

Хьюберт Л. ДРЕЙФУС и Стюарт И. ДРЕЙФУС

СОЗДАНИЕ СОЗНАНИЯ *vs.* МОДЕЛИРОВАНИЕ МОЗГА: ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ ВЕРНУЛСЯ НА ТОЧКУ ВЕТВЛЕНИЯ ¹

[Н]ичто не кажется мне более возможным, чем то, что люди однажды придут к бесповоротному мнению, что в ... нервной системе нет копии, которая соответствовала бы той или иной *конкретной* мысли, или той или иной *конкретной* идее, или воспоминанию.

Людвиг Витгенштейн [1948: i. 504 (66e)]

[И]нформация не хранится ни в каком определенном месте — она хранится везде. Лучше представлять себе дело так, что информацию «истребуют», а не «находят».

Дэвид Румелхарт и Дональд Норман (1981: 3)

В начале 1950-х годов, когда вычислительные машины начинали получать признание, немногие мыслители-новаторы начали осознавать, что цифровые компьютеры могут быть чем-то большим, чем мельницами, перемалывающими числа. Тогда возникли и вступили в борьбу за признание два противоположных видения того, чем могли бы стать компьютеры, — каждое со своей исследовательской программой. Одна партия видела в компьютерах систему манипулирования ментальными символами; другая — средство моделирования мозга. Одна стремилась использовать компьютеры для инстанцирования формальной репрезентации мира, другая — для моделирования взаимодействия нейронов. Одна рассматривала в качестве парадигмы интеллекта решение задач, другая — обучение. Одна использовала логику, другая — статистику. Одна школа была наследницей рационалистической, редукционистской традиции в философии, другая — рассматривала себя как новую идеализированную, холистскую нейронауку.

¹ Dreyfus H. L., Dreyfus S. E. Making a Mind Versus Modelling The Brain: Artificial Intelligence Back at a Branch-Point // The Philosophy of Artificial Intelligence / Boden M. (ed.) Oxford, 1990. Перевод выполнен А. Л. Блиновым. Впервые данная статья опубликована в журнале: «Artificial Intelligence», 117, № 1 (Winter 1988), Camb., Mass. © Daedalus, Journal of the American Academy of Arts and Sciences. — Прим. ред.

Боевой клич первой группы был: и сознания, и компьютеры суть физико-символьные системы. К 1955 году Аллен Ньюэлл и Герберт Саймон, работавшие в РЭНД Корпорейшн, пришли к выводу, что последовательности битов, которыми манипулирует цифровой компьютер, могут стоять вместо чего угодно — вместо чисел, конечно, но также вместо тех или иных черт действительного мира. Более того: программы можно использовать в качестве правил репрезентации отношений между этими символами, так что система могла бы выводить дальнейшие факты о репрезентируемых объектах и отношениях между ними. Как недавно написал Ньюэлл в своем рассказе об истории проблем AI,

Область цифровых компьютеров определяет компьютеры как машины, манипулирующие числами. «Вся грандиозность идеи компьютеров, — говорят их приверженцы, — состоит в том, что числами можно закодировать все что угодно, даже инструкции». Ученые же в области AI рассматривали компьютеры как машины, манипулирующие символами. «Вся грандиозность идеи компьютеров, — говорили они, — в том, что символами можно закодировать все что угодно, даже числа» (*Newell 1983: 196*).

Этот взгляд на компьютеры стал основанием для соответствующего взгляда на сознание. Ньюэлл и Саймон выдвинули гипотезу, что человеческий мозг и цифровой компьютер, будучи абсолютно различными по структуре и механизму, имеют на определенном уровне абстракции одно и то же функциональное описание. На этом уровне и мозг человека, и подходящим образом запрограммированный цифровой компьютер можно рассматривать как две различных инстанции устройства одного и того же рода — устройства, порождающего разумное поведение посредством манипулирования символами с помощью формальных правил. Ньюэлл и Саймон сформулировали свой взгляд в виде гипотезы:

Гипотеза о Системе Физических Символов. Система физических символов обладает необходимыми и достаточными средствами для порождения разумных действий.

Под «необходимыми» мы имеем в виду, что любая система, выказывающая общую разумность, окажется — по рассмотрении ее — системой физических символов. Под «достаточными» мы имеем в виду, что любую систему физических символов, имеющую достаточно большие размеры, можно далее организовать так, чтобы она выказывала общую разумность (*Newell and Simon 1981: 41*).

Ньюэлл и Саймон возводят истоки своей гипотезы к Готлобу Фреге, Берtrandу Расселу и Альфреду Норту Уайхеду (1981: 42), но, конечно, Фреге и компания сами были наследниками давней атомистической, рационалистической традиции. Уже Декарт выдвинул предположение, что: (1) понимание заключается исключительно в построении подходящих репрезентаций и манипулировании ими; (2) эти репрезентации можно разложить на простые элементы (*naturas simplices*); и (3) все эти феномены можно понимать как сложные комбинации простых символов. И у Гоббса, современника Декарта, имелось неявное допущение, согласно которому элементы суть формальные компоненты, связанные друг с другом чисто синтаксическими операциями, так что рассуждения можно свести к вычислениям. «Когда человек *рассуждает*, он делает не что иное, как выводит целое из сложения частей, — пишет Гоббс, — ибо РАЗУМ ... есть не что иное, как расчет...» (1958: 45). Наконец, Лейбниц, разрабатывая классическую идею матезиса — формализации всего, имел в виду обеспечить тем самым основания для построения универсальной системы символов, так чтобы «мы могли каждому объекту приписать соответствующее ему определенное число» (1951:18). Согласно Лейбницу, в процессе понимания мы разлагаем понятия на более простые элементы. Чтобы избежать регресса ко все более и более простым элементам, должны иметься простейшие, в терминах которых можно понять все более сложные понятия. Более того, если мы хотим, чтобы понятия были приложимы к миру, то должны иметься некие простые признаки предметов, репрезентируемые этими элементами. Лейбниц имел в виду построить «алфавит человеческих мыслей» (1951: 20), «буквы [которого] должны демонстрировать, будучи употреблены в доказательствах, связь, группировку и порядок, имеющие место также и в объектах» (1951: 10).

Людвиг Витгенштейн, опираясь на Фреге и Рассела, сформулировал в своем «Логико-философском трактате» в чистом виде этот синтаксический, репрезентационистский взгляд на отношение сознания к действительности. Он определил мир как совокупность логически независимых атомарных фактов:

1.1. Мир есть совокупность фактов, а не вещей.

Факты же, считал он, можно исчерпывающим образом разложить на простые объекты.

2.01. Атомарный факт есть объединение объектов...

2.0124. Если даны все объекты, то *тем самым* даны также и все атомарные факты.

Эти факты, их составные части и логические отношения между ними, утверждал Витгенштейн, репрезентируются в сознании.

2.1. Мы создаем для себя образы фактов.

2.15. То, что элементы образа соединяются друг с другом определенным способом, показывает, что так же соединяются друг с другом и вещи (1960).

AI можно трактовать как попытку отыскать в субъекте (в человеке или в компьютере) простые элементы и логические отношения, отражающие те простые объекты и отношения между ними, из которых сделан мир. В сущности, гипотеза Ньюэлла и Саймона о системе физических символов превращает витгенштейновское видение (само являющееся кульминацией философской традиции классического рационализма) в некую эмпирическую заявку и основывает на ней исследовательскую программу.

Противопоставляемая этому интуиция, согласно которой нам следует создавать искусственный интеллект, моделируя мозг, а не символичные репрезентации мира в сознании, черпала свои побудительные мотивы не из философии, а из того, что вскоре назвали нейронаукой. Непосредственным источником ее мотиваций были работы Д. О. Хебба, предположившего в 1949 году, что некоторая масса нейронов способна обучаться в том случае, если при одновременном возбуждении нейрона *A* и нейрона *B* это возбуждение увеличивало силу связи между ними.

Эта идея была подхвачена Франком Розенблаттом, рассудившим так: раз разумное поведение, основывающееся на нашей репрезентации мира скорее всего трудно формализуемо, то AI должен вместо этого пытаться автоматизировать те процедуры, с помощью которых сеть нейронов научается различать образы (*patterns*) и соответственно реагировать на них. Как писал Розенблатт,

неявное допущение [исследовательской программы манипулирования символами] состоит в том, что относительно легко специфицировать поведение, которое мы хотели бы получить от машины, а самое трудное — построить такое устройство или механизм, который бы эффективно реализовал это поведение... [И] легче и полезнее аксиоматизировать *физическую систему*, а затем аналитически исследовать эту систему, чтобы детерминировать ее поведение, чем аксиоматизировать *поведение*, а затем строить физическую систему с помощью техники логического синтеза (1962b: 386).

По-другому различие между этими двумя исследовательскими программами можно сформулировать так: те, кто ориентировались на

символьные репрезентации, искали такую формальную структуру, которая дала бы компьютеру возможность решать определенный класс проблем или различать определенные образы. Розенблатт же хотел построить такое физическое устройство (или смоделировать такое устройство на компьютере), которое затем смогло бы генерировать свои собственные способности.

Многие обсуждавшиеся модели связаны с вопросом о том, какую логическую структуру должна иметь система, если мы хотим, чтобы она обладала некоторым свойством *X*. По существу, это вопрос о статической системе...

Альтернативный подход к этому вопросу таков: какая система могла бы *развить* свойство *X*? Я думаю, что для ряда интересных случаев можно показать, что второй вопрос можно решить, не имея ответа на первый (1962b: 387).

