

Федеральное государственное бюджетное учреждение науки Институт  
философии Российской академии наук

На правах рукописи

Михайлов Игорь Феликсович

## **Когнитивные основания социальности**

Специальность 09.00.01 —  
онтология и теория познания

Диссертация на соискание учёной степени  
доктора философских наук

Научный консультант:  
академик, доктор философских наук, профессор  
Лекторский Владислав Александрович

Москва — 2021

## Оглавление

	Стр.
<b>Введение</b> . . . . .	6
<b>Глава 1. Философско-методологические основания</b> . . . . .	16
1.1 О соотношении философского и научного знания . . . . .	16
1.1.1 Философия как концептуальный анализ. Наука как порождение гипотез или моделей, соответствующих фактам	16
1.1.2 Онтологические презумпции . . . . .	17
1.1.3 Понятие научных онтологий . . . . .	21
1.1.4 Вывод из раздела 1.1 . . . . .	32
1.2 Общая онтология когнитивных и социальных наук . . . . .	32
1.2.1 Необходимость анализа онтологий когнитивных и социальных теорий . . . . .	32
1.2.2 Понятие метаонтологии . . . . .	34
1.2.3 Сведение свойств к отношениям влечёт сетевой подход . .	38
1.2.4 Возможные контраргументы и ответы на них . . . . .	40
1.2.5 Вывод из раздела 1.2 . . . . .	43
1.3 Философия сознания и её значение для когнитивных наук . . . .	44
1.3.1 Классификация позиций в философии сознания. Функционализм . . . . .	44
1.3.2 Коммуникативный функционализм. Пропозициональные отношения как коммуникационные модальности . . . . .	54
1.3.3 Невыразимость субъективного и перспектива первого лица	56
1.3.4 Проблема «я». Может ли «я» умереть? . . . . .	60
1.3.5 Сознание и осознание . . . . .	68
1.3.6 Интенциональность. Коммуникативная интенциональность	74
1.3.7 Вывод из раздела 1.3 . . . . .	90
1.4 Проблема феноменального опыта . . . . .	90
1.4.1 Важность проблемы qualia . . . . .	90
1.4.2 Квалиа как репрезентации . . . . .	95
1.4.3 Аргументы от qualia против функционализма . . . . .	96

	Стр.
1.4.4	Интерпретация мысленных экспериментов . . . . . 100
1.4.5	(Не)выразимость и (не)репрезентативность квалиа . . . . . 102
1.4.6	Вывод из раздела 1.4 . . . . . 104
1.5	Вывод из главы 1 . . . . . 105
<b>Глава 2.</b>	<b>Концептуальные основы когнитивных наук . . . . . 107</b>
2.1	Развитие психологии в направлении вычислений . . . . . 107
2.1.1	Кант и проблема научности психологии . . . . . 107
2.1.2	Качественно-субстанциальный и количественно-функциональный подходы . . . . . 110
2.1.3	Объектная и функциональная онтологии . . . . . 112
2.1.4	Коммуникативное понимание знания . . . . . 113
2.1.5	Вычислительный образ науки . . . . . 116
2.1.6	Вычислительный этап развития когнитивной психологии . 131
2.1.7	Вывод из раздела 2.1 . . . . . 133
2.2	Основные когнитивные модели . . . . . 134
2.2.1	Символизм . . . . . 135
2.2.2	Коннекционизм . . . . . 137
2.2.3	Теория когнитивного К. В. Анохина как разновидность сетевых подходов . . . . . 141
2.2.4	Воплощённый разум и «динамические системы» . . . . . 144
2.2.5	Предсказывающий разум . . . . . 145
2.2.6	Теория функциональных систем как предтеча «предсказывающего разума» . . . . . 151
2.2.7	Вывод из раздела 2.2 . . . . . 159
2.3	Проблема вычислений в когнитивной науке . . . . . 159
2.3.1	Вычисления и проблема их определения . . . . . 160
2.3.2	Уровни вычислений . . . . . 165
2.3.3	Типы вычислений . . . . . 170
2.3.4	Антикомпьютеризм . . . . . 172
2.3.5	Термодинамическая стоимость вычислений . . . . . 173
2.3.6	Вывод из раздела 2.3 . . . . . 176
2.4	Репрезентации в когнитивных науках . . . . . 176
2.4.1	Понятие репрезентации в исторической перспективе . . . . . 176

2.4.2	Трудности и парадоксы в современном понимании репрезентаций . . . . .	182
2.4.3	Позиция «слабого» репрезентационализма . . . . .	185
2.4.4	Вывод из раздела 2.4 . . . . .	188
2.5	Дискуссии о биологической или вычислительной природе мозга .	189
2.5.1	Школа Коштыянца и «биологический мозг» . . . . .	189
2.5.2	Когниции как эмерджентные эффекты сетевых вычислительных архитектур . . . . .	199
2.5.3	Вывод из раздела 2.5 . . . . .	202
2.6	Вывод из главы 2 . . . . .	202
<b>Глава 3. От когнитивного к социальному . . . . .</b>		<b>204</b>
3.1	Социальная онтология . . . . .	204
3.1.1	Номиналистический взгляд на общество . . . . .	204
3.1.2	Сеть как предельный случай социальной онтологии . . . . .	209
3.1.3	Вывод из раздела 3.1 . . . . .	212
3.2	Концепция метасети . . . . .	213
3.2.1	Социология сетевых обществ . . . . .	213
3.2.2	Когнитивные надстройки над социальными сетями . . . . .	224
3.2.3	Метасетевой подход к социально-когнитивным процессам .	227
3.2.4	Вывод из раздела 3.2 . . . . .	234
3.3	Сетевые и вычислительные подходы в биологии . . . . .	234
3.3.1	Онтологии жизни . . . . .	235
3.3.2	Математические подходы в биологии . . . . .	238
3.3.3	Информационные пути и сети в живых тканях . . . . .	240
3.3.4	«Принцип свободной энергии» в биологии . . . . .	242
3.3.5	Вычислительный подход в биологии . . . . .	245
3.3.6	Вывод из раздела 3.3 . . . . .	246
3.4	Когниции в социальной среде . . . . .	247
3.4.1	Общество как на результат когнитивной эволюции . . . . .	247
3.4.2	Общество как подсистема когнитивного механизма . . . . .	250
3.4.3	Мультиагентные системы . . . . .	251
3.4.4	Взгляд на нейроэтику . . . . .	266
3.4.5	Вывод из раздела 3.4 . . . . .	269

	Стр.
3.5 Люди как (ир)рациональные агенты . . . . .	270
3.5.1 Мыслим ли мы рационально? . . . . .	270
3.5.2 Логично ли мышление? . . . . .	275
3.5.3 Вывод из раздела 3.5 . . . . .	277
3.6 Как возможна когнитивная теория общества? . . . . .	278
3.6.1 Когнитивность как методологическое обязательство . . . . .	278
3.6.2 Образы когнитивно-социальных наук . . . . .	280
3.6.3 Проект когнитивной науки об обществе . . . . .	288
3.6.4 Социальная организация как естественный параллельный компьютер . . . . .	289
3.6.5 Вывод из раздела 3.6 . . . . .	294
3.7 Вывод из главы 3 . . . . .	294
<b>Заключение . . . . .</b>	<b>296</b>
<b>Список сокращений и условных обозначений . . . . .</b>	<b>305</b>
<b>Словарь терминов . . . . .</b>	<b>306</b>
<b>Список литературы . . . . .</b>	<b>307</b>
<b>Список рисунков . . . . .</b>	<b>339</b>
<b>Список таблиц . . . . .</b>	<b>340</b>

## Введение

В ряде эмпирических наук о человеке — в когнитивных науках, нейрофизиологии, социологии и др. — наблюдается возрастающий интерес к изучению влияния нервно-психической организации человека на его социальную организацию. Появляются факты, свидетельствующие об определяющем воздействии нейронных механизмов мозга, химического синтеза гормонов и нейромедиаторов, когнитивных способностей (памяти, склонностей и т.п.), способностей к обучению, распознаванию и категоризации и др. — на скорость формирования, количественные и структурные характеристики социальных связей. Т. е., если до сих пор философы рассматривали в основном детерминацию когнитивной сферы со стороны социальных структур и процессов, то *новые эмпирические данные свидетельствуют об обратной детерминации социальности со стороны когнитивных способностей и, напротив, ограничений, которая может оказаться не менее, а более существенной.* Таким образом, философия оказывается перед новым вызовом со стороны конкретных наук: с одной стороны, появляется новый материал для подтверждения и/или уточнения известных философских положений о единстве биологического и социального в человеке, социокультурной детерминации сознания и познавательной деятельности, для дальнейших размышлений над фундаментальными проблемами творческой сущности сознания, свободы воли и автономии человеческой личности. Вместе с тем, с другой стороны, перед философией открывается новое поле для методологической работы: имея возможность взглянуть на проблему со стороны и немного «сверху», она должна предложить адекватную методологию интеграции усилий отдельных дисциплин — нейрофизиологии, биологии и психологии индивида, — направленных на объяснение социальных фактов и обоснование социальных наук.

