эту операцию удаётся произвести в виду того, что в двоичной системе увеличение разряда соответствует умножению на 2. Если информационное пространство из двух различных состояний хранит 1 бит, то N таких пространств хранит N бит, так как полное число состояний 2^N. Если N равно бесконечности, получается континуум состояний, каждое из которых может включаться в подпространство с двумя состояниями, за счёт смешения первых двух типов усложнения. Если увеличить число элементов в подпространстве до бесконечности, то число состояний также станет бесконечным. Такое усложнение помогает перейти от дискретных пространств к континуальным. Такие пространства очень похожи на топологические многомерные пространства весовых коэффициентов глубокой нейросети. Каждому информационному состоянию в такой сети можно найти уникальный набор весовых коэффициентов. Д. Чалмерс расширяет понятие информационного пространства до такой степени, что оно совпадает с динамическим пространством весовых коэффициентов глубокой ИНС. И в этом смысле

любое феноменальное различие, которое можно пронумеровать методом Д. Чалмерса и которое может зафиксировать глаз, ухо, язык и др., может быть сопоставлено с точкой на таком пространстве весовых коэффициентов.

Возможно представить себе, что последовательная реализация когнитивной функции может выполнять процедуру поиска соответствия между Y и X без необходимости размещения отдельных команд параллельно в одном информационном пространстве. Тогда алгоритм такой последовательной процедуры поиска соответствия между Y и X выполняет сличение поочерёдно, в виде линейной последовательности команд и игнорирует их естественные отношения, связанные с временной и пространственной реализацией на носителе (мозге), которые естественно организуют внутренние причинные отношения Y и X. Программа нахождения функции отличная от тех что используются в мозге имеет иные внутренние состояния. Значит такая программа имеет состояния сознания отличные от тех которые знакомы нам из нашего опыта.

Возможно также представить, что параллельная реализация когнитивной функции может выполнять процедуру поиска соответствия между Y и X одномоментно, без какого-либо времени поиска. Тогда одношаговый алгоритм такой параллельной процедуры поиска соответствия между Y и X также игнорирует их естественные отношения, связанные с временной и пространственной реализацией на носителе (мозге), которые естественно организуют внутренние причинные отношения Y и X. И также имеет иные состояния сознания отличные от тех, что известны нам из нашего опыта.

Так как для мозга характерна параллельная и распределённая обработка информации то внутренние состояния системы в случае человека организуются соответственно. Для ментальных феноменов имеется параллельное и распределённое сосуществование в сознании. Значит параллельное и распределённое представление когнитивной функции

является необходимым условием для сличения ментальных и физиологических явлений применительно к человеку.

Коннекционизм предсказывает соответствие типов ментальных феноменов с типами информационных систем мозга. И в случае обработки перцептивной информации в мозге, и в случае представления различных квалитативных характеристик в феноменальном пространстве (которое нумеруется согласно соответствующему ему информационному пространству и имеет тип функциональных связей, согласный с выводами из первой главы диссертации) наблюдается параллельное и распределённое представление информации, которое очевидно в силу наличия разных типов сенсорных систем и соответствующих им типов модальностей. Например, сенсорная система зрительного тракта соответствует зрительному восприятию, то же самое и со слухом. И в сознании, и в мозге информация об услышанном и увиденном представлена одновременно и в разных локусах (параллельно и распределённо). Современное развитие науки показывает, как параллельно и распределённо обрабатываются различные типы информации мозгом и эти типы соотносятся с тем, как в сознании представляются различные характеристики опыта (тоже параллельно и распределённо). **Поэтому коннекционизм – необходимое условие для сличения ментальных и физиологических характеристик любой теории, которая предлагает способ сличения мозгового механизма с некоторым состоянием сознания.**

На это важнейшее свойство, объединяющее ментальные явления и процессы обработки информации в мозге, в частности, обращает внимание нейронаучная теория сознания К. В. Анохина[172]. В этой теории сетевое взаимодействие становится общим знаменателем как для видов опыта,

---

[172] Когнитом: нейронаучная теория сознания К.В. Анохина [Электронный ресурс]. URL: https://cmi.to/%d0%ba%d0%be%d0%b3%d0%bd%d0%b8%d1%82%d0%be%d0%bc-%d0%bd%d0%b5%d0%b9%d1%80%d0%be%d0%bd%d0%b0%d1%83%d1%87%d0%bd%d0%b0%d1%8f-%d1%82%d0%b5%d0%be%d1%80%d0%b8%d1%8f-%d1%81%d0%be%d0%b7%d0%bd%d0%b0%d0%bd/ (дата обращения: 08.10.2021).

представленных ментально, так и для физиологических процессов в нейронных сетях мозга, соответствующих данному виду опыта.

Это же самое необходимое условие скрыто присутствует в теории интегрированной информации Д. Тонони. В этой теории количество информации, содержащейся в сознании при восприятии объекта, становится ключевым для наличия самого феномена сознания. Само необходимое условие включено в постулат о составе сознательного состояния, в котором всегда находятся несколько характеристик и в постулат об единстве, который предполагает не последовательное, а одновременное представление в сознании различных типов опыта.

Если сетевые принципы для мозга являются общепринятыми, то относительно сетевых принципов устройства разума (о параллельности и распределённости ментальных феноменов) первым заявил коннекционизм. В данном параграфе показывается, что в виду сетевого устройства мозга, для теорий, которые претендуют на отождествление мозговых процессов с ментальными феноменами, следствием будет необходимость коннекционизма. А также важное следствие из этого параграфа — специфика феноменологии сознания в силу его параллельного и распределённого воплощения в мозге. Предполагается, что информационные состояния, вычисляемые иным образом (например, последовательно), будут иметь иную феноменологическую структуру и таким образом не будут обладать «человеческими» квалиа.