Оба подхода вскоре привели к поразительным успехам. К 1956 году Ньюэлл и Саймон преуспели в программировании компьютера с помощью символьных репрезентаций так, чтобы он решал простые головоломки и доказывал теоремы пропозиционального исчисления. Эти первые впечатляющие результаты наводили на мысль, что гипотеза о системах физических символов близка к подтверждению, и нетрудно понять, что Ньюэлл и Саймон были в эйфорическом настроении. Саймон объявил:

Не собираюсь удивлять или шокировать вас... Но проще всего подвести итоги было бы так: теперь в мире существуют машины, которые мыслят, обучаются и творят. Больше того, все эти их способности будут быстро расти до той поры, когда — в обозримом будущем — диапазон проблем, с которыми они способны справляться, совпадет с тем диапазоном проблем, за которые вообще брался когда-либо ум человека (1958: 6).

Он и Ньюэлл объясняли:

[У] нас теперь имеются элементы теории эвристического (а не алгоритмического) решения задач; и мы можем использовать эту теорию как для понимания эвристических процессов у человека, так и для моделирования таких процессов с помощью цифровых компьютеров. Интуиция, инсайты и обучение не являются более только собственностью людей: любой большой высокоскоростной компьютер мы можем запрограммировать так, чтобы он выказывал эти свойства. (1958: 6)²

² Эвристические правила — это такие правила, о которых, когда они используются людьми, говорят, что они основаны на опыте или суждении. Часто такие правила приводят к правдоподобным решениям проблем

Розенблатт воплотил свои идеи в устройство, названное им перцептроном³. К 1956 году Розенблатт сумел научить перцептрон распознавать подобие паттернов определенных видов и отделять их от других, неподобных, паттернов. К 1959 году он также ликовал и чувствовал, что его подход оправдал себя:

Представляется ясным, что... с перцептроном на свет явились новые автоматы по обработке информации: впервые у нас имеется машина, спо-

или увеличивают эффективность процедуры решения задачи. Тогда как алгоритмы гарантируют правильное решение (если таковое вообще имеется) за некоторое конечное время, эвристики лишь увеличивают вероятность нахождения правдоподобного решения.

³ Румелхарт и Маклелланд (1986) так описывают перцептрон: «Подобные машины состоят из так называемой *сетчатки* — набора бинарных входов, упорядоченно размещенных, как иногда трактуется, в двумерном пространстве; множества *предикатов* — множества бинарных пороговых элементов, у которых фиксированы связи с некоторым подмножеством элементов в сетчатке, так что каждый предикат вычисляет некую локальную функцию над тем подмножеством элементов, с которым он соединен; и с одним или несколькими децизиональными элементами, связи которых с предикатами можно изменять» (i. III). Они противопоставляют способ хранения информации в модели с параллельной распределенной обработкой (ПРО-модели) — например, в перцептроне — способу ее хранения с помощью символической репрезентации: «В большинстве моделей знание хранится в качестве статической копии некоего паттерна. Поиск информации заключается в нахождении данного паттерна в долгосрочной памяти и копировании его в буфер или в оперативную память. Нет серьезных различий между репрезентацией, хранимой в долгосрочной памяти, и репрезентацией, активно используемой в оперативной памяти. В ПРО-моделях же дело обстоит не так. В этих моделях сами паттерны не хранятся. А хранятся *силы связей* между элементами, позволяющие воссоздать эти паттерны (i. 31)... [З]нание об индивидуальных паттернах не хранится в связях некоторого конкретного паттерна, специально сопоставленного с данным паттерном, а распределено по множеству связей между многими обрабатываемыми элементами» (i. 33). Это новое понятие репрезентации непосредственно привело к Розенблаттовой идее о том, что такие машины должны обретать свои способности по ходу обучения, а не в результате того, что их запрограммировали некоторыми характеристиками и правилами: [е]сли знание заключается [в] силах связей, то обучение должно состоять в нахождении правильных сил связей, так что при подходящих обстоятельствах будут продуцироваться нужные паттерны активации. Это чрезвычайно важное свойство данного класса моделей, ибо оно открывает для механизмов, обрабатывающих информацию, возможность научиться — подстраивая подходящим образом свои связи — ухватывать взаимозависимости между активациями, которым оно подвергается в ходе обработки» (i. 32).

собная обладать оригинальными идеями. В качестве аналога биологического мозга, перцептрон, или, точнее, теория статистической отделенности, по-видимому, приближается к тому, чтобы удовлетворить требования функционального объяснения нервной системы больше, чем любая другая из ранее предлагавшихся систем ... В концептуальном плане, по-видимому, в перцептроне воплощена, без всякого сомнения, реальная осуществимость, так и принципы отличных от человека систем, способных осуществлять функции человеческого познания ... По-видимому, будущее информационных устройств, работающих не на логических, а на статистических принципах, обозначилось достаточно ясно (1958: i.449)

В начале шестидесятых оба подхода выглядели в равной степени обещающими, и оба в равной степени подставлялись критике, выдвигая завышенные претензии. Тем не менее, результаты внутренней войны между этими двумя исследовательскими программами оказались удивительно несимметричными. К 1970 году подход моделирования мозга, с перцептроном в качестве парадигмы, был сведен к немногим разрозненным, малодатируемым усилиям, в то время как люди, предложившие использовать цифровые компьютеры в качестве символьных манипуляторов, обрели беспспорный контроль над ресурсами, аспирантскими программами, журналами и симпозиумами, что в совокупности и составляет процветающую исследовательскую программу.

Реконструкция того, как произошла такая перемена, осложняется мифом о неотвратимости того, что произошло, — мифом, порождаемым любой активно действующей исследовательской программой. Итак, победителям представляется, что символьная обработка информации победила потому, что была на правильном пути, а нейронные сети или коннекционистский подход проиграл потому, что он попросту не работал. Однако такая концепция истории данной области исследований — не более чем ретроспективная иллюзия. У обеих исследовательских программ имелись как достойные разработки идеи, так и глубокие, остававшиеся в тени проблемы.

У каждой из позиций были свои хулители, и говорили они, в сущности, одно и то же: данный подход показал свою способность решать некоторые легкие проблемы, но нет оснований полагать, что данная исследовательская группа сможет успешно распространить свои методы на сложные объекты. И в самом деле, имелись свидетельства, что по мере возрастания сложности проблем, вычисления, требуемые обоими подходами, будут расти по экспоненте и, стало быть, скоро станут нереалистическими. В 1969 году Марвин Минский и Сеймур Пейперт писали о Розенблаттовом перцептроне:

Розенблаттовы схемы быстро принялись, и вскорости появилось, похоже, не меньше сотни групп и группок, экспериментирующих с этой моделью...

Результаты этих сотен проектов и экспериментов были в общем и целом разочаровывающими, а объяснения — половинчатыми. Обычно машины прекрасно работали с очень простыми проблемами, но быстро ухудшали свои показатели по мере роста трудности ставившихся перед ними задач.

Тремя годами позже сэр Джеймс Лайтхилл, отрецензировав работы по использованию эвристических программ, например работы Саймона и Минского, пришел к поразительно сходным негативным выводам:

Большинство людей, работающих в AI и соседних областях, признаются в четко выраженном чувстве разочарования в том, что достигнуто за последние 25 лет. Люди пришли в эту область около 1950-го и даже около 1960-года с большими надеждами, которые в 1972-м году очень далеки от осуществления. Ни в одном из разделов этой области до сих пор сделанные открытия не произвели того воздействия, которое было в свое время обещано...

[О]дна достаточно общая причина испытанных разочарований: признание всех последствий «комбинаторного взрыва». Это общее препятствие на пути построения ... системы, основанной на крупной базе знаний, источником которого является взрывной рост — при увеличении размеров базы знаний — любого комбинаторного выражения, репрезентирующего число возможных способов группировки элементов базы знаний в соответствии с теми или иными конкретными правилами.

Как лапидарно подытожили это Дэвид Румелхарт и Дэвид Зипсер, «рано или поздно вас настигает комбинаторный взрыв, хотя иногда скорее в различных вариантах параллельной, а не последовательной обработки информации» (*Rumelhart and McClelland 1986: i.158*). Обе стороны, как однажды выразился Джерри Фодор, ввязались в игру в трехмерные шахматы, воображая, что они играют в крестики-нолики. Отчего же на столь ранней стадии игры, когда еще столь мало познано и столь много еще только предстоит узнать, одна исследовательская команда празднует победу за счет полного поражения другой стороны? Почему на этой решающей развилке проект символической репрезентации стал единственной игрой в нашем городе?

Каждый знающий историю этой области сможет указать на ближайшую причину. Около 1965-го года Минский и Пейперт, руководившие лабораторией в Массачусетском Институте Технологии, работавшей над проблематикой символично-манипуляционного подхода и поэтому соревновавшейся с перцептронными проектами, начали рас-

пространять первые варианты книги, нападавшей на идею перцептрона. В этой книге они ясно заявили о своей научной позиции:

Перцептроны были широко разрекламированы как машины, «распознающие образы (patterns)», или обучающиеся машины, и в качестве таковых они обсуждались в многочисленных книгах, журнальных статьях и объемистых «отчетах». Большая часть этих писаний ... лишена научной ценности (1969: 4).

Но их нападки были также и философской кампанией. Они правильно понимали, что традиционному использованию редукции к логическим исходным элементам брошен вызов со стороны некоего нового холизма:

Оба автора, пишущих эти строки (сначала по отдельности, а потом вместе), вовлеклись в некую процедуру терапевтического принуждения с целью рассеять то, что мы с опасением считали первыми предвестниками «холистического» или «гештальтного» заблуждения, угрожавшего нашествием на поля технических исследований и исследований в области искусственного интеллекта, подобным более раннему нашествию на биологию и психологию (1969: 19).