На фоне бурного развития нано-, био-, информационных, когнитивных и социальных наук, результаты которых имеют огромное прикладное значение, существенным вызовом для философии является, во-первых, определение форм и методов интеграции указанных направлений, и во-вторых, исследование этических и гуманитарных аспектов данного этапа научной революции. Предлагаемое диссертационное исследование концентрируется в основном на первом вызове.

Предлагаемое исследование соответствует приоритетным целям отечественной науки, общества и государства, к которым относится углубление знаний о когнитивных способностях человека с целью дальнейшего развития человеческого капитала, лучшее понимание реальных социальных механизмов и динамики с целью повышения и укрепления общественной безопасности, междисциплинарная интеграция науки с целью обеспечения технологических прорывов и обеспечения международной конкурентоспособности страны.

Научные области, находящиеся в сфере интереса данного исследования, выиграют от междисциплинарного синтеза, сопровождающегося профессиональным философским анализом концептуальных рамок и онтологий, лежащих в основе их взаимодействия. Интенсифицируются исследования когнитивных функций, лежащих в основе человеческой социальности; рамок, налагаемых обществом и культурой на когнитивные способности человека, и наоборот — ограничений, налагаемых когнитивными особенностями человеческих особей на их социальные связи; языковой семантики и прагматики в связи с ролью языка в качестве интерфейса между мозгом и обществом.

**Степень разработанности темы.** В российской философской литературе тема соотношения когнитивного и социального разрабатывалась и продолжает разрабатываться в различных аспектах. Традиции деятельностного подхода, развитого в советской философии и психологии, продолжает на новой основе В. А. Лекторский [1; 2]. Среди работ, знакомящих отечественного читателя с классикой мировой философии сознания, наиболее заметны книги Н. С. Юлиной [3], В. В. Васильева [4] и Д. И. Иванова [5]. Новый взгляд на жёсткую социокультурную детерминацию когнитивной сферы предлагает А. В. Смирнов [6; 7]. Социальный контекст научного познания исследуется в работах по социальной эпистемологии, прежде всего в работах И. Т. Касавина [8]. Концептуальные взаимосвязи теории сетей с синергетикой и квантовой механикой исследуют В. И. Аршинов и В. Г. Буданов [9]. Решительную полемику с социально-конструктивистскими крайностями в эпистемологии и философии сознания ведёт Е. О. Труфанова [10]. Информационная природа психической реальности полнее всего раскрывается в работах Д. И. Дубровского [11]. Заметной русскоязычной работой по философии когнитивных наук (как области, отличной от философии сознания и эпистемологии) в их связи с исследованиями природы социальности стала монография В. А. Бажанова [12]. Важным шагом в акцепции отечественными философами когнитивного энактивизма

стала монография Е. Н. Князевой [13]. Большая работа проделана П. Н. Барышниковым по определению сильных и слабых сторон вычислительного подхода к сознанию в философии и когнитивных науках [14–17]. В последние годы выходили интересные исследования А. Ю. Алексеева в области философских проблем искусственного интеллекта [18] и Д. В. Винника, посвящённые квантовым теориям сознания [19].

Конкретно-научные исследования последних лет в области наук о мозге (К. В. Анохин [20; 21], Т. В. Черниговская [22]), психологии (В. М. Аллахвердов [23], А. И. Назаров [24], В. А. Ключарев [25; 26], М. В. Фаликман [27], И. С. Уточкин [28], Д. В. Люсин [29], В. Ф. Спиридонов [30]), общей теории сетей (А. В. Олескин [31; 32], О. П. Кузнецов [33]) значительно расширили наше понимание когнитивных механизмов человека и некоторых принципов развития сложных, в том числе социальных, систем. К сожалению, новые результаты и концепции концепции пока не привели к созданию эффективной научной методологии, способной интегрировать частные результаты, полученные в рамках отдельных дисциплин, в новаторские междисциплинарные проекты, которые, как показывает мировой опыт, закладывают сегодня фундамент будущей, значительно более эффективной науки.

Представители отдельных дисциплин — нейрофизиологи, генетики, психологи, социологи, экономисты, философы — встречаются друг с другом на совместных научных мероприятиях, говорят о важности междисциплинарной интеграции и комплексных исследований человека, но примеры успешных проектов в этой области редки и малоизвестны. Успехи отдельных исследователей, научных школ и дисциплин можно отнести на счёт таланта и трудолюбия отдельных учёных и коллективов, но вряд ли — на счёт какого-либо методологического прорыва, способного изменить науки о человеке настолько, чтобы они реально могли включиться в общественные и производственные процессы и создавать добавленную стоимость. Не говоря уже о задачах общественной безопасности и совершенствования человеческого капитала.

Между тем, прагматично ориентированные зарубежные коллеги сумели создать междисциплинарные делящиеся проекты, которые уже теперь можно считать новыми комплексными научными дисциплинами. В качестве примеров можно назвать когнитивную социологию, теорию когнитивных социальных сетей (КСС), социальную нейронауку и нейроэкономику.



Когнитивная социология, получившая своё название благодаря одноимённой работе Аарона Сикуреля [34], изучает, с одной стороны, влияние социальных структур и культурных норм на когнитивные процессы, а с другой стороны, обратное влияние когнитивных структур общественных индивидов на их социальную жизнь, параметры сообществ, культурные процессы (здесь можно выделить работы Э. Зерубавеля [35], К. Серуло [36], П. ДиМаджио [37]).

Теория КСС изучает, как изменения в информационных последовательностях, количестве источников информации (узлов в информационной сети) и в видах источников (человеческие или технические) могут влиять на доверие к получаемой информации и на процессы принятия решений в сетевой среде (Р. Швайкерт [38], Р. Брэндс [39], Й. Менгес и М. Килдафф [40]). Здесь используются когнитивные модели для прогнозирования поведения идеального человека-исполнителя, измеряется функционирование реального человека в соотношении с этими идеальными моделями, и в результате определяется, как обратная связь и обучение могут быть использованы для улучшения поведения человека в области принятия решений. Такие исследования финансируются, в частности, Министерством обороны США [41].

Социальная нейронаука (Дж. Касиоппо [42], К. Фрит [43], М. Гаццанига [44], Дж. Риццолатти [45]) концентрируется на роли нейронных и гормональных механизмов в формировании социальных связей и социальных структур. Этот междисциплинарный проект работает на концептуальное взаимообогащение нейрофизиологии, биологии и социологии. Нейроэкономику многие рассматривают как раздел социальной нейронауки, изучающий нейронные и когнитивные механизмы, участвующие в экономическом поведении и экономическом выборе индивидов.