Выводы:

Двухсистемная теория А. Ребера, выделяющая эксплицитную систему усваивания закономерностей вербальным образом, может быть ориентиром для ограничения событий, привлекающих процесс аналитического дедуктивного вывода. Обязательным компонентом ментальной манипуляции символами и образами является сознание и от его свойств зависит процесс обработки информации. Именно возможность свободного манипулирования

различными объектами, символами и синтаксическими конструкциями плохо описывается коннекционистскими моделями. Рассматривая коннекционистскую архитектуру для процессов, в которых возможно обрабатывать последовательности, показано, как нейросети могут эмулировать любого рода программы классического вида и как нейросети применяют для работы с последовательными данными, такими как письмо и речь. Последовательная речь и письмо человека могут быть симулированы нейронной сетью, но эта симуляция не предполагает осмысленной работы, и, поэтому, не может отразить когнитивный процесс манипуляции символами. Есть основания полагать, что ментальные манипуляции (как свободный мыслительный процесс) в должной мере не описываются ни в классическом, ни в коннекционистском подходе, хотя изучаются когнитивной психологией[173]. Скорее всего мысленные образы нельзя рассматривать как один процесс, они складываются из множества различных функций и интегрируют задачи моторных, зрительных и ассоциативных зон коры мозга[174]. Возможно, вычислительные модели не обладают достаточной мощностью, но, что более вероятно, пока не раскрыты механизмы ментальной манипуляции. Ответ на вопрос о том, «когда я подсчитываю в уме мелочь, какой когнитивный процесс ответственен за такое вычисление?», состоит в том, что подсчитывание мелочи в уме — это не один когнитивный процесс. Стоит отметить, что на современном этапе применения коннекционистского подхода к задаче логико-дедуктивного вывода, трудным местом останется моделирование ментальных феноменов.

В разработках коннекционистов мало места уделено теме сознательной обработки сигнала. Однако сознание является невычитаемым элементом функционирования психических (мыслительных) процессов. Важную роль

---

[173] Иваницкий А. М. Мозговая основа субъективных переживаний: гипотеза информационного синтеза // Журнал высшей нервной деятельности. 1996. Т.46. № 2

[174] Иваницкий А.М. Информационный синтез в ключевых отделах коры как основа субъективных переживаний // Журнал высшей нервной деятельности. 1997. Т. 47. № 2. С. 209–225.

имеет рассмотрение возможности интерпретации сознательных феноменов с точки зрения коннекционистской методологической программы.

Для исследования сознания с точки зрения коннекционистской архитектуры имеет смысл заострить внимание на эволюции динамики весовых коэффициентов нейронной сети. Информационным событием становится выборочное действие в конце цепочки каузальных связей всех элементов сети, *ответа*[175] сети на выходном нейроне. Показано, что многомерные пространства весовых коэффициентов ИНС могут быть применимы для теории информационных пространств Д. Чалмерса. Однако, такого функционального соотнесения любого рода феноменального различия с входными данными мало для «оживления» мысленного процесса. Возможно, сам Д. Чалмерс, как панпсихист, указал бы на наличие в ИНС феноменов сознательной деятельности. Также и Д. И. Дубровский мог бы согласиться с тем, что нейродинамика весовых коэффициентов содержит информацию, которая воплощена в нейронной сети, она не зависит от субстрата, может содержаться на разных носителях, а также приватна, доступна только обученной нейронной сети. Но этих свойств недостаточно для приписывания ИНС феноменов субъективной реальности если не принимать во внимание функции, которые реализуются сознательными состояниями.

[175] Ф. Розенблатт понимает термин ответ как любое различимое состояние организма, которое может включать или не включать внешне определяемую мышечную активность.

Заключение

Становление коннекционистской программы когнитивных исследований происходило в рамках вычислительной теории разума. Коннекционизм интерпретирует понятие когнитивной функции как постепенную настройку весовых коэффициентов внутри группы связанных искусственных нейронов, во время которой сама функция возникает непосредственно из связанных групп нейронов. Для построения когнитивной функции в нейронных сетях операции в виде элементарных задач, как в архитектуре фон Неймана, не выделяются, промежуточные шаги не группируются. Искусственные нейронные сети не претендуют на морфологическое подобие и изофункционализм на физическом уровне, но в попытке репродуцировать сетевые взаимодействия биологических нейронов они пытаются отразить алгоритмы, управляющие процессам обработки информации в нервной системе, формируя правила настройки коэффициентов в сети, подобные присутствующим в мозге. Обучение биологических нейронных сетей связывают с «правилом Хебба», которое свидетельствует о высоких адаптивных способностях нейронов. Именно относительно воспроизводства этого правила в искусственных сетях нейронов происходят жаркие споры в среде разработчиков, которые предлагают биологически правдоподобные алгоритмы настройки коэффициентов для своих моделей работающей нервной системы.

Адаптивные возможности биологических нейронов генетически определены и достаточно устойчивы во времени. Сегодня существует большое количество работ по формализации таких адаптаций в виде простых вычислительных процессов в рамках вычислительной нейронауки. В данной работе исследуются элементарные вычислительные операции, такие, как поиск производной по времени и нормализация, для обоснования связи этих адаптивных возможностей нейронов с настройкой коэффициентов в

искусственных нейронных сетях. Однако в отличие от вычислительных моделей нейронауки, коннекционистские модели имеют ряд методологических отличий, которые обусловлены тем, что такие модели в качестве функциональных эквивалентов выбирают сети биологических нейронов низкого уровня специфичности и не исследуют отдельные молекулярные взаимодействия в каждом нейроне, что приводит к невозможности избирательного моделирования конкретных адаптаций нейронов. Поэтому с точки зрения некоторых биологов такие модели выглядят крайне упрощёнными, это отмечает Ф. Крик. В то же самое время, такой выбор функциональных примитивов даёт принципиально новый взгляд на работу нервной ткани, в основе которого лежат алгоритмы настройки сети. Нейронные сети не воспроизводят и не моделируют особенности строения нейрона, не создают компьютерную модель нервной системы организма, не являются математической моделью мозга, поэтому не могут быть достаточными для объяснения функционирования конкретной нервной системы. Но нейросети воспроизводят некие фундаментальные принципы, необходимые для осуществления процедуры вычисления когнитивных функций, отражая особенности функционирования сетей нервных клеток, и показывают как биологические нейросети могут статистически обрабатывать большие потоки данных. Такой взгляд подкрепляется моделями искусственных нейронных сетей, которые оказываются применимыми для обработки информации, состоящей из квази-натуралистических сенсорных сигналов, таких как пиксели изображений или слуховые спектрограммы. Одно из достаточных условий для работы нервной системы – это система имплицитного статистического вывода, которая позволяет нейронной сети предсказывать паттерны в новых данных исходя из предыдущего опыта. Эта особенность функционирования нейронных сетей может объяснить формирование наших аподиктических знаний, привычек, предвосхищений, а также помочь понять устройство естественного интеллекта. Исходя из этого предположения коннекционистские модели для когнитивной науки