Они были вполне правы. Искусственные нейронные сети могут — хотя не обязаны — допускать интерпретацию своих скрытых узлов ⁴ в терминах таких признаков, которые могли бы быть распознаны человеком и использованы им для решений данной проблемы. Если само моделирование нейронной сети не связывает себя ни с какой точкой зрения, можно показать, что ассоциация не *требует*, чтобы скрытые узлы были интерпретируемы. Такие холисты, как Розенблатт, выдвинули удачное допущение, что индивидуальные узлы или паттерны узлов не выбирают фиксированных признаков данной предметной области.

Минскому и Пейперту так сильно хотелось устранить всякое соперничество, и они были столь безмятежно убеждены в правоте атомистической традиции, идущей от Декарта к раннему Витгенштейну, что в их книге гораздо больше намеков, чем реальных доказательств. Они взялись за анализ способностей однослойного перцептрона ⁵ и в то же время в математическом разделе своей книги абсолютно проигнорировали главы Розенблатта о многослойных машинах и его доказательство о сходимости вероятностного обучающегося алгоритма, ос-

⁴ Скрытые узлы — это узлы, которые не детектируют непосредственным образом вход в сеть, а с другой стороны, и не образуют собой ее выхода. Но они прямо или косвенно связаны — посредством соединений, обладающих регулируемыми силами, — и с узлами, которые детектируют вход, и с узлами которые образуют собой выход.

⁵ У однослойной сети нет скрытых узлов, а у многослойной есть.

нованного на обратном размножении ⁶ ошибок (1962а: 292) ⁷. Как пишут Румелхарт и Маклелланд,

Минский и Пейперт поставили себе целью показать, какие функции могут, а какие не могут вычислять [однослойные] машины. Они, в частности, показали, что подобные перцептроны неспособны вычислять такие математические функции, как функцию четности (находится ли на сетчатке четное или нечетное число точек) или топологическую функцию связности (все ли точки на сетчатке связаны со всеми остальными точками на сетчатке либо напрямую, либо через какие-то другие точки, которые также находятся на сетчатке) без того, чтобы использовать до нелепости большое число предикатов. Их анализ чрезвычайно изящен; он демонстрирует важность математического подхода к анализу вычислительных систем (1986: i. 111).

Однако следствия этого анализа достаточно ограничены. Румелхарт и Маклелланд пишут далее:

В общем и целом, ... хотя Минский и Пейперт были абсолютно правы в своем анализе *однослойного перцептрона*, их теоремы неприменимы к таким системам, которые хотя бы немного сложнее. В частности, они неприменимы ни к многослойным системам, ни к системам, допускающим циклы обратной связи (1986: i. 112).

И несмотря на это, в заключении к *Перцептронам*, когда Минский и Пейперт задаются вопросом: «Рассматривали ли вы перцептроны со многими слоями?», они, хотя и риторически, оставляют вопрос открытым, создавая впечатление, что решили его:

Да, мы рассмотрели машины Гамба, которые можно описать как «два слоя перцептрона». Мы не обнаружили (размышляя над этим вопросом и просматривая имеющуюся литературу) никакого другого по-настоящему

⁶ Для обратного размножения ошибок требуется рекурсивное вычисление, начиная с тех узлов, что расположены на выходе, воздействует изменения сил связей на разность между желаемым выходом и тем выходом, который производится входом. Затем, по ходу обучения, веса подгоняются так, чтобы сократить эту разность.

⁷ См. также: «Добавление четвертого слоя сигналопереносящих блоков или поперечное сдвигание (cross-coupling) А-блоков трехслойного перцептрона позволяет решать проблемы обобщения над произвольными группами преобразований... В перцептронах с обратным сдвиганием (back-coupled perceptrons) может иметь место выборочное внимание к знакомым объектам в некоей сложной области. Для такого перцептрона возможно также выборочно следить за объектами, дифференциально движущимися относительно своего фона» (Rosenblatt 1962а: 576).

интересного класса многослойных машин, — во всяком случае, никакого такого класса, принципы которого были бы существенно связаны с принципами перцептрона ... [М]ы считаем, что важной исследовательской проблемой является задача прояснения (или отвержения) нашего интуитивного суждения, что расширение в направлении многослойности бесплодно (1969: 231–232).

Их атака на гештальтное мышление в AI преуспела так, как им и не снилось. Лишь очень немногие, чьи голоса остались неслышанными, — в том числе Стивен Гроссберг, Джеймс А. Андерсон и Теуво Кохонен — занялись этой «важной исследовательской проблемой». В сущности, почти каждый работавший в AI считал, что нейронные сети похоронены навсегда. Румелхарт и Маклелланд отмечают:

Анализа Минским и Пейпертом ограниченностей однослойного перцептрона, вкуче с некоторыми первыми успехами подхода в AI, базирующегося на символической обработке, оказалось достаточно для того, чтобы внушить большому числу работающих в этой области, что у перцептронных вычислительных устройств нет никакого будущего в искусственном интеллекте и когнитивной психологии (1986: i. 112).

Но почему этого оказалось достаточно? Оба подхода выдали некоторые многообещающие результаты и некоторое количество необоснованных обещаний⁸. Было слишком рано подводить итоги. И все же в книге Минского и Пейперта оказалось нечто такое, что задело струну, готовую к ответу. Дело выглядело таким образом, будто люди, работающие в AI, разделяли те квазирелигиозные философские предрассудки против холизма, которые лежали в основании нападок Минского и Пейперта. Силу этой традиции можно увидеть, например, в статье Ньюэлла и Саймона о физико-символьных системах. Статья начинается с научной гипотезы о том, что сознание и компьютер разумны в силу того, что они манипулируют дискретными символами, а заканчивается неким откровением:

«Исследование логики и компьютеров открыло нам, что разумность коренится в физико-символьных системах» (1981: 64).

Холизм не совладал с такими сильными философскими убеждениями. Розенблатт был дискредитирован вместе с сотнями менее ответственных исследовательских групп, которые были оснащены его работой. Его поток спонсирования иссяк, и ему было нелегко опубликовать свои работы. К 1970 году в том, что касается AI, нейронные

⁸ Оценку действительных успехов символично-репрезентационного подхода к 1978 году см. в кн.: *Dreyfus* (1979).

сети были «мертвы». В своей истории AI Ньюэлл говорит, что спор «символы против числа» «определенно не имеет места сейчас и в течение уже долгого времени» (1983: 10). Розенблатт даже не упоминается в историях AI, написанных Джоном Хогелэндом (1985) и Маргарет Боден (1977)⁹.

Однако сваливать полный провал коннекционистов на антихолистские предрассудки значило бы слишком упрощать дело. Философские допущения повлияли на интуицию и привели к неоправданно высокой оценке важности первых результатов символично-процессорного подхода и на некоем более глубоком уровне. В то время дело выглядело так, что перцептронщикам приходилось выполнять огромный объем математического анализа и вычислений для решения даже самых простых проблем распознавания образов, например, для отличия горизонтальных линий от вертикальных в различных частях поля восприятия, в то время как символично манипулятивный подход относительно легко решал трудные когнитивные проблемы, например доказательство логических теорем и решение комбинаторных задач. И что еще важнее, казалось, что при вычислительных возможностях того времени исследователи нейронных сетей могли заниматься лишь спекулятивной нейронаукой и психологией, в то

⁹ Работа по нейронным сетям маргинально продолжалась в психологии и в нейронауке. Джеймс А. Андерсон из Броунского Университета продолжил защищать нейронную модель в психологии, хотя ему приходилось жить за счет грантов других исследователей, а Стивен Гроссберг разработал изящную математическую реализацию элементарных когнитивных способностей. О позиции Андерсона см.: *Anderson (1978)*. Примеры работ Гроссберга, сделанных в эту мрачную годину, см. в его книге (1982). О ранней работе Кохонена рассказывается в книге «*Associative Memory — A System-Theoretical Approach*» (Berlin: Springer-Verlag, 1977). В Массачусетском Технологическом Институте Минский продолжал читать лекции по нейронным сетям и назначал своим студентам темы дипломных работ об исследовании логических свойств нейронных сетей. Но согласно Пейперту, Минский поступал так только потому, что сети обладают интересными математическими свойствами, а насчет свойств символьных систем нельзя доказать ничего интересного. Кроме того, многие исследователи AI исходили из допущения, что раз машины Тьюринга суть символьные манипуляторы, а Тьюринг доказал, что машины Тьюринга могут вычислять все что угодно, он доказал тем самым, что все умопостижимое может быть схвачено логикой. Если принять эту точку зрения, то для холистского (а в те дни — для статистического) подхода нужно искать оправданий, а для подхода символьного AI не нужно. Однако эта самоуверенность основывалась на смешении неинтерпретированных символов машины Тьюринга (нулей и единиц) с семантически интерпретированными символами искусственного интеллекта.

время как простые программы символьных репрезентационистов уже вскоре обещали приносить пользу. Эта оценка ситуации основывалась на допущении, что мышление и распознавание образов — это две разных области, и мышление важнее. Как мы увидим ниже в нашем обсуждении проблемы обыденного знания, так смотреть на вещи — значит не учитывать как преобладающую роль распознавания образов в общем наборе человеческих способностей, так и тот фон обыденного понимания, который является предпосылкой настоящего обыденного мышления людей. А вполне вероятно, что для того, чтобы учесть этот фон, понадобится распознавание образов.