К сожалению, судя в том числе по зарубежным публикациям, вклад философов в новые намечающиеся междисциплинарные направления пока минимален. Ведущие зарубежные философы, занимающиеся когнитивной тематикой (П. и П. Чёрчленд [46], Дж. Сёрл [47], Д. Деннет [48; 49], Д. Чалмерс [50; 51], Э. Кларк [52; 53] и др.) крайне мало внимания уделяют связи когнитивного с социальным и, соответственно, разработке методологий соответствующих междисциплинарных исследований. Между тем, участие философов представляется решающим именно в методологической части междисциплинарных проектов, поскольку речь идёт о взаимодействии предметно и методологически

различных теорий, и принципы такого взаимодействия нуждаются в философском анализе.

В последние десятилетия в мире многократно вырос поток публикаций, посвящённых тематике когнитивных наук. Причём этот комплекс наук пережил последовательно уже три исследовательские программы: *классицизм* (он же *символизм*), ярче всего защищаемый Дж. Фодором и его единомышленниками [54]; *коннекционизм* (Румельхарт, Маклиланд и др. [55; 56]), нашедший горячих сторонников среди элиминативных материалистов (Патрисия и Пол Чёрчленды [57—59]); и, наконец *энактивизм*, тесно связанный с программой *embodied cognition*, запущенной Ф. Варелой [60], и в значительной степени nasledующий философской феноменологии М. Мерло-Понти. В последнее время внимание когнитивных теоретиков в большей степени приковано к многообещающей теоретической платформе известной как *предиктивный процессинг* или *предиктивное кодирование*, которая основана на представлении о когнитивных актах как статистических предсказаниях [61—63].

В то же время, прогресс в области когнитивных наук пока ненамного приблизил решение «вечных» философских проблем, связанных с сознанием, пониманием, субъектом и субъективностью. Дискуссии по этим проблемам кристаллизуются вокруг таких тем, как *qualia* (субъективность феноменального мира и её отношение к мозгу), *self* (осознание себя как «я» и своих ментальных действий как своих сознательных и свободных) и *awareness* (представленность субъекту его собственных ментальных актов).

Предлагаемое исследование направлено на философский анализ эмпирических когнитивных теорий феноменального сознания (*qualia*), осознанности и самосознания (*awareness, consciousness, self*) и соотнесение их с философскими позициями. В процессе исследования используются также достижения и техники философии языка, с тем чтобы определить, какие из философских и научных терминов, относящихся к сознанию, относятся к реальным психическим сущностям, а какие представляют собой языковые («грамматические») конструкции.

В основе исследования лежит следующая **рабочая гипотеза**: *если «когниция»<sup>1</sup> может быть понята как вычислительный процесс той или иной архитектуры, то, учитывая некоторую общность методов когнитивных и социальных дисциплин, общественная жизнь и социальные взаимодействия*

<sup>1</sup>См. определение на с. 306 .

*также могут быть объяснены как вычислительные процессы в самом широком смысле слова. Тогда язык и – шире – символические системы, будучи важной формой социализации индивидов, могут быть поняты как своего рода программный интерфейс между когнитивными и социальными процессами.*

Если эта гипотеза подтверждается в процессе исследования, она вполне может стать точкой междисциплинарного синтеза когнитивных, социальных наук, а также логики, семиотики и лингвистики.

**Целью** данной работы, таким образом, является оценка возможности подтверждения рабочей гипотезы, сформулированной выше.

Для достижения поставленной цели необходимо было решить следующие **задачи**:

- проанализировать основные направления когнитивной науки, их онтологические и концептуальные основы;
- провести анализ онтологической и концептуальной основы социальных наук, чтобы определить возможность использования точных методов и компьютерного моделирования;
- уточнить понимание вычислений, чтобы показать возможность использования этого понятия в науках о природе, познании и обществе;
- идентифицировать концептуально-онтологические пересечения указанных областей знания, определить перспективные точки и методологии их междисциплинарного синтеза.

Я вижу два принципиальных вопроса, на которые должно дать ответ философское исследование проблемы:

- (1) говоря о «когнитивных основаниях», имеем ли мы в виду только ограничения, которые накладывает биологическая и психологическая организация людей на их социализацию, или можно говорить об активном достраивании сложных (биологических) систем самих себя до ещё более высокого (социального) уровня сложности?
- (2) какая онтология биосоциальной реальности оптимальна для сквозного использования формальных языков, средств компьютерного моделирования в ходе междисциплинарного синтеза научных направлений, участвующих в решении этой проблемы?

**Научная новизна:**

1. Выработана комплексная методология междисциплинарной интеграции научных направлений, исследующих когнитивные основания человеческой социальности.
2. Предложена онтология (концептуальная схема), позволяющая применять одни и те же или схожие формальные средства для описания когнитивных и социальных явлений, открывая тем самым путь к существенной интеграции наук о человеке.
3. Показана применимость современного понимания вычислительных процессов в исследовании биологических, когнитивных и социальных систем.
4. В отечественный научный оборот введена проблематика, широко обсуждаемая в многочисленных зарубежных публикациях по философским проблемам когнитивных и социальных наук, но сравнительно мало известная широкому кругу российских исследователей.

**Практическая значимость** В результате исследования должны сложиться условия для ускоренного развития комплекса когнитивных и социальных наук. Позитивное развитие в этом направлении может привести к появлению «единой науки о человеке», идея которой была сформулирована молодым Марксом.

Полученные результаты могут быть использованы во всех направлениях деятельности, связанных с повышением качества человеческого капитала, качества жизни и социальных институтов, а также с повышением общественной безопасности. Идеи и результаты исследования могут быть использованы в преподавании курсов философии, истории и философии науки, психологии и социологических дисциплин.

**Методология и методы исследования.** Современные научные теории, наряду с традиционной формализацией и математизацией, всё шире используют компьютерное моделирование как одно из средств достижения точности и однозначности получаемых знаний. Интересующие нас области — когнитивная наука, социология (в широком смысле слова) и лингвистика — не исключение. Тестирование гипотез на компьютерных моделях, в частности с использованием таких платформ, как ACT-R и CLARION, стало традицией в когнитивных науках. В социологии компьютерные программы уже давно используются для обработки эмпирических данных, но моделирование социальных структур и отношений пришло в научную практику относительно недавно — с появле-

нием технологий мультиагентных программных систем (МАС). Лингвистика выигрывает от того факта, что наиболее популярные области применения искусственного интеллекта — распознавание речи, машинный перевод, боты — относятся к её сфере компетенции. Кроме того, программные разработки используются для реконструкции различных состояний языков и т.п.

Все эти обстоятельства облегчают задачу индивидуального исследователя по поиску общей методологической почвы трёх изначально разнородных научных областей. Таким образом, общий подход предлагаемого исследования состоит в оценке возможностей компьютерного моделирования во всех трёх группах дисциплин, анализе конкретных вычислительных техник и методик, применяемых там, и поиск основы междисциплинарного синтеза в этой области, поиске возможностей их интеграции. Со строго научной точки зрения такой подход выглядит вполне обоснованным, поскольку возможность компьютеризации науки напрямую зависит от степени её формализации и математизации, а также способствует прогрессу обеих тенденций.

Исследование проведено с применением:

- (1) концептуального анализа;
- (2) изучения и философского осмысления результатов конкретных исследований;
- (3) синтеза научных результатов в новых методологических подходах и онтологиях.

#### **Основные положения, выносимые на защиту:**

1. Для междисциплинарного синтеза когнитивных и социальных наук необходима общая онтология, благодаря которой существенная взаимосвязь когнитивных и социальных процессов может быть теоретически продемонстрирована.
2. На уровне онтологии объекты когнитивных и социальных наук лучше всего описываются как сетевые структуры, состоящие из узлов и связей между ними.
3. В нейронных и социальных сетях реализуются параллельные распределённые вычисления, упрощёнными моделями которых можно считать, соответственно, искусственные нейросети и мультиагентные системы.
4. Когнитивные способности, обеспечиваемые телесной организацией индивида, имеют ассоциативно-вероятностный характер, а линейное

вербально-категориальное мышление порождается коммуникативными практиками и оформляется языком.