предлагают объяснения функционирования языка, образного мышления и последовательного дедуктивного вывода с опорой на внешние системы знаков, средствами реализации которых будут зависимости или отношения между элементарными сигналами, статистически вероятно появляющиеся вместе. Для таких моделей «знание» удобно описывать в терминах информационных событий или ответов в терминологии Ф. Розенблатта, или различий в терминологии Г. Бейтсона, или информационных состояний в терминах Д. Чалмерса, или информации в терминах Д. И. Дубровского. Общим сходством этих описательных моделей будет идея выбора между несколькими вариантами (в самом простом случае между двумя вариантами), запись о таком выборе будет закодирована во всей сети и будет являться эмерджентным свойством всей сети целиком. Описания такой информации достаточно для автоматических процессов обучения нервной сети, но может быть недостаточно для реализации высших когнитивных функций, таких как понимание и сознание, так как в нейронной сети можно закодировать такое знание без программы или алгоритма субъективного взгляда «изнутри».

В работе была высказана мысль о функциональном значении «субъективного взгляда» как особенного свойства самостоятельных агентов, имеющих тело. Таким агентам необходимы внутренние модели своего тела для планирования действий, но обученной нейронной сети этого не требуется. Большинство современных нейронных сетей «бестелесны» и время, которое необходимо для их обучения, не ограничено индивидуальной жизнью или средой обитания, их функциональных свойств недостаточно для описания высших психических функций. Эта их особенность позволяет сказать, организму приходится существовать в рамках условий среды, которая определяет время обучения и время реакции. В целом временные масштабы, необходимые для реализации некоторых когнитивных функций, могут многое нам рассказать об особенности поведения организмов. Именно биологическая определённость (конкретная анатомия, предзаданная

эволюцией и физическими законами) нашего когнитивного аппарата не может быть правильным образом дедуцирована из концепта абстрактных моделей нейронных сетей. Поэтому настройка весовых коэффициентов в сети нейронов является одним из достаточных условий для работы ЕНС, но не достаточна для объяснения реализации наших когнитивных способностей, что подкрепляется тем фактом, что современные модели вычислительной нейронауки отказываются от модели коннекционистского типа в пользу построения копий «реальной нервной системы».

Таким образом, искусственные нейронные сети являются грубой моделью для любой нервной системы. Они обладают способностью распознавать объекты, формируют память и обучаются, но не способны воспроизвести множество деталей и признаков, характерных для нервной системы. Несмотря на биологически неправдоподобные способы обучения, эти системы могут решать задачи принципиально схожие с теми, что решает мозг. Такие системы предлагают в новом свете подойти к пониманию работы любой нервной системы, как системы, накапливающей информацию о среде и статистически обрабатывающей эту информацию для организации поведения. Эти системы также прилагают объяснения для такой обработки информации.

Список литературы

1. Алексеев А. Ю. Машина Корсакова (1832 г.) как прототип мультиагентного суперкомпьютерного автомата // Искусственные общества. 2019. Т. 14. Выпуск 1.

2. Алексеев А.Ю. Философия искусственного интеллекта: концептуальный статус комплексного теста Тьюринга. ФГБОУ ВО «Московский государственный университет имени М.В. Ломоносова», 2016.

3. Алексеев А.Ю. Протонейрокомпьютер Корсакова // Нейрокомпьютер: разработка, применение. №7. 2013. С. 6-17.

4. Анохин К. В. Коды Вавилонской библиотеки мозга /; подгот. Валерий Чумаков // В мире науки. 2013. № 5. С. 82-89.

5. Арутюнян В. Г. Структура ментальных репрезентаций: извлечение текста из памяти, нейронная сеть и искусственный интеллект // Вестник Пермского университета. Российская и зарубежная филология. 2013. №4. 240 с.

6. Барышников П. Н. Методологические возможности и границы вычислительных моделей сознания. Москва, 2018.

7. Бейтсон Г. Экология разума. Избранные статьи по антропологии, психиатрии и эпистемологии / Пер. с англ. М.: Смысл. 2000. 476 с.

8. Винер Н. Кибернетика, или Управление и связь в животном и машине. М.: Советское радио. 1958.

9. Герович В.А. Динамика исследовательских программ в области искусственного интеллекта. Москва. 1991.

10. Гибсон У. Нейромант: Фантаст.роман / Пер. с англ. Е. Летова, М. Пчелинцева. — М.: Аст; СПб.: Terra Fantastica, 2000. 317с.

11. Гоббс Т. Сочинения в 2 т. Т.1. М. Мысль, 1989.- 622с.- (Филос.насл. Т.107.)- С.518-519.

12. Декарт Р. Сочинения в 2 т.-Т. 1. М.: Мысль, 1989. 654 с. (Филос. наследие; Т. 106). С.423-460.

13. Джеймс У. Научные основы психологии. СПб.: Санкт-Петербургская электропечатня. 1902. 192 с.

14. Дрейфус Х. Чего не могут вычислительные машины? М: Прогресс, 1978. 336 с.

15. Дубровский Д. И. Проблема «Сознание и мозг»: Теоретическое решение. М.: «Канон+» РООИ «Реабилитация», 2015. 208 с.

16. Дубровский Д.И. Критический анализ теории сознания Пенроуза–Хамероффа. Часть 2 // Философия науки и техники 2017. Т. 22. № 2. С. 89–102

17. Дубровский Д.И. Проблема «Сознание и мозг»: Теоретическое решение. М.: «Канон+» РООИ «Реабилитация», 2015. 208 с.

18. Иваницкий А. М. Информационный синтез в ключевых отделах коры как основа субъективных переживаний // Журн. высшей нервной деятельности. 1997. Т. 47. № 2. С. 209–225.

19. Иваницкий А. М. Мозговая основа субъективных переживаний: гипотеза информационного синтеза // Журнал высшей нервной деятельности. 1996. Т.46. № 2.