Эта мысль возвращает нас к философской традиции. За символьной обработкой информации стоял не просто Декарт и его последователи, но вся западная философия. Согласно Хайдеггеру, традиционная философия с самого начала задана своим сосредоточением на фактах в мире, в то время как «игнорируется» сам мир как таковой (Heidegger 1962; §§14–21; Dreyfus 1988). Это означает, что философия с самого начала систематическим образом игнорировала или искажала обыденный контекст человеческой деятельности¹⁰. К тому же та ветвь философской традиции, которая идет от Сократа через Платона, Декарта, Лейбница и Канта к классическому AI, считает само собой разумеющимся, что понимание некоторой предметной области состоит в том, что у нас есть *теория* этой области. Теория формулирует отношения между объективными, *бесконтекстными* элементами (простыми, исходными признаками, атрибутами, факторами, дискретными данными, сигналами и т. д.) в терминах абстрактных принципов (покрывающих законов, правил, программ и т. д.).

Платон считал, что в таких теоретических областях, как математика и, быть может, этика, люди, которые мыслят, пользуются явно выраженными, бесконтекстными правилами теорий, которые они узнали в иной жизни, за пределами обыденного мира. Будучи познанными, такие теории функционируют в нашем мире, управляя сознанием человека, который мыслит, — неважно, осознает ли сам человек эти теории или нет. Концепция Платона была приложима не к обыденным навыкам и умениям, а только к таким областям, в которых имеется априорное знание. Однако успех теорий в естественных науках подкреплял ту идею, что в любой упорядоченной области должен иметься некий набор бесконтекстных элементов и некоторых абстрактных отношений между этими элементами, ответственных за упорядоченность этой области и за способность человека разумно дейст-

¹⁰ Согласно Хайдеггеру, ближе всех других философов к пониманию важности обыденной деятельности подошел Аристотель, но даже и он не избежал искажения феномена обыденного мира, неявно содержащегося в здравом смысле.

воват в ее пределах. Так, Лейбниц смело обобщил рационалистский подход на все формы разумной деятельности, даже и обыденной практики:

[Н]аиболее важные наблюдения и умения всех ремесел и профессий еще не записаны. Этот факт подтверждается тем опытом, который мы получаем, когда переходим от теории к практике и хотим нечто осуществить. Конечно, мы можем и записать эту практику, поскольку, в сущности, она есть лишь еще одна теория, более сложная и конкретная... [курсив наш — Дрейфусы] (1951: 48).

Подход, основывающийся на обработке символической информации, черпает свою уверенность именно из этого переноса на все сферы тех методов, которые были развиты философами и оказались успешными в естественных науках. Поскольку, в согласии с этой точкой зрения, любая область формализуема, в любой области AI осуществляется тем, что отыскиваются бесконтекстные элементы и принципы и на этом теоретическом анализе базируется некая формальная, символическая репрезентация. В этом духе Терри Виноград описывает свою работу в области AI в терминах, заимствованных у физики:

Нам нужно построить такой формализм, или «репрезентацию», посредством которой мы могли описывать ... знание. Нам нужно отыскать «атомы» и «частицы», из которых оно строится, и «силы», которые на него действуют (1976: 9).

Несомненно, теории о Вселенной часто строятся постепенно, шаг за шагом, сначала моделируются относительно простые и изолированные системы, а затем модель шаг за шагом усложняется и связывается в одно целое с моделями других предметных областей. Это возможно потому, что все явления происходят, как считается, от законоподобных отношений между «структурными примитивами», как называют их Пейперт и Минский. Поскольку никто не *аргументирует* в пользу атомистической редукции в AI, кажется, что те, кто работает в AI, просто неявно *допускают*, что абстрагирование элементов от их обыденного контекста, задающее философию и работающее в естествознании, должно работать и в AI. Вполне возможно, что это допущение объясняет, почему гипотеза о физико-символической системе так быстро превратилась в откровение и почему книга Пейперта и Минского с такой легкостью разгромила холизм перцептрона.

Один из нас — Хьюберт, — преподавая философию в Массачусетском Технологическом Институте в середине 60-х годов, скоро был втянут в споры о возможности искусственного интеллекта. Было очевидно, что такие исследователи, как Ньюэлл, Саймон и Минский бы-

ли наследниками этой философской традиции. Но с учетом выводов позднего Витгенштейна и раннего Хайдеггера, это не казалось хорошим предзнаменованием для редукционистской исследовательской программы. Оба эти мыслителя поставили под вопрос саму ту традицию, на которой основывалась обработка символической информации. Оба были холистами, оба были под впечатлением важности обыденных практик и оба считали, что у человека не может быть теории обыденного мира.

Ирония интеллектуальной истории заключалась в том, что сокрушительная критика Витгенштейном его собственного «Трактата», — его «Философские исследования», — были опубликованы в 1953 году, как раз в то время, когда AI воспринял ту абстрактную, атомистическую традицию, на которую нападал Витгенштейн. Написав свой «Тракtrat», Витгенштейн провел годы, занимаясь тем, что он называл феноменологией (1975), в тщетных поисках атомарных фактов и базовых объектов, требовавшихся его теорией. В конце концов он отбросил свой «Тракtrat» и всю рационалистическую философию. Он аргументировал, что анализ обыденных ситуаций в терминах фактов и правил (а именно с этого, как считают большинство традиционных философов и исследователей AI, должна начинаться теория) сам имеет смысл лишь в том или ином контексте и для той или иной цели. Так, выбор элементов уже отражает те цели и задачи, для которых они отбирались. Когда мы пытаемся отыскать окончательные бесконтекстные, независимые от конкретной цели элементы, — а мы должны это сделать, если мы хотим отыскать исходные символы для ввода в компьютер, — то на самом деле мы пытаемся освободить аспекты нашего опыта как раз от той прагматической организации, которая позволяет нам разумно их использовать при столкновении с обыденными проблемами.

В «Философских исследованиях» Витгенштейн прямо критиковал логический атомизм «Трактата»:

«На чем основана идея, что имена на самом деле обозначают простые [элементы]?» — говорит Сократ в «Теэтете»: «Если я не ошибаюсь, я слышал, некоторые люди говорят так: нет определения тех первичных элементов — так сказать — из которых состоим мы и все остальное... Но в точности подобно тому, как то, что состоит из этих первичных элементов, само сложно, так и имена этих элементов становятся описательным языком, если их сложить вместе». И Расселовы «индивиды» и мои «объекты» («Логико-философский трактат») были такими первичными элементами. Но что представляют собой те простые составляющие, из которых составлена действительность?.. Совершенно бессмысленно говорить абсолютным образом о «простых частях стула» (1953: 21).

Уже в 1920-х годах Мартин Хайдеггер аналогичным образом реагировал против своего учителя Эдмунда Гуссерля, который считал свое учение кульминацией картезианской традиции и который был поэтому дедушкой AI (*Dreyfus* 1982). Гуссерль аргументировал, что акт сознания, или ноэзис, сам по себе не схватывает объект; такой акт имеет интенциональность (направленность) лишь в силу некоей «абстрактной формы», или смысла (*meaning*), в ноэме, коррелированной с данным актом ¹¹.

Этот смысл (*meaning*), или символическая репрезентация, как ее понимал Гуссерль, есть некая сложная сущность, перед которой стоит трудная задача. В «Идеях к чистой феноменологии» (1982) Гуссерль сделал смелую попытку объяснить, каким образом ноэма решает эту задачу. Референция обеспечивается «предикатами-смыслами», которые, подобно Фрегевым *Sinne*, обладают замечательным свойством выбирать атомарные свойства объектов. Эти предикаты комбинируются в сложные «описания» сложных объектов, как в Расселовой теории дескрипций. Для Гуссерля, близкого в этом вопросе к Канту, ноэма содержит некую иерархию строгих правил. Поскольку Гуссерль представлял себе разумность как детерминированную контекстом целенаправленную деятельность, ментальная репрезентация объекта любой разновидности должна была обеспечить контекст, или «горизонт», ожиданий или «предначертаний» для структурирования входящих данных: «правило, управляющее *возможным* другим сознанием [объекта] как тождественным — возможным, как экземплифицирующим существенно предначертанные типы» (1960: 45). Ноэма должна содержать правило, описывающее все те признаки, которых можно с достоверностью ожидать, исследуя объект того или иного *типа* — признаки, остающиеся «нерушимо одними и теми же: поскольку объективность остается подразумеваемой в качестве *этой* объективности и в качестве объективности этого рода» (1960: 53). Данное правило должно также предписывать предначертания свойств, которые возможны, но не необходимы, признаков объекта этого типа: «Поэтому вместо абсолютно детерминированного смысла всегда имеется некая *рамка* (a frame) *пустого смысла* ...» (1960: 51).

В 1973 году Марвин Минский предложил новую структуру данных для репрезентации обыденного знания, удивительно похожую на Гуссерлеву:

¹¹ «Смысл ..., как мы его определили, не есть конкретная сущность в общем составе ноэмы и некий род живущей в ней абстрактной формы». См.: Husserl (1950). Свидетельство того, что Гуссерль считал, что ноэма ответственна за интенциональность ментальной деятельности, см. в статье Хьюберта Дрейфуса «Перцептуальная ноэма Гуссерля» // (*Dreyfus* 1982).

Фрейм есть структура данных для репрезентации стереотипной ситуации, например нахождения в гостиной определенного рода или похода на день рождения к ребенку...

Фрейм можно представлять себе как сеть узлов и отношений. Верхние уровни фрейма фиксированы и репрезентируют такие вещи, которые всегда истинны относительно данной предполагаемой ситуации. Нижние уровни имеют много *терминалов* – автоматных щелей, которые должны быть заполнены конкретными инстанциями или данными. Для каждого терминала могут иметься особые условия, которым должны удовлетворять его инстанции...