5. Социальная метасеть представляет собой расширение когнитивных возможностей живых организмов за счёт распределения вычислительных задач между ними.

**Достоверность.** Благодаря поддержке РГНФ (проект №15-03-00417) и РФФИ (проекты 13-06-00878 А и №18-011-00316 А), диссертантом была разработана теоретико-методологическая основа междисциплинарной интеграции когнитивных и социальных наук. Эта рабочая рамка представляет собой синтез нейросетевого подхода в когнитивных науках и сетевой концепции общества. Преимущество предлагаемого подхода, вкратце, состоит в том, что он формирует единую «сетевую» онтологию для когнитивных и социальных наук, что принципиально позволяет использовать один и тот же формальный аппарат, осуществляя «сквозные» социально-когнитивные исследования. Выполнение указанных проектов позволило корректно сформулировать эту концепцию на философско-методологическом уровне. Теперь стоит задача превращения её в работающую общенаучную методологию.

Результаты работ по указанным проектам были позитивно оценены экспертами фондов как находящиеся в соответствии с результатами, полученными другими авторами в России и за рубежом.

**Апробация работы.** Основные результаты работы докладывались на:

- Третьей международной конференции по философии сознания, Университет Минью (Португалия) 12 октября 2017 г.
- теоретическом семинаре в Институте психологии РАН (руководитель — акад. Д. В. Ушаков) 18 ноября 2019 г.
- семинаре «Нейронаука и философия» в Сеченовском университете 21 ноября 2019 г.
- IX Международной конференции по когнитивной науке (Москва) 14 октября 2020 г.

**Публикации.** Основные результаты по теме диссертации изложены в 55 печатных изданиях, 33 из которых изданы в журналах, рекомендованных ВАК, 15 — в периодических научных журналах, индексируемых Web of Science и Scopus.

**Объем и структура работы.** Диссертация состоит из введения, 3 глав и заключения. Полный объём диссертации составляет 340 страниц, включая 4 рисунка и 1 таблицу. Список литературы содержит 324 наименования.

## Глава 1. Философско-методологические основания

### 1.1 О соотношении философского и научного знания

#### 1.1.1 Философия как концептуальный анализ. Наука как порождение гипотез или моделей, соответствующих фактам

Настоящее диссертационное исследование является по преимуществу философским, но, в то же время, оно опирается на большой массив данных и широкий спектр концепций конкретных наук — когнитивных и социальных. Поэтому, в интересах академической строгости, необходимы ясные критерии различения методов и результатов, относящихся к этим различным формам познания. Для целей настоящего исследования научными (относящимися к науке) считаются теории и концепции, чья истинность существенно зависит от эмпирических данных, по крайней мере, в негативном смысле: т. е., для которых понятно, какие результаты наблюдения или эксперимента могли бы сделать данную теорию или гипотезу ложной. К сфере философии в контексте данной работы относится концептуальный анализ возможных онтологических конструкций и/или то, что можно было бы назвать метатеориями — исследования языка, логики и методов построения научных теорий. Никакая теория или исследовательская программа, претендующая на описание положений дел в мире — абстрактных или конкретных, локальных или универсальных, — но не предлагающая гарантий собственной эмпирической фальсифицируемости, не рассматривается здесь как содержащая знание в собственном смысле слова (аргументы в пользу этой позиции приводятся в [64]).

Возможно возражение: если философия занимается онтологиями в том или ином виде, то она описывает положения дел в мире, при том что её теории эмпирически не фальсифицируемы. И тогда мы должны отказать ей в каком бы то ни было эпистемическом статусе. Для того, чтобы снять это возражение, необходимо разобраться в статусе научных онтологий.



В общем виде научная деятельность представляет собой порождение правдоподобных гипотез, вывод из них эмпирических следствий и сопоставление последних с фактами при строгом соблюдении принципа фальсифицируемости: мы имеем право утверждать, что имеющиеся факты подтверждают истинность нашей теории, только если мы точно представляем себе, какие факты могли бы сделать её ложной. Философские теории такой проверки пройти не могут, поскольку, если бы могли, они были бы научными теориями.

### 1.1.2 Онтологические презумпции

Когда мы пытаемся узнать что-то о мире при помощи опыта и рассуждения, первый даёт нам новое, но ненадёжное знание, а второе само по себе способно сообщить с достоверностью только нечто и так известное. Эта проблема лежала в основе философских дискуссий XVII – XVIII вв., но не нашла удовлетворительного решения, поскольку эмпиризм не мог совладать с проблемой аподиктичности научного знания, а рационализм мог предложить только псевдорешение, связанное с признанием «самоочевидных» содержательных предпосылок априори. Логично предположить, что должен быть третий элемент, или, возможно, источник научного знания, сообщающий ему искомую аподиктичность. Кант обнаружил таковой в априорных формах чувственности и рассудка, но это решение плохо смотрелось на фоне последовавших научных революций: появления неэвклидовых геометрий, неаристотелевских логик и неклассических теорий в физике. Можно, конечно, спорить, насколько изменившиеся толкования фундаментальных научных понятий действительно противоречат трансцендентальной установке как таковой, но в целом трансцендентализм остался в стороне от мейнстрима философии науки.

Другое решение вышло в рамках аналитической философской традиции, и связано оно с именами Витгенштейна и Куайна. Так, Витгенштейн обнаруживает такую внутреннюю структурность языка, которая выходит за пределы логики и демонстрирует его онтологическую нагруженность.

Об этом идёт речь в его полемических заметках против «доказательства существования внешнего мира» Дж. Мура, изданные под объединяющим названием «О достоверности» [65]. Аргументация Мура основана на рассуждении,

что есть высказывания, истинность которых зависит от объективного существования внешнего мира, и они несомненно истинны, например «Это – моя рука». Витгенштейн возражает, что если отрицания некоторых фактуальных по видимости высказываний непредставимы в рамках данной картины мира, то они сами истинны не в том же самом смысле, в котором истинны собственно фактуальные высказывания («На таком-то расстоянии от Земли находится другая планета»): «Истинность определённых эмпирических предложений относится к нашей системе референций» [65, с. 333]. Витгенштейн поясняет свою мысль: сомнение, вопрошание, выражение потребности в дополнительном знании, есть языковая игра (в том смысле, в котором этот термин определён в его же «Философских исследованиях» [66, с. 83]), которая возможна только при условии наличия чего-то несомненного. «Попробуй я усомниться в том, что Земля существовала задолго до моего рождения, мне пришлось бы усомниться во всем, что для меня несомненно» [65, с. 350].

Форма или видимый характер высказывания может вводить в заблуждение: «я склонен думать, что не всё, чему присуща форма эмпирического высказывания, является эмпирическим высказыванием» [65, с. 358]. Эмпирические (фактуальные) высказывания в строгом смысле – это такие высказывания, истинностные значения которых определяются соответствующими положениями дел в мире. Их отрицание само по себе не составляет проблемы в рамках логического атомизма: «Любой факт может иметь место или не иметь места, а все остальное останется тем же самым» [67, с. 5]<sup>1</sup>. И очевидно, что любое движение мысли в сторону того, что истинность одних элементарных предложений может как-то зависеть от истинности (несомненности) других, Витгенштейну периода «Трактата» показалось бы ересью. Есть веские основания считать, что, отказавшись от логического атомизма, он пришёл к некоему варианту лингвистического холизма, хотя и отличного от холизма де Соссюра. Холизм последнего состоит в том, что язык как целое определяет свои части, тогда как у Витгенштейна язык в виде причудливо переплетенных языковых игр есть часть целостного жизненного мира *homo sapiens*. Мы не в состоянии отрицать некоторые высказывания не потому, что этого не позволяют сделать формальные правила поверхностной грамматики или логики, а потому что тогда рухнет вся система нашего взаимопонимания, и становится невозможна коммуникация – цель и оправдание существования языка.

---

<sup>1</sup>Перевод уточнён мною – И. М.