20. Иванов Д.В. Радикальный энактивизм и проблема субъективности // Вопросы философии. 2016. № 11.

21. Истинные имена. True Names. Повесть, 1981 год. Язык написания: английский. Перевод на русский: А. Новоселов (Истинные имена), 2015. 2 изд.

22. Клюева Н. Ю. Влияние идеи г. Лейбница на развитие компьютерных наук и исследования в области искусственного интеллекта // Вестник Московского университета. Серия 7. Философия. 2017. №4. с. 79-92.

23. Клюева Н. Ю. Отношение теоретических концепций и компьютерных моделей в исследованиях искусственного интеллекта.

ФГБОУ ВО «Московский государственный университет имени М.В. Ломоносова». 2008.

24. Когнитом: нейронаучная теория сознания К.В. Анохина [Электронный ресурс]. URL: https://cmi.to/%d0%ba%d0%be%d0%b3%d0%bd%d0%b8%d1%82%d0%be%d0%bc-%d0%bd%d0%b5%d0%b9%d1%80%d0%be%d0%bd%d0%b0%d1%83%d1%87%d0%bd%d0%b0%d1%8f-%d1%82%d0%b5%d0%be%d1%80%d0%b8%d1%8f-%d1%81%d0%be%d0%b7%d0%bd%d0%b0%d0%bd/ (дата обращения: 08.01.2021).

25. Костелянец Н. Б., Левкович Ю. И. Зрительное распознавание при предваряющей или запаздывающей настройке наблюдателя на определенный набор изображений // Журнал высшей нервной деятельности. 1982. Т. 32. № 2. 292.

26. Кузнецов В. «Аристотелевская теория категорий и прототипический подход» Вестник Московского университета. Серия 7. Философия, no. 1, 2018, pp. 32-44.

27. Лекторский В. А. Реализм, анти-реализм, конструктивизм и конструктивный реализм в современной эпистемологии и науке » ИНТЕЛРОС [Электронный ресурс]. URL: http://www.intelros.ru/intelros/reiting/reyting_09/material_sofiy/6141-realizm-anti-realizm-konstruktivizm-i-konstruktivnyj-realizm-v-sovremennoj-yepistemologii-i-nauke.html (дата обращения: 10.06.2022).

28. Лейбниц Г. В. Порядок есть в природе // Он же. Соч.: в 4 т. М., 1982. Т. 1. С. 234.

29. Мак-Каллок У., Питтс В. Логическое исчисление идей, относящихся к нервной активности // Нейронные сети: история развития теории / Под общей ред. Галушкина А.И., Цыпкина Я.З. М.: ИПРЖР, 2001. С. 5–22.

30. Михайлов И.Ф. Человек. Сознание. Сети. М.: ИФ РАН, 2015. 196 с.

31. Михайлов И. Методологический выбор между субстанциализмом и функционализмом. Человек вчера и сегодня: междисциплинарные исследования. Вып. 6. М.: ИФ РАН, 2012.

32. Овчинникова И. Г. О коннекционистской интерпретации речевой деятельности // Вопросы психолингвистики. 2006. №4. с. 37-47.

33. Панина Е.М. Когнитивная наука как комплекс междисциплинарных исследований. Москва, 2001.

34. Пенроуз Р. Новый ум короля. М.: Едиториал УРСС, 2003. 339 с.

35. Петренко А. К., Петренко О. Л. Машина Беббиджа и возникновение программирования // Историко-математические исследования. 1979. Т. 24. С. 340.

36. Савельев А.В. Научная школа «Психофизиология и нейрокомпьютинг сенсорных систем» ИМПБ РАН / В.Я. Сергин, Е.В. Лосева, А.В. Савельев / Выпуск под ред. В.Я. Сергина, Е.В. Лосевой, А.В. Савельева // Нейрокомпьютеры: разработка, применение. 2015. №11. С. 5-7.

37. Савельев А. В. Философско-методологические основания нейрокомпьютинга. Москва, 2016.

38. Сварник О. Активность мозга: специализация нейрона и дифференциация опыта / Российская акад. наук, Ин-т психологии. М: Институт психологии РАН, 2016.188 с.

39. Сонин А.Г. Моделирование механизмов понимания поликодовых текстов. Москва. 2006.

40. Стасенко С. В., Гордлеева С. Ю., Семьянов А. В., Дитятев А. Э., Казанцев В. Б. Модель астроцитарной координации тормозного и возбуждающего входов интернейрона // Вестник ННГУ. 2014.

41. Сущин М.А. Концепция ситуативного познания в когнитивной науке: критический анализ. Москва, 2014.

42. Тьюринг А. Могут ли машины мыслить? // Информационное общество / Сост. А. Лактионова. М.: ООО «Издательство АСТ», 2004. С. 221–284.

43. Черниговская Т. В. Язык, мозг и компьютерная метафора //Человек. – 2007. – Т. 2. – С. 63-75.

44. Черниговская, Т. В. «Language acquisition device» : Где оно? [Текст]/ Т. В. Черниговская // Детская речь как предмет лингвистического исследования. Матер. Междунар. науч. конф. (Санкт-Петербург, 31 мая-2 июня 2004 г.). - СПб.: Наука, 2004. - С. 280-281.

45. Шарков Ф. И. Общение в Сети и зарождение сетевой киберкультуры. 2013. с. 98-100.

46. Швырков В.Б. Системная детерминация активности нейронов в поведении // Успехи физиологических наук. 1983. Т.14. № 1.

47. Шредингер Э. Анатомия разума: об интеллекте, религии, и будущем [перевод с немецкого] М: Родина, 2020. 208 с.

48. Эшби У. Р. Конструкция мозга. Происхождение адаптивного поведения М.: ИЛ, 1962. 397 с.

49. Ягодинский И. И. Философия Лейбница. Процесс образования системы. СПб., 2007. С. 200.

50. Amrein I. Isler K. LiP H.P. Comparing adult hippocampal neurogenesis in mammalian species and orders: influence of chronological age and life history stage. Europe Journal Neuroscience. 2011. Vol. 34. P. 978-987.

51. Angel I., Dolores del Castillo M., Ignacio J., Serrano Jesus O. Connectionist Models of Decision Making // Chiang J. S. (ed.) Decision SuPort Systems. IntechOpen. 2010.