Немалая часть феноменологической силы теории зависит от включения в нее ожиданий и презумпций других родов. *Терминалы фрейма* обычно уже заполнены руководствами “по умолчанию” (1981: 96).

В описанной Минским модели фрейма «верхний уровень» – это развитая версия того, что, по терминологии Гуссерля, остается в репрезентации «нерушимо одним и тем же», а Гуссерлевы предначертания становятся «инстанциями по умолчанию» – теми дополнительными признаками, которые могут в норме ожидаться. В результате в технике AI был сделан шаг от пассивной модели обработки информации к такой модели, которая пытается учесть взаимодействие познающего субъекта и мира. Таким образом, задачи AI смыкаются с задачами трансцендентальной феноменологии. И тот, и другая должны пытаться отыскать в обыденных областях фреймы, построенные из множества исходных предикатов и их формальных отношений.

Хайдеггер, еще до Витгенштейна, осуществил, отвечая Гуссерлю, феноменологическое описание обыденного мира и таких обыденных объектов, как стулья и молотки. Подобно Витгенштейну, он обнаружил, что обыденный мир нельзя репрезентировать посредством некоторого набора бесконтекстных элементов. Именно Хайдеггер заставил Гуссерля рассмотреть именно эту проблему, указав, что есть иные способы «встречи» с вещами, чем отнесение к ним как к объектам, заданным некоторым набором предикатов. Когда мы используем такой инструмент, как молоток, сказал Хайдеггер, мы актуализируем некое умение (которое необязательно должно быть репрезентировано в сознании) в контексте социально организованной связи инструментов, целей и человеческих ролей (которые необязательно должны быть репрезентированы как некоторое множество фактов). Этот контекст, или мир, и наши обыденные способы умелого приспособления к нему, которые Хайдеггер называл «осмотрительностью» («*circumspection*»), суть не нечто такое, что мы лишь *мыслим*, но часть нашей социализации, формирующей то, *каковы мы суть*. Хайдеггер заключает:

Контекст ... можно понимать формально – как систему отношений. Но [ф]еноменальное содержание этих «отношений» и их «членов» ... та-

ково, что они сопротивляются какой бы то ни было математической функционализации; не являются они также и чем-то таким, что мыслится, впервые постулируется в «акте мышления». Они суть такие связи, в которых уже живет озабоченная осмотрительность как таковая (1962: 121–122).

Здесь расходятся пути Гуссерля и AI, с одной стороны, и Хайдеггера и позднего Витгенштейна — с другой. Решающим становится вопрос: «Может ли иметься теория обыденного мира, как всегда считали философы-рационалисты?» Или же фон, на который опирается здравый смысл, — это комбинация умений, практик, различений и так далее, которые не являются интенциональными состояниями и таким образом, *a fortiori*, не обладают каким-либо репрезентативным содержанием, которое можно было бы или следовало бы эксплицировать в терминах элементов и правил?

Сделав шаг, который скоро стал общим местом в кругах AI, Гуссерль попытался избежать проблемы, поставленной Хайдеггером. Гуссерль заявил, что мир, фон значимости, обыденный контекст, — это просто очень сложная система фактов, коррелированная с очень сложной системой убеждений, которые — поскольку у них имеются условия истинности — он называл *очевидностями*. Он считал, что в принципе человек мог бы на время воздержаться от своего бытия в мире и достичь отстраненного описания системы человеческих убеждений. Таким образом человек мог бы выполнить ту задачу, которая неявным образом присутствовала в философии со времен Сократа: сделать явными те убеждения и принципы, которые лежат в основе всякого разумного поведения. Как формулировал это Гуссерль,

[д]аже тот фон..., который мы всегда совместно сознаем, но который в каждое данное мгновение иррелевантен и остается совершенно незамеченным, все же функционирует в соответствии со своими неявными очевидностями (1970: 149).

Поскольку Гуссерль твердо верил, чем общий для людей фон можно сделать явным в качестве некоей системы убеждений, он опередил свое время, подняв вопрос о возможности AI. Обсудив возможность того, что формальная аксиоматическая система могла бы описать опыт, и указав, что такая система аксиом и исходных элементов — по меньшей мере, как мы знаем ее в геометрии, — неспособна описать такие обыденные формы, как «зубчатый» или «линзообразный», Гуссерль оставил открытым вопрос о том, можно ли все же формализовать такие обыденные концепты. (Это было подобно тому, как в AI был поднят и оставлен открытым вопрос о том, можно ли аксиомати-

зировать обыденную физику.) Восприняв Лейбницева мечту о матезисе всяческого опыта, Гуссерль добавляет:

насущенный вопрос состоит ... в том, не может ли иметься ... некая идеализирующая процедура, заменяющая узрваемые интуицией данные на чистые и строгие идеалы и могущая ... служить ... основным средством для построения матезиса опыта (1952: в. 134).

Но, как предсказал Хайдеггер, задача создания полного теоретического описания обыденной жизни оказалась гораздо более трудной, чем это ожидалось сначала. Проект Гуссерля столкнулся с серьезными трудностями, и есть признаки, что с не менее серьезными трудностями столкнулся и проект Минского. По ходу двадцатипятилетних попыток выявления компонентов субъективной репрезентации обыденных объектов, Гуссерль обнаруживал, что он должен включать в описание все новые и новые порции обыденного понимания субъектом повседневности:

Несомненно, даже те задачи, которые возникают, когда мы берем в качестве ограниченных ключевых данных некие отдельно взятые типы объектов, оказываются чрезвычайно сложными и всегда ведут к обширным дисциплинам по мере того, как мы все глубже проникаем в них. Так обстоит дело, например, с ... пространственными объектами (не говоря уже о Природе) как таковыми, психофизическим бытием и человечеством как таковым, культурой как таковой (1960: 54–55).

Он говорил о «гигантской конкретности» ноэмы (1969: 244) и о ее «необъятной сложности» (1969: 246) и печально заключил, когда ему было семьдесят пять, что он вечный начинающий и что феноменология — «бесконечная задача» (1970: 291).

В статье Минского «Концептуальный аппарат для репрезентации знания» имеются намеки, что он взялся за ту же самую «бесконечную задачу», которая в конце концов сломила Гуссерля:

Просто построить базу знаний — сложнейшая интеллектуальная исследовательская проблема ... Мы еще слишком мало знаем о содержании и структуре обыденного знания. «Минимальная» система здравого смысла должна «знать» что-то о причине, следствии, времени, цели, местоположении, процессе и типах знания В этой области требуются серьезные эпистемологические исследовательские усилия (1981: 124).

Для исследователя современной философии наивность и вера Минского поразительны. Феноменология Гуссерля как раз и была таким исследовательским усилием. В сущности, философы от Сократа до Лейбница и раннего Витгенштейна прилагали в этой области серь-

езные эпистемологические усилия в течение двух тысячелетий без заметных успехов.

В свете радикального изменения взглядов Витгенштейном и сокрушительной Хайдеггеровой критики Гуссерля один из нас — Хьюберт — предсказал трудность обработки символической информации. Как отмечает Ньюэлл в своей истории AI, это предупреждение не было принято в расчет:

Основное интеллектуальное возражение Дрейфуса ... состоит в том, что попытки разложения контекста человеческого действия на дискретные элементы обречены на провал. Это возражение основывается на феноменологической философии. К сожалению, в том, что касается AI, это возражение осталось без внимания. Ответы, возражения и анализы, появившиеся как реакция на писания Дрейфуса, попросту прошли мимо этого вопроса, который на самом деле мог бы стать новшеством, если бы он попал в поле зрения (1983: 222–223).

Но попадания в поле зрения соответствующих трудностей не пришлось долго ждать, когда сам мир повседневности отомстил AI, как он отомстил и традиционной философии. Нам представляется, что исследовательская программа, выдвинутая Ньюэллом и Саймоном, прошла три десятилетних стадии. С 1955 по 1965 гг. в этой области исследований, которую тогда называли «когнитивным моделированием», преобладали две исследовательские темы — репрезентация и поиск. Например, Ньюэлл и Смит показали, как компьютер может решать некий класс проблем, основываясь на одном общем эвристическом принципе поиска, известном как анализ «средство-цель», именно использовать любую доступную ему операцию для сокращения расстояния между описанием имеющейся ситуации и описанием цели. Затем они абстрагировали эту эвристическую технику и включили ее в свой Общий Решатель Проблем (ОРП).

Вторая стадия (1965–1975 гг.), на которой лидировали Марвин Минский и Сеймур Пейперт в Массачусетском Технологическом Институте, была посвящена тому, какие факты и правила использовать для репрезентации. Идея заключалась в том, чтобы развить методы для систематической манипуляции знанием в изолированных областях, названных «микромирами». Среди знаменитых программ, написанных в МТИ около 1970-го года, — SHRDLU Терри Винограда, которая умела выполнять команды, записывавшиеся на естественном языке и относившиеся к некоему упрощенному «миру кубиков»; программа проблем аналогии Томаса Эванса; программа анализа сцен Дэвида Уолтца; и программа Патрика Уинстона, умевшая усваивать понятия из примеров.