Я бы предложил называть обнаруженные Витгенштейном квазиэмпирические высказывания, смысл которых не может быть фальсифицирован в рамках данного жизненного горизонта, онтологическими презумпциями. Сам он в нескольких местах цитируемой работы говорит о «допущениях» (*assumptions* в английском варианте текста), но допущения мы вольны делать или не делать, а речь на самом деле идёт об утверждениях, которые не являются предметом нашего индивидуального произвольного решения (хотя в каком-то – гносеологическом – смысле они могут быть произвольными).

Итак, наличие онтологических презумпций в языке свидетельствует в пользу (не столько лингвистического, сколько социологического) холизма. И объясняется эта позиция коммуникативной природой языка. Язык «Трактата» мог быть построен кем угодно: Богом, суперкомпьютером и т. п. Его предложения изоморфны фактам, именно поэтому они могут сообщать мысли, т. е. участвовать в коммуникации, но для самой означающей функции этого языка наличие или отсутствие коммуникационного процесса полностью irrelevantно. Напротив, язык, описываемый в поздних работах Витгенштейна, существует только в правилосообразном взаимодействии людей, а сам феномен действия в соответствии с правилом возможен, только когда люди опутаны и спаяны целой сетью конвенций.

Вот пассаж, который лучше всего выражает то, как мы реально получаем знания и оперируем с ними:

И то же самое было бы, если бы ученик усомнился в единообразии природы, а значит, и в оправданности индуктивных выводов. — Учитель почувствовал бы, что такое сомнение лишь задерживает их, что из-за этого учеба только застопоривается и не продвигается. — И он был бы прав. Это походило бы на то, как кто-то искал бы в комнате какой-то предмет так: выдвигая ящик и не находя искомого, он бы снова его закрывал и, подождав, опять открывал, чтобы посмотреть, не появилось ли там что-нибудь, и продолжал в том же духе. Он еще не научился искать. Вот так и ученик еще не научился задавать вопросы. Не научился той игре, которой мы хотим его обучить [65, с. 359].

Система скрытых и явных онтологических презумпций образует картину мира. Собственно, картина мира и внутренняя, глубинная грамматика языка —

это две стороны одной медали. И здесь мы – вполне предсказуемо – упираемся в культуру как механизм социального наследования, культуру, которую Ю. М. Лотман называл «коллективным интеллектом» и «коллективной памятью» человеческих сообществ. Витгенштейн поясняет: «я обрел свою картину мира не путем подтверждений ее правильности, и придерживаюсь этой картины я тоже не потому, что убедился в ее корректности<sup>2</sup>. Вовсе нет: это унаследованный опыт, отталкиваясь от которого я различаю истинное и ложное» [65, с. 335].

Если не тождественные, то схожие выводы делает Уиллард Куайн на основе более формальных исследований. Так, говоря о своей знаменитой формуле «существовать — значит быть значением квантифицированной переменной», он подчёркивает:

...эта формула служит скорее критерием согласованности данного замечания или доктрины с *априорным онтологическим стандартом* [курсив мой — *И. М.*]. Мы рассматриваем связанные переменные в связи с онтологией не для того, чтобы знать, что есть, но для того, чтобы знать, что данное замечание или доктрина, наша или чья-то еще, говорит, что же есть; а это, собственно, проблема, затрагивающая язык [68, с. 20].

Аргументы в пользу выделения онтологий в качестве независимых элементов структуры знаний будут рассмотрены в следующем параграфе. Пока заметим, что дальнейшие рассуждения будут основываться на концепции онтологических презумпций — опорных семантических конструкций языка, которые могут проявляться в форме квазиэмпирических высказываний, истинность которых «очевидна», а отрицание не ложно, а бессмысленно. Если принять эту концепцию, то правильная философская стратегия должна состоять в концептуальном прояснении вопроса: что может заставлять язык выстраивать свой внутренний семантический каркас именно таким образом?

---

<sup>2</sup>Положение об особых условиях истинности онтологических высказываний, в отличие от высказываний собственно эмпирических, крайне важно для моих дальнейших рассуждений.

### 1.1.3 Понятие научных онтологий

Традиционно считается — и преподаётся во всех соответствующих курсах, — что научное знание имеет два уровня: теоретический (где гипотезы становятся теориями) и эмпирический (где добываются факты, обеспечивающие эти превращения). Но, на мой взгляд, можно и нужно говорить ещё об одном уровне. Как теоретические, так и эмпирические предложения любой науки формулируются в специфических терминах, которые семантически интерпретируются на определённой предметной области, относясь к её объектам, свойствам и отношениям. Т. е., в основе любой теории лежат представления о том, какие объекты здесь имеют место быть, какими свойствами они в принципе могут обладать и в какие отношения они могут вступать. В современных науках о структурах данных и представлении знаний такие концептуальные схемы принято называть онтологиями [69]. Поэтому определяемый мною третий уровень научного знания я буду называть уровнем научных онтологий.

По умолчанию считается, что учёный получает эти представления непосредственно из наблюдения. Однако это не совсем так. Куайн в уже цитированной классической работе полагает, что принятие той или иной научной онтологии

в принципе сходно с тем, как мы принимаем научную теорию, скажем, систему физики. Мы адаптируем, по крайней мере в той степени, в какой способны понять, простейшую концептуальную схему, в которой разрозненные фрагменты сырого опыта могут быть согласованы и упорядочены. Наша онтология определяется, как только мы зафиксировали общую концептуальную схему, которая должна обеспечивать науку в самом широком смысле <...> В той степени, в какой о принятии любой системы научной теории можно говорить как о проблеме *принятия языка* [курсив мой — И. М.], в той же степени — но не больше — можно говорить о принятии онтологии [68, с. 21].

Онтология, таким образом, может пониматься как стандартная (но не обязательно единственно возможная) семантика языка теории. Отнесение многообразных чувственных впечатлений к постулируемым объектам «упрощает

наше описание опыта». «Физические объекты — это постулируемые сущности, которые упрощают и завершают наше описание потока опыта, так же как введение иррациональных чисел упрощает законы арифметики» [68, с. 22]. В каком-то смысле она является полезным мифом.

Нет принципиальной проблемы в том, чтобы построить формальную модель, в которой значения переменных в некоторых формулах будут определяться вводимыми данными, и от этих значений будет зависеть истинность некоторых других формул. Но человеку-исследователю важно понимать, к какой части его *Umwelt*'а относится данная теория, и, следовательно, где именно должен он искать данные для ввода. Для этого и теоретические положения, и предложения наблюдения, и сами данные должны быть представлены в терминах определённой концептуальной схемы. И такие схемы в науке творчески выбираются или порождаются самими учёными. Удачный или неудачный выбор концептуальной схемы предопределяет успех или неуспех научной теории не в меньшей степени, чем выбор формального аппарата. Свободным и творческим решением Исаака Ньютона было принято, что физический мир состоит из физических тел, которые из интересных свойств обладают только массой, а также способностью двигаться прямолинейно и равномерно, пока на них не действуют какие-либо силы. Удачно найденная концептуальная схема позволила непротиворечивым образом применить математику как к формулируемым гипотезам, так и к данным опыта, и в результате человечество получило одну из самых успешных в его истории научных теорий, практические выходы которой существенным образом изменили лицо планеты.

Если Витгенштейн рассуждает об онтологических презумпциях естественного языка, то Куайн формулирует достаточно чёткий критерий («стандарт», в его терминологии) того, что несомненно относится к онтологическим обязательствам научной теории:

Теория обязывает к тем и только к тем сущностям, на которые должны быть способны указывать связанные переменные этой теории, для того чтобы сделанные в ней утверждения были истинными [68, с. 18].