52. Angel I., Dolores del Castillo M., Ignacio J., Serrano O. Connectionist Models of Decision Making // Chiang J. S. (ed.) Decision SuPort Systems. IntechOpen. 2010.

53. Balcázar J. «Computational Power of Neural Networks: A Kolmogorov Complexity Characterization». IEEE Transactions on Information Theory. 1997. Vol. 43. № 4. P. 1175–1183.

54. Balcázar J. Computational Power of Neural Networks: A Kolmogorov Complexity Characterization // IEEE Transactions on Information Theory. 1997. Vol. 43. №. 4. P. 1175–1183.

55. Bargh J., Chartrand T. The unbearable automaticity of being// American Psychologist. 1999. Vol. 54. No. 7. P. 462–479.

56. Bargh J., Chartrand T. The unbearable automaticity of being// American Psychologist. 1999. Vol. 54. №. 7. P. 462–479.

57. Bechtel W. Natural deduction in connectionist systems // Synthese. 1994. Vol. 101. P. 433–463.

58. Becker J. B., Meisel R. L. Neurochemistry and Molecular Neurobiology of Reward // Handbook of Neurochemistry and Molecular Neurobiology. Springer. 2007. P. 739-774

59. Bengio Y. The Consciousness Prior // arXiv:1709.08568 [cs, stat]. 2019.

60. Bengio Y., Ducharme R., Vincent P. A Neural Probabilistic Language Model // Advances in Neural Information Processing Systems 13. MIT Press. 2001. P. 932–938.

61. Bengtsson S. et al. Extensive piano practicing has regionally specific effects on white matter development // Nature Neuroscience. 2005. Vol.8. P. 1148–1150.

62. Benhamou S., Bovet P. How animals use their environment: a new look at kinesis. Animal Behavior. 1989. Vol. 38. P. 375–383.

63. Beniaguev D., Segev I., London M. Single cortical neurons as deep artificial neural networks // Neuron. 2021. № 17. P. 2727-2739.

64. Biologically Inspired Cognitive Architectures (BICA) for Young Scientists: Proceedings of the First International Early Research Career Enhancement School (FIERCES 2016) // Advances in Intelligent Systems and Computing. 2016. Vol. 449.

65. Bowers J. Grandmother cells and localist representations: a review of current thinking. Lang // Cognitive Neuroscience. 2017. Vol.32. P. 257–273.

66. Bowers J. Grandmother cells and localist representations: a review of current thinking. Lang. Cognitive Neuroscience. 2017. Vol. 32. P. 257–273.

67. Bowers J. Parallel Distributed Processing Theory in the Age of Deep Networks // Trends in Cognitive Sciences. 2017. Vol. 21. №.12. P. 950–961.

68. Cadieu C., Kouh M., Pasupathy A., Conner C., Riesenhuber M., Poggio, T.A. A Model of V4 Shape Selectivity and Invariance // J Neurophysiology. 2007. Vol. 98. P. 1733-1750.

69. Carandini M., Heeger, D. J. Normalization as a canonical neural computation // Nature Reviews Neuroscience, 2011. Vol.13. №. 1. P. 51–62.

70. Casale A., McCormick D. Active action potential propagation but not initiation in thalamic interneuron dendrites // Journal of Neuroscience. 2011. Vol. 31. №.18. P. 289-302.

71. Catastrophic Interference in Connectionist Networks: The Sequential Learning Problem - ScienceDirect [Electronic resource]. URL: https://www.sciencedirect.com/science/article/pii/S0079742108605368 (accessed: 21.09.2020).

72. Chalasani S., Chronis N., Tsunozaki M., Gray J., Ramot D., Goodman M., Bargmann C. Dissecting a circuit for olfactory behavior in Caenorhabditis elegans // Nature. 2007. Vol. 450. P. 63–70. Cho C.E., Brueggemann C., L'Etoile N.D., Bargmann C.I. Parallel encoding of sensory history and behavioral preference during Caenorhabditis elegans olfactory learning. Elife. 2016.

73. Chomsky N. Three models for the description of language, in IRE Transactions on Information Theory. Vol. 2. №. 3, P. 113-124.

74. Church A. An Unsolvable Problem of Elementary Number Theory // American Journal of Mathematics. 1936. Vol. 58. №. 58. P. 345—363.

75. Churchland M. A Neurocomputational Perspective: The Nature of Mind and the Structure of Science. Cambridge. MA: MIT Press. 1989.

76. Clark A. Whatever next? Predictive brains, situated agents, and the future of cognitive science // Behavior Brain Science. 2013. Vol. 36. № 3. P. 181–204.

77. Consciousness: A Philosophic Study of Minds and Machines. Random House, 1969. 273 pages. Hard cover: Peter Smith. 1972.

78. Crick F. Astonishing Hypothesis: The Scientific Search for the Soul / F. Crick, Simon and Schuster, 1995. P. 340.

79. Crick F. The Astonishing Hypothesis. Touchstone. 1995.

80. Crick F. The recent excitement about neural networks // Nature. 1989. Vol 337. P. 129–132

81. Damasio A. R., Grabowski T. J., Bechara A., Damasio H., Ponto L.L.B., Parvizi J., Hichwa R.D. Subcortical and cortical brain activity during the feeling of self-generated emotions" 2000. Nature Neuroscience №3. P. 1049–1056.

82. de Vries, S.E.J., Lecoq, J., Buice, M. et al. A large-scale standardized physiological survey reveals functional organization of the mouse visual cortex // Nature Neuroscience. 2020. Vol. 23. P. 138–151.

83. de Vries, S.E.J., Lecoq, J.A., Buice, M.A. et al. A large-scale standardized physiological survey reveals functional organization of the mouse visual cortex. Nature Neuroscience. 2020. Vol. 23. P. 138–151

84. Dechter R. Learning while searching in constraint-satisfaction problems. University of California, Computer Science Department, Cognitive Systems Laboratory. 1986.

85. Deepak Path Pulkit Agrawal, Alexei A. Efros and Trevor Darrell. Curiosity-driven Exploration by Self-supervised Prediction. NTSL. 2017.