Надеялись, что эти ограниченные и изолированные микромиры можно будет постепенно делать более реалистическими и комбинированными, так чтобы приблизиться к пониманию в масштабах настоящего мира. Однако исследователи спутали две области, которые, по Хайдеггеру, следует различать — «вселенную» и «мир». Множество взаимосвязанных фактов способно конституировать некую *вселенную*, подобную физической Вселенной, но оно не конституирует *мира*. Мир, например, мир бизнеса, мир театра или мир физиков, — это организованный массив объектов, целей, умений и практик, на основе которых получает свой смысл или становится осмысленной человеческая деятельность. Чтобы понять, в чем различие, можно сопоставить *лишенную смысла* физическую Вселенную с *наделенным смыслом* миром дисциплины физики. Мир физики, мир бизнеса и мир театра получают смысл лишь на фоне общих человеческих забот и интересов. Они суть локальные конкретизации одного общего для всех нас мира здравого смысла. То есть подмиры связаны друг с другом не так, как изолированные физические системы с теми объемлющими системами, которые они *составляют*, а скорее как локальные конкретизации некоего целого, которое они *предполагают в качестве своей предпосылки*. Микромиры же это не миры, а изолированные, лишенные смысла области, и постепенно стало ясно, что их нельзя объединить друг с другом и расширить так, чтобы получить в результате мир обыденной жизни.

На третьей стадии, примерно с 1975-го года по настоящее время, AI борется с тем, что получило название проблемы обыденного знания. Репрезентация обыденного знания всегда была одной из центральных проблем в AI, но два первых периода — когнитивное моделирование и микромиры — характеризовались попыткой избежать проблемы обыденного знания и попытаться получить максимум достижений с минимумом наличного знания. Однако к середине 1970-х годов этой проблемой уже стало необходимо заняться вплотную. Без особого успеха были испытаны различные структуры данных, например фреймы Минского и сценарии Роджера Шенка. Проблема обыденного знания не позволяла AI даже приступить к выполнению двадцатилетней давности предсказания Саймона о том, что «не более чем через двадцать лет машины смогут выполнять любую работу, которую способен выполнить человек» (1965: 96).

В сущности, проблема обыденного знания не дала AI за последнее десятилетие продвинуться ни на шаг вперед. Виноград одним из первых увидел ограниченности SHRDLU и всех этих попыток расширить рамки микромирового подхода с помощью сценариев и фреймов. «Разуверившись» в AI, он теперь занят тем, что преподает Хайдеггера в рамках своего курса по computer science в Стэнфорде и указывает на «трудность формализации того обыденного фона, который детер-

минирует, какие именно сценарии, цели и стратегии имеют отношение к делу и как именно они взаимодействуют друг с другом» (1984: 142).

В чем находит надежду и силы AI, забравшись в этот тупик, так это в том убеждении, что раз люди, очевидно, как-то решили проблему обыденного знания, то, стало быть, она должна иметь решение. Но люди, в норме, могут и вообще не использовать обыденное знание. Как указали Хайдеггер и Витгенштейн, обыденное понимание вполне может состоять в *обыденном ноу-хау (everyday know-how)*. Под «ноу-хау» мы понимаем не процедурные правила, а знание, что делать, в огромном множестве частных случаев¹². Например, обыденную физику оказалось чрезвычайно трудно обстоятельно объяснить в форме некоторого набора фактов и правил. Если попытаться сделать это, то окажется, что либо для понимания найденных тобой фактов и правил тебе опять же потребуется здравый смысл, либо в результате получаются формулы такой сложности, что представляется крайне маловероятным, чтобы они помещались в сознании ребенка.

Занятия теоретической физикой тоже требуют фоновых умений, которые могут оказаться неформализуемыми, однако саму эту область можно описать с помощью абстрактных законов, которые не апеллируют к фоновым умениям. Из этого обстоятельства AI-исследователи ошибочно заключают, что и обыденная физика должна быть выразима в виде некоторого набора абстрактных принципов. Но ведь может быть и так, что проблема нахождения *теории* обыденной физики неразрешима потому, что у этой области вообще нет никакой теоретической структуры. Каждый день в течение нескольких лет играя со всяческими жидкостями и твердыми телами, ребенок, возможно, просто научается различать прототипические случаи твердых тел, жидкостей и т. д. и научается типичным адекватным реакциям на их типичное поведение в типичных обстоятельствах. То же может быть верно и для социального мира. Если и в самом деле фоновое понимание — это некое умение, а умения основаны на цельных образах, а не правилах, то можно ожидать, что символьные репрезентации окажутся неспособными ухватить наше обыденное понимание.

В свете этого тупика, классический, символичный AI становится все более и более похожим на превосходный пример того, что Имре Лакатош (1978) назвал вырождающейся исследовательской программой. Как мы видели, AI начался при благоприятных предзнаменованиях с работы Ньюэлла и Саймона в РЭНДе и к концу 1960-х годов превратился в процветающую исследовательскую программу. Минский предсказывал, что «еще при жизни этого поколения проблема

¹² Эта концепция умения нашла свое выражение и защиту в книге: *Dreyfus and Dreyfus* (1986).

создания «искусственного интеллекта» будет в общем и целом решена» (1977: 2). Затем, довольно-таки внезапно, вся эта область столкнулась с неожиданными трудностями. Сформулировать теорию здравого смысла оказалось гораздо труднее, чем ожидалось. Вопреки надеждам Минского, дело состояло не в том, чтобы каталогизировать несколько сотен тысяч фактов. Центральное значение приобрела проблема обыденного знания. Умонастроение Минского решительным образом поменялось за пять лет. Он сообщил одному репортеру, что «проблема AI — одна из труднейших, за какие когда-либо принималась наука» (Kolata 1982: 1237).

В конце концов, рационалистская традиция предстала перед судом эмпирической проверки — и провалилась. Идея построения формальной, атомистской теории мира обыденного здравого смысла и репрезентации такой теории в символьном манипуляторе столкнулась в точности с теми трудностями, которые были обнаружены Хайдеггером и Витгенштейном. Интуитивное убеждение Франка Розенблатта в том, что формализовать мир и таким образом формально специфицировать разумное поведение будет безнадежно трудно, — это убеждение оправдалось. Его исследовательская программа (состоявшая в том, чтобы использовать компьютер для воплощения холистской модели идеализированного мозга), которая была загнана в подполье, но никогда не была по-настоящему опровергнута, вновь стала предметом реально возможного выбора.

Журналисты в своих описаниях истории AI приводят примеры того, как анонимные клеветники поливали грязью Розенблатта как «продавца змеиного жира»:

Нынешние исследователи помнят, как о Розенблатте и о работе его машины важным тоном делались резкие заявления. «Он был мечтой рекламного агента, — говорит один ученый, — настоящий шаман. Послушать его, так Перцептрон способен делать прямо-таки фантастические вещи. И может быть, так оно и есть. Но вот работа Франка этого не доказывала» (McCorduck 1979: 87).

На самом же деле, он гораздо яснее говорил о способностях и ограничениях различных типов перцептронов, чем Саймон и Минский о своих символьных программах¹³. Теперь его реабилитируют.

¹³ Вот некоторые типичные цитаты из Розенблаттовых «Принципов нейродинамики»: «По ходу эксперименту обучения перцептрону, как правило, дается некая последовательность паттернов, содержащих представителя каждого типа или класса, который должен быть различен, и в соответствии с некоторым правилом модификации памяти «подкрепляется» подходящий выбор ответной реакции. Затем перцептрон подвергают тес-

товому стимулированию, и устанавливается вероятность правильного ответа для данного класса стимулов... Если тестовый стимул активирует множество сенсорных элементов, которые полностью отличаются от тех элементов, что были активированы в предыдущих случаях подвержения перцептрона стимулам того же самого класса, то эксперимент представляет собой тест на «чистое обобщение». Простейший из перцептронов не обладает способностью к чистому обобщению, но можно показать, что вполне удовлетворительно ведет себя в экспериментах по различению, особенно если тестовый стимул почти тождествен одному из ранее воспринятых паттернов (с. 68) ... Рассматривавшиеся до сих пор перцептроны по своим способностям распознавания фигур и тенденциям к организации гештальтов мало чем похожи на человеческие субъекты (с. 71) ... Распознавание последовательностей в рудиментарной форме вполне под силу подходящим образом организованным перцептронам, однако проблема фигуральной организации и сегментации становится здесь столь же серьезной, как и в случае восприятия статического паттерна (с. 72) ... В случае простого перцептрона паттерны распознаются прежде «отношений»; в сущности, такие абстрактные отношения, как «*A* выше, чем *B*» или «этот треугольник находится внутри окружности», вообще не абстрагируются в качестве таковых, но могут быть усвоены лишь с помощью процедуры полного заучивания на память, по ходу которой каждый случай, в котором данное отношение имеет место, преподается перцептрону отдельно (с. 73) ... Сеть, состоящая из менее чем трех слоев сигналопередающих блоков, или сеть, состоящая исключительно из линейных элементов, линейно соединенных друг с другом, неспособна обучиться различать классы паттернов в изотропной среде (где любой паттерн может появиться во всех возможных местоположениях сетчатки, без пограничных эффектов) (с. 575) ... В предыдущих главах был описан ряд умозрительных моделей, которые, вероятно, способны научиться секвенциальным программам, разложению речи на фонемы, а также усваивать содержательный «смысл» имен существительных и глаголов с простыми сенсорными денотатами. Такие системы представляют верхний предел абстрактного поведения у перцептронов, рассматривавшихся по сей день. К их недостаткам относится отсутствие удовлетворительной «временной памяти», неспособность не слишком сложно воспринимать абстрактные топологические отношения и неспособность изолировать осмысленные фигуральные сущности или объекты, кроме как при некоторых особых обстоятельствах (с. 577) ... Приложения, с наибольшей вероятностью осуществимые с перцептронами, описанным в этой книге, — это распознавание букв и «читающие машины», распознавание речи (для случаев членораздельно и отдельно друг от друга произносимых слов) и крайне ограниченные способности распознавания рисунков или распознавание объектов на простом фоне. К «восприятию» в более широком смысле, возможно, окажутся способны потомки нынешних моделей, но предстоит добыть еще очень много фундаментального знания, прежде чем достаточно изощренное строение по-

Дэвид Румелхарт, Джеффри Хинтон и Джеймс Маклелланд отражают эту новую оценку его новаторских работ:

Работы Розенблатта в то время были очень спорны, а конкретные модели, предложенные им, не оправдывали надежд, которые он возлагал на них. Но его видение человеческой системы по обработке информации как динамической, интерактивной, самоорганизующейся системы лежит в основе подхода PDP (1986: i.45).