Для переменной быть связанной — значит стоять под квантором «все» или «некоторые». Как разъясняет Куайн, «<с> точки зрения категорий традиционной грамматики этот подход, грубо говоря, означает, что быть — значит

попадать в область референции местоимения» [68, с. 17]. Таким образом, местоимения, выражающие квантификацию — «все», «некоторые», «никакие», — в истинных предложениях теории стоят перед именами, которые указывают на сущности, признаваемые данной теорией реальными. Но сами эти сущности выполняют функцию упрощения описания «потока опыта», их выбор — это выбор той или иной стратегии экономии описательных средств, и в некоторых случаях различные онтологические обязательства могут взаимно конвертироваться без ущерба для истинностных значений использующих их теоретических высказываний.

Важно отметить следующее:

1. Научные онтологии логически независимы от теоретического и эмпирического уровня науки, и их выбор является творческим актом исследователя — автора научной теории.
2. Любая научная онтология служит общей областью интерпретации теоретических и эмпирических предложений данной науки.
3. «Предложения» онтологии (например, «даны объекты  $O$  и  $P$ ») не могут быть истинными или ложными, в отличие от собственных предложений теории («из наличия свойства  $A$  у объекта  $O$  следует наличие свойства  $B$  у объекта  $P$ »). Они могут только оцениваться на корректность (и здесь может помочь философия) и на эффективность — насколько они способствуют хорошей работе теории: не затерялся ли в их составе, например, какой-нибудь излишний «флогистон».

Первый пункт требует следующих пояснений. Предположим, создаётся некая социальная теория. Её автор говорит, например: в моей теории речь будет идти о личностях и социальных институтах. Личности будут обладать такими свойствами, как пол, возраст и набор социальных ролей. Социальные институты понимаются как устойчивые системы социальных ролей, реализуемых внутри и вовне института. Институты характеризуются (историческим) возрастом, степенью проникновения в общество и способом организации <sup>3</sup>...

Всё до сих пор произнесённое относится к научной онтологии, а не к теории, поскольку здесь вводятся типовые объекты, их возможные свойства и возможные отношения между ними, и ничего пока не утверждается о том, как обстоят дела в реальности. Здесь только задаётся структура предметной области, к которой будут относиться (на которой будут интерпретироваться)

<sup>3</sup>Подобную социальную онтологию предлагает Тони Лоусон (см. с. 30).

будущие теоретические и эмпирические положения. Но когда будет сказано что-нибудь вроде: «Личности, связанные институтом брака, живут в среднем на 15% дольше» (в нашей модели это можно выразить, например, через сравнительную долю состоящих в браке среди великовозрастных личностей), исследователь приступит к формулированию собственно положений теории. Это и подобные высказывания можно выразить на естественном языке — но тогда, действительно, трудно отличить их от высказываний онтологии по чисто формальным признакам. Но их можно выразить также и средствами какого-либо формального языка. Тогда принадлежность высказывания к той или иной эпистемической категории станет более явной, поскольку положения онтологии будут формулироваться на некотором метаязыке. Очевидно, что между высказываниями в языке теории и высказываниями в метаязыке не может быть логических связей. В языке теории мы формулируем общие теоретические положения, выводим из них эмпирически значимые следствия, устанавливаем истинность или ложность последних путём наблюдения и транслируем их истинностные значения обратно, в сторону теоретических обобщений, насколько это позволяют сделать логические правила. Язык теоретических законов и язык описания эмпирических фактов — это один и тот же язык, что именно и позволяет совершать логический вывод «сверху вниз» и транслировать обоснование «снизу вверх». Следовательно, суждения онтологии — онтологические презумпции [70, с. 16–19] — логически независимы от теоретических и эмпирических высказываний в рамках одной и той же научной теории.

Однако Куайн прямо утверждает наличие логических связей между онтологическими и дескриптивными высказываниями: «Онтологические высказывания непосредственно следуют из всего комплекса случайных высказываний о заурядном факте» [68, с. 15]. Моё возражение сводится к следующему. Онтологические допущения как пресуппозиции, содержащиеся в дескриптивных высказываниях, содержатся также и в отрицаниях этих высказываний. Высказывание «Теплород распределяется равномерно по всему объёму физического тела» предполагает существование теплорода. Однако его отрицание может быть интерпретировано двояко: «Теплород распределяется по объёму физического тела неравномерно» или «Теплорода не существует». Поскольку, как показал Рассел [71, с. 19], отрицательное высказывание о несуществующем предмете может быть истинным, неверно считать, что истинность онтологических высказываний следует «из всего комплекса» истинных дескриптивных выска-



званий о соответствующих сущностях. Как метафорически выразил эту идею Гидеон Розен, «<с>пекулятивная онтология — это палеонтология без окаменелостей» [72, р. 72]

Относительно второго пункта: может оказаться, что формула, выражающая теоретический закон, применима и к другой предметной области: например, выяснится, что, если заменить личностей на львов, а институт брака на львиный прайд, формула останется истинной. Тогда ранее описанная предметная область окажется одной из возможных моделей, на которых может быть интерпретирована теория. Концепция моделей, разработанная Тарским, является важной частью семантики логики предикатов [73, с. 416–417]. Я понимаю научные онтологии в очень похожем смысле.

Относительно третьего пункта: часто, к сожалению, все высказывания о существовании / несуществовании чего бы то ни было рассматриваются как часть мироописания и на этом основании оцениваются как истинные или ложные. Это происходит потому, что высказывания типа «Пегасов не существует» и «Эфира не существует» имеют одинаковую грамматику в естественном языке. Однако, как уже отмечено в [74, с. 117], высказывание о Пегасах относится к экземплярам типа и обладает эмпирической истинностью. Поэтому, конечно, нет никакого противоречия между «Пегас есть лошадь с крыльями» и «Пегасов нет на свете»: первое предложение характеризует тип, второе — его экземпляры. Высказывание об эфире относится, напротив, к типу и означает буквально следующее: теории, обходящиеся без эфира как объекта своей онтологии, работают лучше, чем теории, исходящие из его существования. Приемлемость и признанность этого высказывания научным сообществом носит, на самом деле, конвенциональный характер.

У философии нет экспериментальных средств, чтобы проверять выводы рассудка, у неё также нет математической компетенции, чтобы участвовать в строительстве теории. Но, поскольку наука не исчерпывается эмпирическим и теоретическим уровнем, а опирается ещё и на независимо от них выработанную онтологию своей предметной области, философия может предложить методы и эвристики для создания концептуальных схем, т. е. объектных моделей для интерпретации предложений науки. У неё за плечами более чем двухтысячелетний опыт выполнения работы. Смысл этой деятельности не всегда понимался философами адекватно: когда других наук не было, они считали, что непосредственно открывают истину при помощи исключительно рассужде-

ния. Затем, когда науки появились, они долгое время полагали, что за ними остаётся метафизика, учение о «началах» сущего, важной частью которой должна быть общая онтология — рационально обоснованный ответ на вопрос «что есть?», как его впоследствии сформулировал Куайн. Далее последовал непродолжительный период почти всеобщего отказа философов от метафизики, а следовательно, и от онтологии.

Сегодня, когда метафизика возвращается даже в штаб-квартиры аналитической философии, вновь встают вопросы об уместности философской онтологии, онтологической относительности, онтологическом плюрализме и т. п. (см. об этом [75]). В рамках моей концепции философская онтология нерелевантна. В обоснование приведу следующее рассуждение. Онтология есть концептуальная модель системы объектов, на которой интерпретируется научная теория. Если есть некая философская онтология, то на ней должна интерпретироваться какая-то теория. Если это научная теория, то что делает такую онтологию философской? А неуместность философских объектных «теорий» следует из отсутствия у философии механизмов эмпирической фальсификации её предложений. Онтология же вне какой бы то ни было теории является исключительно предположением, какие объекты с какими свойствами могли бы существовать, чтобы быть предметом осмысленных высказываний. В любом случае, предполагаемая философская онтология повисает в воздухе, лишённая каких бы то ни было семантических и прагматических опор.

И тем не менее, основная сфера приложения философии лежит именно в сфере онтологии. Как это возможно?