86. Elman J. Finding structure in time // Cognitive Science. 1990. Vol. 14. P. 179–212.

87. Fodor J. The Mind Doesn’t Work This Way; The Scope and Limits of Computational Psychology, MIT Press. 2000.

88. Fodor, Jerry A. The Language of Thought, Cambridge. Massachusetts. Harvard University Press, 1975.

89. Fredkin E. An Introduction to Digital Philosophy // International Journal of Theoretical Physics. 2003. Vol. 42 №. 2. P. 189–247.

90. Fried I., Rutishauser U., Cerf and M. Kreiman G. Single Neuron Studies of the Human Brain: Probing Cognition. The MIT Press: Cambridge Massachusetts and London, England. 2014.

91. Ganis G., Thompson W., Kosslyn M. Brain areas underlying visual mental imagery and visual perception: an fMRI study // Cognitive Brain Research. 2004. Vol. 20. №. 2. P. 226-241.

92. Gauthier I., Tarr M. J. Visual object recognition: Do we (finally) know more now than we did? Annual review of vision science. 2016. Vol. 2. P. 377-396.

93. George A. Miller The Magical Number Seven, Plus or Minus Two // The Psychological Revie. 1956. Vol. 63. P. 81—97.

94. Gibson J. The Perception of the Visual World. Boston: Houghton Mifflin. 1950.

95. Giese M., Poggio T. Neural mechanisms for the recognition of biological movements and action // Nature Review Neuroscience 2003. Vol. 4. P. 179–192.

96. Glöckner A., Betsch T. Modeling option and strategy choices with connectionist networks: Towards an integrative model of automatic and deliberate decision making // Judgment and Decision Making. 2008. Vol. 3. №. 3. P. 215–228.

97. Grossberg S. Competitive learning: From interactive activation to adaptive resonance. Cognitive Science. 1987. Vol. 11. P. 23-63.

98. Grush R. The emulation theory of representation: motor control, imagery, and perception // Behavior Brain Science. 2004. Vol. 27. №. 3. P. 377-442.

99. Güntürkün, O., & Bugnyar, T. Cognition without Cortex. Trends in Cognitive Sciences. 2016. № 20(4). P. 291–303.

100. Haber S.N., The place of dopamine in the cortico-basal ganglia circuit. Neuroscience. 2014. № 282. P. 248-257.

101. Halford G., Wilson W., Phillips S. Processing capacity defined by relational complexity: implications for comparative, developmental, and cognitive psychology. Behavior Brain Science.1998. Vol. 2. P. 803-31.

102. Haraway D. Manifesto for Cyborgs: Science, Technology, and Socialist Feminism in the 1980s // Socialist Review. 1985. Vol. 80. P. 65-108.

103. Hassabis, D., Kumaran, D., Summerfield, C., & Botvinick, M. Neuroscience-inspired artificial intelligence. Neuron. 2017. Vol. 95. № 2. P. 245-258.

104. Hebb, D. O. The organization of behavior: A neuropsychological theory. New York: John Wiley. P. 62. 1964.

105. Hilton A. M. The Social Implications of Mechanization, Automation and Cybernation in Agriculture // Front Cover, 1967.

106. Houk J. C., Adams C. M., Barto, A. G. A model of how the basal ganglia generate and use neural signals that predict reinforcement. in Models of Information Processing in the Basal Ganglia (eds. Houk, J.C., Davis, D.G.) MIT Press. Cambridge. MA. USA. 1995. P.249–270.

107. Huang Y., Rao R. Predictive coding // Wiley Interdiscipte Review Cognitive Science. 2011. Vol. 2. P. 580–593.

108. Hubel D., Wiese T. Brain mechanisms of vision // Scientific American. 1979. Vol. 241. P. 150—162.

109. Iino Y, Yoshida K. Parallel use of two behavioral mechanisms for chemotaxis in Caenorhabditis elegans. Neuroscience. 2009. Vol. 29. № 17. P. 5370-5380.

110. Indiveri, G., Linares-Barranco, B. et al. Neuromorphic Silicon Neuron Circuits // Frontiers in Neuroscience. Vol 5.

111. Jastorff J., Kourtzi Z., Giese M.A. Learning to discriminate complex movements: biological versus artificial trajectories // Journal of Vision. 2006. Vol. 6, № 8. P. 791–804.

112. John Von Neumann The computer and the brain. Yale University Press. First edition 1958.

113. Johnson J.G., Busemeyer J.R. A dynamic, stochastic, computational model of preference reversal phenomena // Psychological Review. 2005. Vol. 112. №. 4. P. 841-861.

114. Kamin L. Selective association and conditioning. In Fundamental Issues in Associative Learning (Mackintosh, N.J. and Honig, F.W.K., eds).1969. P. 42–64

115. Kandel E. R. A Cell - Biological Approach to Learning, New York: Society for Neuroscience. 1978.

116. Kandel E. R. The biology of memory: a forty-year perspective. Neuroscience. Vol. 2. №. 41. P. 12748–12756.

117. Kell, A. J. E., Yamins, D. L. K., Shook, E. N., Norman-Haignere, S. V. & McDermott, J. H. A task-optimized neural network replicates human auditory behavior, predicts brain responses, and reveals a cortical processing hierarchy // Neuron. 2018. Vol. 98. P. 630–644.

118. Kellogg R.T. Fundamentals of cognitive psychology, 2nd edn. SAGE, Thousand Oaks. 2012

119. Kesner R., Rolls E. A computational theory of hippocampal function, and tests of the theory: new developments. Neuroscience Biobehavioral Review. 2015. Vol. 48. P. 92-147.

120. Khaligh-Razavi, S., Kriegeskorte N. Deep supervised, but not unsupervised, models may explain IT cortical representation. PLoS Computation Biology. 2014. Vol. 10.

121. Kinouchi Y., Mackin K. J. A Basic Architecture of an Autonomous Adaptive System With Conscious-Like Function for a Humanoid Robot // Frontiers in Robotics and AI. 2018. № 5. P. 30.

122. Kollias P., McClelland J. L. Context, cortex, and associations: a connectionist developmental aProach to verbal analogies // Front. Psychol. 2013. Vol. 4. P. 857.

123. Koch, C., Massimini, M., Boly, M. et al. Neural correlates of consciousness: progress and problems. Nature Review Neuroscience. 2016. № 3, P. 307–321.