Ясно, что исследования перцептрона ... предвосхитили многие результаты, которые используются сегодня. Критику перцептронов Минским и Пейпертом широко и ошибочно интерпретировали как разрушающую доверие к ним, в то время как их работа всего лишь показывала ограниченность способностей самого ограниченного класса перцептроноподобных механизмов и ничего не говорила о более сильных, многослойных моделях (1986: ii.535).

Обманутые в своих ожиданиях исследователи AI, уставшие цепляться за исследовательскую программу, которую Джерри Летвин в начале 1980-х годов охарактеризовал как «единственную соломинку на плаву», толпой набросились на новую парадигму. В первый день поступления на рынок было распродано шесть тысяч экземпляров книги Румелхарта и Маклелланда «Параллельная распределенная обработка информации», и сейчас печатается еще тридцать тысяч. Как сказал Пол Смоленский,

за последние лет пять коннекционистский подход к когнитивному моделированию из мало кому известного культа, к которому причислила себя горстка верных, превратился в движение столь мощное, что последние конференции Общества когнитивной науки стали походять на тусовки коннекционистов (В печати).

Если многослойные сети успешно выполняют свои обещания, то исследователям придется расстаться с убеждением Декарта, Гуссерля и раннего Витгенштейна, будто единственный способ продуцировать разумное поведение — это отобразить мир посредством некоей формальной теории. И хуже того, возможно, придется расстаться с еще более фундаментальным интуитивным взглядом, коренящимся в самых истоках философии, согласно которому у каждого аспекта действительности должна иметься своя теория, то есть должны иметься элементы и принципы, в терминах которых можно описать и объяснить умопостижимость любой области. Нейронные сети могут показать, что Хайдеггер, поздний Витгенштейн и Розенблатт были правы

зволит перцептрону соревноваться с человеком в условиях нормальной среды» (с. 583).

в том, что мы разумно ведем себя в этом мире, не имея никакой теории мира. Если наличие теории не *необходимо* для объяснения разумного поведения, то мы должны быть готовы рассмотреть, вопрос о том, *возможно* ли вообще такое теоретическое объяснение в обыденных областях.

Люди, занимающиеся моделированием нейронных сетей, под влиянием символично-манипуляторного AI, прилагают значительные усилия — натренировав свои сети исполнять ту или иную задачу, — к поиску признаков, репрезентируемых индивидуными узлами и множествами узлов. Результаты, полученные до сей поры, неоднозначны. Возьмем сеть Хинтона (1986), обучающуюся понятиям посредством распределенных репрезентаций. Эту сеть можно натренировать кодировать такие связи в предметной области, которую люди концептуализируют в терминах признаков, и при этом сети не сообщаются признаки, используемые людьми. Хинтон дает примеры случаев, в которых некоторые узлы в натренированной сети можно интерпретировать как соответствующие признакам, используемым людьми, хотя эти узлы лишь приблизительно соответствуют данным признакам. Однако большая часть узлов вообще не поддается никакой семантической интерпретации. Признак, используемый в символической репрезентации, либо присутствует, либо отсутствует. В сети же, хотя некоторые узлы более активны, когда в предметной области наличествует некий признак, степень активности не только меняется в зависимости от присутствия или отсутствия данного признака, но на эту степень активности влияют также и другие признаки.

Хинтон выбрал такую предметную область — семейные отношения, — которую люди трактуют именно в терминах признаков, как правило, замечаемых людьми — принадлежность к тому или иному поколению и национальности. Затем он разбирает такие случаи, в которых, начиная с некоторых случайно выбранных величин сил в исходной конфигурации связей, некоторые узлы можно, после обучения, толковать как репрезентирующие данные признаки. Вычисления, использующие Хинтонову модель, показывают, однако, что даже его модель, как видно, научается своим ассоциациям в отношении некоторых случайно выбранных величин сил в исходной конфигурации связей, не используя сколько-нибудь очевидным образом эти обыденные признаки.

В некотором очень узком смысле любую успешно натренированную многослойную сеть можно интерпретировать в терминах признаков — не обыденных признаков, а тех, что мы будем называть высоко-абстрактными признаками. Рассмотрим простой случай слоев бинарных элементов, активируемых посредством связей, но не боковых или обратных связей, а связей, направленных вперед. Чтобы сконструировать подобное объяснение на основе сети, обученной определен-

ным ассоциациям, каждый узел, который размещается на один уровень выше входных узлов, можно интерпретировать как распознающий наличие входного паттерна, принадлежащего некоторому набору таких паттернов. (Некоторые из этих паттернов использовались в ходе обучения сети, а другие никогда не использовались.) Если этому набору входных паттернов, распознаваемых некоторым данным узлом, дать какое-то выдуманное имя (почти наверняка никакое имя из нашего словаря не будет ему подходить), то данный узел можно было бы интерпретировать как распознающий тот высоко-абстрактный признак, которому мы присвоили это имя. Таким образом, каждый узел, который размещается на один уровень выше входных узлов, можно охарактеризовать как распознаватель некоторого признака. Точно так же, каждый узел, который размещается на один уровень выше этих узлов, можно охарактеризовать как распознаватель признака более высокого порядка, определяемого как наличие среди распознавателей признаков первого уровня паттерна из некоторого заданного множества. И так далее вверх по иерархии.

То обстоятельство, что интеллект, определяемый как знание некоторого множества ассоциаций в данной предметной области, всегда можно описать в терминах отношений между высоко-абстрактными признаками данной области – это обстоятельство, однако, не сохраняет в себе рационалистическую интуицию, согласно которой эти объяснительные признаки должны схватывать существенную структуру данной области – так, чтобы, основываясь на них, можно было построить теорию данной области. Если обучить сеть еще одной ассоциации между входом и выходом (где до обучения данный вход продуцировал иной выход, чем тот, которому мы предполагаем научить сеть), то интерпретацию по меньшей мере некоторых узлов пришлось бы изменить. Таким образом, оказалось бы, что признаки, соответствовавшие – до последнего тура обучения – некоторым узлам, не являлись инвариантными структурными признаками данной области.

Раз мы отбросили философский подход классического AI и приняли а-теоретический подход моделирования нейронных сетей, остается один вопрос: в какой мере такая нейронная сеть может смоделировать обыденный интеллект? Исследователи AI в классической парадигме тут же скажут, – как, на самом деле, указывал уже и Розенблатт, – что люди, моделирующие нейронные сети до сих пор сталкивались с трудностями в области пошагового решения проблем. Коннекционисты отвечают, что они уверены, что со временем решат эту проблему. Однако этот ответ слишком уж напоминает тот ответ, который в шестидесятых годах давали сторонники символьной манипуляции, реагируя на критические высказывания, что их программы де очень слабы в распознавании образов. Продолжается старинная

борьба между интеллектуалистами, которые считают, что раз они занимаются бесконтекстной логикой, то они ухватывают обыденное познание, но которые слабы в понимании восприятия, и гештальтистами, у которых имеются начатки объяснения восприятия, но отсутствует какое-нибудь объяснение обыденного познания¹⁴. Можно предположить, используя метафору правого и левого [полушария] мозга, что, быть может, мозг или сознание применяет, когда нужно одну из этих двух стратегий, а когда нужно — вторую. В таком случае проблема состоит в том, как объединить эти две стратегии. Нельзя просто переключаться с одной стратегии на другую и обратно, ибо, как показали Хайдеггер и гештальтисты, решающую роль в определении релевантности играет прагматический фон, даже в обыденной логике и решении проблем; и специалисты в любой области, вплоть до логики, ухватывают операции в терминах их функциональных сходств.

Рассматривать объединение этих стратегий преждевременно, ибо по сей день ни за одной из двух не числится достаточно свершений, чтобы можно было считать ее солидно обоснованной. Быть может, моделирование нейронных сетей попросту использует сейчас заслуженный им шанс на поражение, как ранее его использовал символичный подход.

И все же по мере того, как каждый из двух подходов пробивает себе путь вперед, следует принимать в расчет одно важное различие между ними. Подход физико-символьных систем проваливается, по видимому, потому, что попросту неверно предполагать, что для каждой области должна иметься теория этой области. Но моделирование нейронных сетей не связывает себя с каким-нибудь иным философским допущением. Тем не менее, быть может, слишком трудно построить интерактивную сеть, которая была бы существенно подобна той сети, какую развил наш мозг. На самом деле, проблема обыденного знания, в течение пятнадцати лет блокировавшая прогресс технологии символической репрезентации, возможно, маячит на горизонте нейронных сетей, хотя исследователи еще и не распознали ее. Все люди, занимающиеся моделированием нейронных сетей, соглашаются, что для того, чтобы сеть обладала интеллектом, она должна обладать способностью к обобщениям; то есть, если задано достаточно примеров входов, ассоциируемых с одним конкретным выходом, она должна ассоциировать другие входы того же типа с тем же самым выходом. Встает, однако, вопрос: что считать тем же самым типом? Конструктор сетей имеет в виду некоторое конкретное определение типа, тре-

¹⁴ Об одной новейшей влиятельной концепции восприятия, отрицающей надобность ментальных репрезентаций, см.: *Gibson* (1979). Гибсон и Розенблатт в 1955 году были соавторами научной статьи, написанной для американских военно-воздушных сил см.: *Gibson, Olum, and Rosenblatt* (1955).