Онтология, в моём понимании, это не описание мира как он есть, а описание мира каким он мог бы быть при условии истинности соответствующей научной теории, т. е. в ситуации, когда эмпирические следствия теории подтверждаются фактами при интерпретации её на данной онтологической модели. Излишне говорить, что истинность теории не влечёт с необходимостью адекватность её онтологии миру как он есть, хотя бы потому, что моделей для интерпретации данной теории всегда может быть больше, чем одна (считается, что некоторые работающие и сегодня законы термодинамики были сформулированы ещё для теплорода). И спектр философских позиций может простираться здесь от наивного реализма — мир именно таков, каким его представляет онтология нашей истинной теории — до крайнего скептицизма и релятивизма, оснований для которых предостаточно. И здесь для философии есть две ра-

боты: одна трудная, другая полегче. Первая — её, перефразируя Чалмерса, я назвал бы трудной проблемой метаонтологии — состоит в поисках «трансцендентальных» оснований научных онтологий, в попытках определить, какое отношение могут иметь онтологии успешных научных теорий к подлинной природе мира. И вторая — лёгкая проблема метаонтологии — состоит в выработке общезначимых методологий и эвристик «строительства» научных онтологий, а также рациональной критики уже построенных. Это как раз та работа, которая реально помогла бы науке, ускоряя порождение новых теорий и делая более эффективными уже существующие.

Термин «метаонтология» не является моим изобретением. Если верить авторам книги [76], первенство принадлежит Петеру ван Инвагену [77], по мнению которого, если главный вопрос онтологии, согласно Куайну, это «что есть?», то метаонтология отвечает на вопросы «что мы имеем в виду, спрашивая, что есть?» и «какова корректная методология онтологии?». В уже упоминавшейся статье [74] я также ввожу этот термин, приводя примеры метаонтологических рассуждений в «Метафизике» Аристотеля и «Логико-философском трактате» Витгенштейна, и даю определение метаонтологии как философской квази-теории, трактующей возможность объектов и субстанции как формы данности мира.

Логические отношения между теоретическими и эмпирическими предложениями, как уже говорилось, могут быть вычислены в том числе искусственным интеллектом. Объектная интерпретация, напротив, является средством «одомашнивания» научной теории в рамках «жизненного мира» человека. Однако создание онтологий, которое иногда выглядит как концептуальный анализ высокой степени абстракции, имеет за собой многовековую традицию: от размышлений Парменида и Демокрита, диалогов Платона и метафизики Аристотеля, через схоластику — к «Логико-философскому трактату» и аналитической метафизике. Таким образом, философский аспект настоящего исследования состоит в анализе концептуальных схем — онтологий — когнитивных и социальных теорий и рациональной критики их с целью поиска путей их возможной оптимизации.

Отдельно стоит прояснить вопрос о том, как соотносится так определённое понятие научных онтологий с тем, что В. С. Стёпин называл научной картиной мира, считая её одним из трёх оснований науки [78, с. 191–204], [79, с. 188–230]. Особенно учитывая, что понятие научных онтологий также при-

существовало в его работах [80, с. 165–184]. В книге «Цивилизация и культура» он прямо отождествляет научные онтологии и научные картины мира, выделяя дисциплинарные онтологии, обобщающие образы природы и общества, и общенаучную картину мира [80, с. 165]. В более специальной работе «Теоретическое знание» В. С. Стёпин утверждает, что научная картина мира, упрощая реальный мир до его существенных связей, фиксирует последние «в виде системы научных принципов, на которые опирается исследование и которые позволяют ему активно конструировать конкретные теоретические модели, объяснять и предсказывать эмпирические факты» [79, с. 189]. Здесь же он утверждает, что научные онтологии «являются особым типом научного теоретического знания» и функционируют как «специфическая форма систематизации научного знания» или система представлений о мире, «которая вырабатывается в результате синтеза знаний, получаемых в различных областях научного исследования» [79, с. 192]. Научная картина мира, синонимичная научной онтологии, «принадлежит и к внутренней структуре науки, репрезентированной взаимосвязями между эмпирическим и теоретическим знанием» [79, с. 192].

Эта концепция вызывает у меня следующие сомнения. Во-первых, если научная картина мира предстаёт как система научных принципов (к которым, наверное, относятся принцип причинности, принцип сохранения и т. п.), то непонятно, каким образом такая система может служить основой конструирования теоретических моделей и предсказания эмпирических фактов. Такие принципы представляют собой скорее ограничения или запреты на определённые типы объяснений: ничто не есть без причины, ничто не возникает из ничего, ничто не действует на другое вне непосредственного соприкосновения и т. п. Очевидно, что общеотрицательные суждения не могут составлять порождающую модель для конкретных объяснений. Да, из того, что ничто не происходит без причины, следует, что все события имеют причину. Но из этой позиции невозможно вывести, какие события вообще могут происходить, и какие типы причин соответствуют определённым типам событий. Этот же аргумент справедлив для других научных принципов. Во-вторых, утверждая, что научные онтологии одновременно являются особым типом теоретического знания и специфической формой систематизации научного знания, автор, на мой взгляд допускает противоречие. Следуя принципу предметности в логике, мы не можем предполагать, что некое высказывание о предмете может содержать также способ собствен-

ной систематизации, истинностной оценки и т. п., поскольку всё это должно относиться к метаязыку или метатеории.

Третью позицию рассмотрим чуть подробнее. Она предполагает, что внутреннюю структуру науки составляют взаимосвязи между эмпирическими и теоретическими положениями. Т. е., в качестве таковой структуры мы имеем (1) теоретические высказывания научной дисциплины, (2) её же эмпирические высказывания и (3) взаимосвязи между ними. Научная картина мира, согласно В. С. Стёпину, «принадлежит» этой структуре, т. е. является её частью. Очевидно, что (3) представляют собой логические (гипотетико-дедуктивные) взаимосвязи, а логика, если и может претендовать на статус какой-либо картины мира, то во всяком случае не специфически научной. Вряд ли также научную картину мира можно отнести к (2), поскольку эмпирические высказывания не всегда имеют собственно научный характер — они могут быть частью обыденного знания. Кроме того, В. С. Стёпин всегда подчёркивает интегративный, синтезирующий характер научных онтологий, что не может относиться к эмпирическим фактам. Следовательно, единственно возможная локализация научной картины мира в концепции В. С. Стёпина — это теоретическая часть научных дисциплин, что и подтверждается его характеристикой их как «особого типа научного теоретического знания». Таким образом, единственно возможный вывод из всех приведённых положений автора состоит в том, что научные картины мира (научные онтологии) суть некоторый особый тип теоретических высказываний науки. Среди этих положений есть только одно, которое даёт намёк на то, в чём состоит их типологическое отличие. Это положение об их роли в синтезе теоретического знания, которое вырабатывается в различных областях науки. Таким образом, можно предположить, что научные картины мира, по В. С. Стёпину, состоят из междисциплинарных теоретических высказываний. И тогда встаёт вопрос: возможны ли теоретические высказывания, не относящиеся к теоретическим корпусам сложившихся дисциплин?

Как мне представляется, намёк на ответ содержится в статье Э. М. Мирского в «Новой философской энциклопедии». По его мнению, сложившаяся практика междисциплинарных исследований свидетельствует, что ключевую роль играет их методологическое обеспечение, «которое предполагает создание предметной конструкции, *функционально аналогичной предметной конструкции дисциплины* [курсив мой — И. М.]» [81]. Т. е., если речь идёт не о «мягких» междисциплинарных проектах, где встречаются эксперты для временного об-

суждения проблем, а о создании устойчивых эпистемических конструкций для их решения, их методологические основания не могут не принимать квазидисциплинарные формы. Мнение автора, занимающегося проблематикой междисциплинарных исследований минимум с 1980 г., имеет ценность в моих глазах, несмотря на то, что, по свидетельству И. Т. Касавина, ему противостоит подход «международных веб-сообществ, занимающихся широким кругом проблем, в основном в области гуманитарных и общественных наук», которые усматривают необходимость «выходить за пределы рассмотрения науки как чисто теоретической деятельности, как языковой игры» [82, с. 15–16]. Собственно, если не считать постмодернистские «дискурсы» серьёзной философией науки, то очевидно, что рамки *дисциплины* задаются жёсткой взаимозависимостью между дедуктивной структурой формул некоторого теоретического языка, с одной стороны, и объясняемыми ею эмпирическими фактами, с другой. Поэтому тавтологичным выглядит предположение, что если теоретическое высказывание связано жёсткими логическими связями с эмпирически истинными высказываниями — т. е. является научным в собственном смысле слова, — то оно или относится к существующей научной дисциплине, или закладывает фундамент новой.