124. Larsch J., Flavell S., Liu Q., Gordus A., Albrecht D., Bargmann C. A Circuit for Gradient Climbing in C. elegans Chemotaxis. Cell Reports. 2015. Vol. 12 №11. P. 1748-1760.

125. LeCun Y., Boser B., Denker J. et al. Backpropagation Applied to Handwritten Zip Code Recognition // Neural Computation. 1984. Vol. 1. №. 4. P. 541-551.

126. Lillicrap, T. P., Santoro, A., Marris, L., Akerman, C. J., Hinton, G. Backpropagation and the brain. Nature reviews // Neuroscience. 2020. Vol. 216. № 6. P. 335–346.

127. London M., Roth A., Beeren L., Häusser M., Latham P.E. Sensitivity to perturbations in vivo implies high noise and suggests rate coding in cortex // Nature. Vol. 466. P. 123–127

128. Lucas J. Machines and Gödel. Philosophy. 1961. Vol. 36 (XXXVI). P. 112–127.

129. Marcus G. (a). Deep learning: A critical appraisal. 2018. ArXiv Preprint ArXiv:1801.00631.

130. Marcus G. (b). Innateness, AlphaZero, and artificial intelligence. 2018. ArXiv Preprint ArXiv:1801.05667.

131. Marr D., Poggio T. From Understanding Computation to Understanding Neural Circuitry. Neurosciences Research Program Bulletin. 1979. Vol. 15. №.3. P. 470-488.

132. Masquelier T., Thorpe S. J. Unsupervised learning of visual features through spike timing dependent plasticity // PLoS Computation Biology. 2007.

133. Maxwell M. Psycho-Cybernetics. Simon and Schuster. 1960.

134. McClelland J. L., Elman J. L. The TRACE model of speech perception // Cognitive Psychology. 1986. Vol. 18. P. 1–86.

135. McClelland J. L., Rumelhart D. E. An interactive activation model of context effects in letter perception: I. An account of basic findings // Psychology Review. 1981. Vol. 88. P. 375–407.

136. McClelland J. L., Rumelhart D. E., Hinton, G. E. The appeal of parallel distributed processing. In A. M. Collins & E. E. Smith (Eds.), Readings in cognitive science: A perspective from psychology and artificial intelligence 1988. P. 10-11.

137. McClelland, J. L., & Rogers, T. T. The Parallel Distributed Processing Approach to Semantic Cognition. Nature Reviews Neuroscience. 2003. Vol. 4. №. 4. P. 310–322.

138. Mermillod M., Bugaiska A., Bonin P. The stability-plasticity dilemma: investigating the continuum from catastrophic forgetting to age-limited learning effects. Front. Psychology. 2013. Vol. 4. P. 504.

139. Minsky M., Papert S. M.I.T. Press. Cambridge. Mass. 1969.

140. Newell A., Simon H. Computer Science as Empirical Inquiry: Symbols and Search // Communications of the Associations for Computing Machinery. 1975. Vol. 19. № 3. P. 113–126.

141. Nick A. Bostrom and Anders Sandberg, «Whole Brain Emulation: A Roadmap». 2008. P.130.

142. Niv Y., Duff M.O., Dayan P. Dopamine, uncertainty and TD learning. Behavioral Brain Function. 2005. Vol. 1. P. 6.

143. O'Reilly R.C. Biologically Plausible Error-Driven Learning Using Local Activation Differences: The Generalized Recirculation Algorithm // Neural Computation. 1996. Vol. 8. P. 895–938.

144. Oaksford M., Chater N. Dual processes, probabilities, and cognitive architecture. Mind Soc. 2012. Vol. 1. P. 15–26

145. Olds J., Milner P. "Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain". Journal of Comparative and Physiological Psychology. 1954. Vol. 47 (6). P. 419–427.

146. Pakkenberg B., Pelvig D., Marner L., Bundgaard M., Gundersen H., et al. Aging and the human neocortex. Expirience Gerontology. 2003 Vol. 38. №. 1-2. P. 95.

147. Pashler H. Dual-task interference in simple tasks: Data and theory // Psychological Bulletin. 1994. Vol. 116. №. 2. P. 220–244.

148. Perrinet L.U. An Adaptive Homeostatic Algorithm for the Unsupervised Learning of Visual Features // Vision (Basel). 2019. Vol. 3, № 3. P. 47.

149. Pierce-Shimomura J.T., Morse T.M., Lockery S.R. The fundamental role of pirouettes in Caenorhabditis elegans chemotaxis. Neuroscience. 1999. Vol. 19. P. 955-997.

150. Place U. T. Is Consciousness a Brain Process? // British Journal of Psychology. 1956. № 1 (47). С. 44–50.

151. Priebe N., Ferster D. Inhibition, spike threshold, and stimulus selectivity in primary visual cortex. Neuron. 2008. Vol. 57. №4. P. 482-497.

152. Rao R., Ballard D. Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. Nature Neuroscience, 1999. Vol. 2 №. 1. P. 79–87.

153. Ratcliff R. Connectionist models of recognition memory: constraints imposed by learning and forgetting functions // Psychological review. Psychol Rev, 1990. Vol. 97, № 2. P. 285-308.

154. Reicher G. Perceptual recognition as a function of meaningfulness of stimulus material. Journal Experience Psychology. 1965. Vol. 81. P. 275.

155. Robert E. Ornstein The Evolution of Consciousness: Of Darwin, Freud, and Cranial Fire: The Origins of the Way We Think. 1991. P. 320.

156. Rolls E.T., Tromans J., Stringer S. Spatial scene representations formed by self-organizing learning in a hiPocampal extension of the ventral visual system. Eur. J. Neuroscience. 2008. Vol. 28. P. 2116–2127.

157. Rosenblatt F. The perceptron: A probabilistic model for information storage and organization in the brain Cornell Aeronautical Laboratory Psychological Review 1958. Vol. 65. №. 6.

158. Rosenblatt F. The perceptron: A probabilistic model for information storage and organization in the brain Cornell Aeronautical Laboratory Psychological Review 1958. Vol. 65. №. 6.