буемое для разумного обобщения, и считает успехом, если сеть осуществляет обобщения на другие образцы того же самого типа. Но когда сеть производит некую неожиданную ассоциацию, можем ли мы сказать, что она потерпела провал в обобщении? Точно так же можно было бы сказать, что все время до сих пор сеть действовала в соответствии с неким иным определением типа, и сейчас просто раскрылось это различие в определениях. (Все вопросы вида «продолжи последовательность», которые можно отыскать в тестах на уровень интеллекта, имеют более одного возможного ответа, но у большинства людей совпадают интуитивные представления о простоте, разумности и, стало быть, приемлемости.)

Люди, занимающиеся моделированием нейронных сетей, пытаются избежать этих двусмысленностей и заставляют сеть выдавать «разумные» обобщения, заранее задавая ей некоторое допустимое семейство обобщений, т. е. допустимых преобразований, которые будут считаться допустимыми обобщениями (пространство гипотез). Эти модельщики пытаются затем построить архитектуру своих сетей таким образом, чтобы они преобразовывали входы в выходы лишь теми способами, которые присутствуют в пространстве гипотез. И тогда обобщение будет возможно только на условиях автора архитектуры. Правда нескольких примеров будет недостаточно, чтобы единственным образом идентифицировать подходящий элемент пространства гипотез, однако после достаточного числа примеров останется одна-единственная гипотеза, объясняющая все эти примеры. И тогда сеть обучается подходящему принципу обобщения. То есть какой бы вход мы ни взяли, он продуцирует выход, подходящий с точки зрения автора сети.

Проблема здесь в том, что автор сети с помощью архитектуры сети сделал так, что некоторые возможные обобщения останутся никогда не найденными. Все это прекрасно для игрушечных проблем, в которых не встает вопрос о том, что такое разумное обобщение, но в ситуациях настоящей жизни немалая часть человеческого интеллекта как раз и заключается в обобщении, уместном в данном контексте. Если же автор сети ограничивает ее неким предзаданным классом подходящих ответов, то сеть будет выказывать тот интеллект, который встроено в нее автором для одного данного контекста, но не будет обладать здравым смыслом, который позволил бы ей приспособиться к другим контекстам, как смог бы приспособиться истинно человеческий интеллект.

Возможно, сеть должна иметь размеры, строение и конфигурацию начальных связей, подобные размерам, строению и конфигурации человеческого мозга, если мы хотим, чтобы у нее были человеческие представления о том, что такое подходящее обобщение. Если мы хотим, чтобы она научалась из своего собственного «опыта» делать

ассоциации, подобные тем, что делает человек, а не того, чтобы ее научили делать ассоциации, специфицированные ее тренером, то в таком случае у сети должны быть наши, человеческие, представления о том, что такое подходящий выход, а это значит, что она должна разделять с нами наши нужды, желания и эмоции и иметь тело, подобное телу человека с соответствующими физическими движениями, способностями и уязвимостью.

Если правы Хайдеггер и Витгенштейн, люди гораздо холистичнее нейронных сетей. Интеллект должен быть мотивирован целями и задачами организма, и в том числе и теми целями, которые организм черпает из наличной культуры. Если минимальная единица анализа — целостный организм, сцепленный с некоторым целостным миром культуры, то нейронным сетям, как и символично программированным компьютерам, предстоит пройти еще очень долгий путь.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

- Anderson J. A.* (1978). *Neural Models with Cognitive Implications // Basic Processing in Reading / Berse D. La, Samuels S. J. (eds.), NJ: Erlbaum.*
- Boden M.* (1977) *Artificial Intelligence and Natural Man.* New York: Basic Books.
- Dreyfus H.* (1979) *What Computers Can Do, 2nd edn.* New York: Harper & Row.
- (1988). *Being-in-the-World: A Commentary on Division I of «Being and Time».* Cambridge, Mass.: MIT Press.
- (ed.) (1982). *Husserl, Intentionality and Cognitive Science.* Cambridge, Mass.: MIT Press.
- and *Dreyfus S.* (1986). *Mind Over Machine.* New York: Macmillan.
- Gibson J. J.* (1979). *The Ecological Approach to Visual Perception.* Boston: Houghton- Mifflin.
- Gibson J. J., Olum P., Rosenblatt F.* (1955). *Parallax and Perspective During Aircraft Landing // American Journal of Psychology, № 68, pp. 372–85.*
- Grossberg S.* (1982). *Studies of Mind and Brain: Neural Principles of Learning, Perception, Development, Cognition and Motor Control.* Boston: Reidel Press.
- Haugeland J.* (1985). *Artificial Intelligence: The Very Idea.* Cambridge, Mass.: MIT Press.
- Hebb D. O.* (1949). *The Organization of Behavior.* New York: Wiley.
- Heidegger M.* (1962). *Being and Time.* New York: Harper & Row.
- Hinton G.* (1986). *Learning Distributed Representations of Concepts // Proceedings of the Eighth Annual Conference of the Cognitive Science Society.* Amherst, Mass.: Cognitive Science Society.
- Hobbes T.* (1958). *Leviathan.* New York: Library of Liberal Arts.
- Husserl E.* (1950). *Ideen Zu Einer Reinen Phänomenologie und Phänomenologischen Philosophie.* The Hague: Nijhoff.

- (1952). *Ideen Zu Einer Reinen Phänomenologie und Phänomenologischen Philosophie*. bk. 3 in Vol. 5, *Husserliana*. The Hague: Nijhoff.
- (1960). *Cartesian Meditations*, trans. D. Cairns. The Hague: Nijhoff.
- (1969). *Formal and Transcendental Logic*, trans. D. Cairns. The Hague: Nijhoff.
- (1970). *Crisis of European Sciences and Transcendental Phenomenology*, trans. D. Car. Evanston: Northwestern University Press.
- (1982). *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy*, trans. F. Kersten. The Hague: Nijhoff.
- Kohonen T.* (1977). *Associative Memory: A System-Theoretical Approach*. Berlin: Springer-Verlag.
- Kolata O.* (1982). *How Can Computers Get Common Sense?* // *Science*, 24 Sept, № 217, p. 1237.
- Lakatos I.* (1978). *Philosophical Papers* / Worrall J. ed. Cambridge: Cambridge University Press.
- Leibniz C.* (1951). *Selections* / Wiener P. ed. New York: Scribner.
- Lighthill Sir James* (1973). *Artificial Intelligence: A General Survey* // *Artificial Intelligence: A Paper Symposium*. London: Science Research Council.
- McCorduck P.* (1979). *Machines Who Think*. San Francisco: W. Freeman.
- Minsky M.* (1977). *Computation: Finite and Infinite Machines*. New York: Prentice-Hall.
- (1981). *A Framework for Representing Knowledge* // *Mind Design* / Haugeland J. (ed.), pp. 95–128. Cambridge, Mass.: MIT Press.
- and *Papert S.* (1969). *Perceptions: An Introduction to Computational Geometry*. Cambridge, Mass.: MIT Press.
- Newell A.* (1983). *Intellectual Issues in the History of Artificial Intelligence* // *The Study of Information: Interdisciplinary Messages*, Machlup F., Mansfield U. (eds.), pp. 196–227. New York: Wiley.
- and *Simon H.* (1958). *Heuristic Problem Solving: The Next Advance in Operations Research* // *Operations Research* 6 (Jan.–Feb.), p. 6.
- (1981). *Computer Science as Empirical Inquiry: Symbols and Search* // *Mind Design* / Haugeland J. (ed.), pp. 35–66. Cambridge, Mass.: MIT Press.
- Rosenblatt F.* (1958). *Mechanisation of Thought Processes: Proceedings of a Symposium Held at the National Physical Laboratory*, v. 1. London: HMS Office.
- (1962a). *Principles of Neurodynamics: Perceptions and the Theory of Brain Mechanisms*. Washington, BC: Spartan Books.
- (1962b). *Strategic Approaches to the Study of Brain Models* // *Principles of Self-Organization* / Foerster H. von (ed.), Elmsford, NY: Pergamon Press.
- Rumelhart D. E., McClelland J. L.* (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, 2 vols. Cambridge, Mass.: MIT Press.
- and *Norman D. A.* (1981). *A Comparison of Models* // *Parallel Models of Associative Memory* / Hinton G., Anderson J. (eds.), pp. 3–6. Hillsdale, NJ: Erlbaum.

Simon H. (1965). *The Shape of Automation for Men and Management*. New York: Harper&Row.

Smolensky P. (1988). On the Proper Treatment of Connectionism // *Behavioral and Brain Sciences*, № 11, p. 1074.

Winograd T. (1976). *Artificial Intelligence and Language Comprehension // Artificial Intelligence and Language Comprehension*. Washington, DC: National Institute of Education.

– (1984). *Computer Software for Working with Language // Scientific American*, Sept., p. 142ff.

Wittgenstein L. (1948). *Last Writings on the Philosophy of Psychology*, Vol. I, trans. corrected. Chicago: University of Chicago Press, 1982.

– (1953). *Philosophical Investigations*. Oxford: Basil Blackwell.

– (1960). *Tractatus Logico-Philosophicus*. London: Routledge & Kegan Paul.

– (1975). *Philosophical Remarks*. Chicago: University of Chicago.