159. Rossato J. I. et al. Silent Learning // Current Biology. 2018. № 21. P. 3508-3515.

160. Rumelhart D. E. Parallel distributed processing: explorations in the microstructure of cognition / David E. Rumelhart, James L. McClelland, and the PDP Research Group. / D. E. Rumelhart, Cambridge, Mass: MIT Press, 1986.

161. Rumelhart D.E., Hinton G.E., Williams R.J., Learning Internal Representations by Error Propagation. In: Parallel Distributed Processing Cambridge, MA, MIT Press. 1986. Vol. 1. P. 318—362.

162. Sandberg A. and Bostrom N. Whole Brain Emulation: A Roadmap. Future of Humanity Institute, Oxford University, Technical Report #2008-3. 2008.

163. Savinov N. et all. Episodic Curiosity through Reachability // arXiv:1810.02274 [cs, stat]. 2019.

164. Schaffner J. et al. Neural codes in early sensory areas maximize fitness //bioRxiv. – 2021.

165. Schultz W. Dopamine reward prediction error coding. Dialogues Clinical Neuroscience. 2016. Vol. 18. №1. P. 23-32.

166. Schultz W. Dopamine reward prediction error coding. Dialogues Clin Neuroscience. 2016. Vol. 18(1). P. 23-32.

167. Searle J. Is the Brain's Mind a Computer Program?, Scientific American T. Vol. 262. №. 1. P. 26–31.

168. Seidenberg M. Sublexical structures in visual word recognition: Access units or orthographic redundancy? In M. Coltheart (Ed.), Attention and performance XII: The psychology of reading 1987. P. 245 – 263.

169. Serre T. Models of visual categorization. Wiley Interdisciplinary Reviews: Cognitive Science. 2016. Vol. 7. №. 3. P. 197–213.

170. Serre, T., Oliva, A., Poggio, T. A feedforward architecture accounts for rapid categorization // Proceedings of the National Academy of Sciences. 2007. Vol. 104. №. 15. P. 6424–6429.

171. Shepard, R.N. Cooper L. Mental Images and their Transformations. MIT Press. 1982.

172. Silver D. et all. Reward is enough // Artificial Intelligence. 2021. (299). C. 103535.

173. Silver D., Schrittwieser, J., Simonyan, K. et al. Mastering the game of Go without human knowledge. Nature. 2017. Vol. 550. P. 354–359.

174. Simon D., Snow C.J., Read S. The redux of cognitive consistency theories: evidence judgments by constraint satisfaction // Journal of Personality and Social Psychology. 2004. Vol. 86. P. 814–837.

175. Single Neuron Studies of the Human Brain: Probing Cognition Eds. I. Fried, U. Rutishauser, M. Cerf and G. Kreiman, 2014 The MIT Press: Cambridge Massachusetts and London, England.

176. Sompolinsky H., Crisanti H., Sommers Chaos in Random Neural Networks // Physical Review Letters. 1988. Vol. 61. P. 259.

177. Sun, R. The CLARION cognitive architecture: Extending cognitive modeling to social simulation. In: Ron Sun (Ed.), Cognition and Multi-Agent Interaction. Cambridge University Press: New York. 2004.

178. Suri R. E., Schultz W. A neural network model with dopamine-like reinforcement signal that learns a spatial delayed response task // Neuroscience. 1999. Vol. 91. № 3. P. 871–890.

179. Sutton J. Philosophy and Memory Traces: Descartes to Connectionism. Cambridge University Press, 1998. 372 p.

180. The Logic Theory Machine -- A Complex Information Processing System. Allen Newell & Herbert A. Simon - 1956 - IRE Transactions on Information Theory. Vol. 2. № 3. P. 1-79.

181. The Social Implications of Mechanization, Automation, and Cybernation in Agriculture. Front Cover. Alice Mary Hilton. 1967.

182. Thiele T. R., Faumont S., Lockery S. R. The Neural Network for Chemotaxis to Tastants in Caenorhabditis elegans Is Specialized for Temporal Differentiation // Journal Neuroscience. 2009. Vol. 29. № 38. P. 11904–11911.

183. Timothy T.R. Neural networks as a critical level of description for cognitive neuroscience // Current Opinion in Behavioral Sciences. 2020. Vol. 32. P. 167-173.

184. Timoty C. May. [Электронный ресурс]. URL. https://nakamotoinstitute.org/static/docs/cyphernomicon.txt дата обращения 05.09.2020

185. Von der Malsburg C. Am I thinking assemblies? In Brain Theory. Springer. 1986. P. 161–176.

186. Walls J. The Philosophy of David Hartley and the Root Metaphor of Mechanism: A Study in the History of Psychology, Journal of Mind and Behavior. 1982. Vol. 3. P. 259–74.

187. Wang, J. X., Kurth-Nelson, Z., Kumaran, D., Tirumala, D., Soyer, H., Leibo, J. Z., … Botvinick, M. Prefrontal cortex as a meta-reinforcement learning system. Nature Neuroscience. 2018. Vol. 21(6). P. 860–868.

188. Watanabe, E., Kitaoka, A., Sakamoto, K., Yasugi, M. Tanaka, K. Illusory motion reproduced by deep neural networks trained for prediction // Frontal Psychology. 2018. Vol. 9. P. 345.

189. White J. G., Southgate E., Thomson J. N., Brenner S. The structure of the nervous system of the nematode Caenorhabditis elegans. Philosophical Transactions of the Royal Society A. 1986. Vol. 314. P. 1-340.

190. Wiechert M., Judkewitz B., Riecke H., Friedrich R. Mechanisms of pattern decorrelation by recurrent neuronal circuits. Nature Neuroscience. 2010. Vol. 13. №8. P. 1003-1010.

191. Wiener N. God and Golem, Inc: A Comment on Certain Points where Cybernetics Impinges on Religion. The M.I.T. paperback series. M.I.T. Press, 1966.

192. Yamins D. L., DiCarlo J. J. Using goal-driven deep learning models to understand sensory cortex. Nature neuroscience. 2016. Vol. 19. №3. P. 356.

193. Zador, A.M. A critique of pure learning and what artificial neural networks can learn from animal brains. Nature Communication. 2019. Vol. 10. P. 3770

194. Zorzi M, Testolin A, Stoianov IP. Modeling language and cognition with deep unsupervised learning: a tutorial overview. Frontires of Psychology. 2013. №. 4. P. 515.