Таким образом, единственная рациональная интерпретация сложносоставной характеристики научных онтологий, данной В. С. Стёпиным, состоит в том, что их стоит понимать как разновидность или составную часть научных теорий. Этот подход довольно детально разрабатывается Тони Лоусоном, идеи которого относительно социальной онтологии мне уже приходилось критически анализировать [83]. Так, Лоусон определяет научную онтологию как «отрасль исследовательской деятельности, связанную с тем, что есть или что существует, которая исследует природу конкретных сущностей» и которая «если и не сводима к науке как таковой, часто осуществляется в её рамках» [84, р. 20]. Здесь Лоусон смешивает две важные, но принципиально различные вещи: исследование объектов, которые действительно составляют исследуемую область, и попытку ответить на вопрос: что может быть принято в качестве объекта конкретной науки. В первом случае выводы исследования могут быть истинными или ложными, тогда как во втором они конвенциональны. Предложения первого типа относятся к научной теории, а суждения о возможных объектах исследования, собственно, и составляют *онтологию конкретной науки*. В последнем случае исследователь задаётся вопросом, подходит ли определенный тип объ-

ектов для конкретного научного описания. В первом случае он эмпирически определяет, заполнен ли этот тип какими-либо реальными экземплярами. С эпистемологической точки зрения эти положения настолько различны, насколько это возможно. Если бы онтология сама по себе была своего рода «теорией», тогда философия, которая строит онтологии уже 2500 лет, могла бы напрямую конкурировать с наукой. Но на самом деле, если случайно изобретённые философами онтологии и имеют какой-либо смысл в своём существовании, он состоит в том, что, возможно, когда-нибудь их используют для интерпретации собственно теоретических и / или фактических высказываний.

Приведённое соображение показывает, на мой взгляд, неоперациональность понятия научной онтологии как разновидности теории: как проговаривается Лоусон в приведённой цитате, в таком виде оно может быть неотличимо от самой науки. Однако В. С. Стёпин в более ранней работе даёт, как мне кажется, более точную характеристику специальной (дисциплинарной) научной картины мира (научной онтологии) как «особой подсистемы идеальных объектов», которая «вводит представления о главных системно-структурных характеристиках предмета соответствующей науки. Отображение на нее как эмпирических, так и теоретических схем обеспечивает связь представленных в этих схемах различных образов реальности и их отнесение к единой предметной области» [85, с. 49]. Это определение я готов принять. Но остаётся эпистемологический вопрос об источнике «легитимности» онтологических высказываний. И, характеризуя пост-максвелловскую электродинамическую картину мира, В. С. Стёпин утверждает: «Эту картину можно рассматривать в качестве предельно обобщённой модели тех природных объектов и процессов, которые были предметом физического исследования в последней трети XIX в.» [85, с. 50]. Это положение может быть интерпретировано в духе концепции Лоусона: научные онтологии как часть или результат исследовательской деятельности. Слабые стороны этого подхода рассмотрены выше.

Таким образом, принципиальное отличие моего подхода от подходов как Тони Лоусона, так и В. С. Стёпина, заключается в следующем. Конечно, в реальных исторических обстоятельствах определённые онтологии могут быть навеяны прошлыми успешными теориями, даже в других областях науки, общими мифами и повседневным опытом, эстетическими аналогиями или даже эзотерическими доктринами, к которым исследователь неравнодушен лично. Но на нормативно-логическом уровне онтология не выводится из теории, а ско-

рее предписывается ей для того, чтобы служить моделью ее интерпретации. Их «брак» может быть удачным или неудачным, но ни в том, ни в другом случае нет оснований для приписывания онтологическим высказываниям (презумпциям, обязательствам) значений истины или лжи.

#### 1.1.4 Вывод из раздела 1.1

Наряду с теоретическим и эмпирическим уровнями в структуре научного знания необходимо признать наличие **онтологий — концептуальных схем, описывающих типы объектов предметной области, их возможные свойства и отношения**. Онтологии не выводимы и не зависимы ни от положений теории, ни от эмпирических данных. Напротив, предложения теории и описания фактов интерпретируются на предметной онтологии как на модели. Подбор подходящей онтологии есть творческий акт теоретика.

## 1.2 Общая онтология когнитивных и социальных наук

### 1.2.1 Необходимость анализа онтологий когнитивных и социальных теорий

В современном состоянии когнитивных (в широком смысле) и социальных наук есть схожие черты, к каковым я бы отнёс:

- (1) доктринальную разобщённость — представители различных школ и направлений не соглашаются друг с другом не только в решении общепризнанных проблем, что нормально, но и относительно самого состава этих проблем, равно как и в определении целей и задач своих наук;
- (2) наличие нефальсифицируемых теорий — очень часто предпочтение отдаётся впечатляюще сформулированным теориям, чудодейственным мето-



дам и универсально применимым интерпретациям, которые объясняют всё, а следовательно — ничего;

- (3) низкую технологическую эффективность, которая, я полагаю, не требует пояснений.

Вместе с тем, наши интуиции подсказывают, что речь идёт об одной и той же реальности — о человеке как животном разумном и животном политическом (общественном), — признаках, которые, как написал один шутник в Твиттере, к сожалению, не проявляются одновременно. Необходимо методологическое исследование, которое, во-первых, помогло бы преодолеть указанные недостатки этих групп наук, а во-вторых, ответило бы на вопрос: нет ли между ними более глубокой общности, чем мы до сих пор подозревали.

Такое исследование должно ответить на следующие вопросы:

1. Чем вызваны указанные недостатки социальных и когнитивных наук, нет ли у них общей причины?
2. Если, по крайней мере, часть причин общая для обеих групп наук, то каким могло бы быть общее решение?
3. Если есть общее решение для обеих групп наук, то есть ли основания — и какие, если есть — для их более глубокой интеграции?

Последовательный поиск ответов на поставленные вопросы стоит, на мой взгляд, начать с некоторых общих соображений о структуре научных теорий.

Если говорить о технологической эффективности, о способности производить добавленную стоимость и вносить реальную и осязаемую лепту в общественный прогресс, мало кто сможет аргументированно усомниться в том, что в этом отношении естествознание оставляет общественные науки далеко позади. Причём эта ситуация сохраняется с эпохи научной революции XVII–XVIII вв. Основываясь на исторических данных можно предположить, что технологическая эффективность науки как-то связана с её математизацией и — шире — формализацией. Если проанализировать ценности и методы, составляющие то, что мы называем «научностью», мы увидим, что их сущность и «энтелехия» состоит, помимо прочего, в создании простой и непротиворечивой онтологии, доступной для формализации и, желательной, математизации. Логические и/или математические формализмы, непротиворечивым образом интерпретируемые на некоторой корректно построенной предметной области, представляют собой

мощный инструмент объяснения и прогнозирования. И, если такая теория состоялась, дело остаётся за малым – чтобы её выводы и прогнозы соответствовали фактам. Из всех предложенных вариантов теорий выбирается тот, который лучше работает в этом направлении. Европейская наука встала на путь ускоренного развития именно после того, как такая эффективно формализуемая онтология, доступная для формализации и математизации, была предложена Ньютоном для объяснения механических движений. Социальная наука пока не достигла этого уровня, чем и вызвана необходимость анализа онтологических оснований социальной науки.

Согласно классическому подходу, конкурирующие теории должны базироваться на единственной «правильной» онтологии. Постклассический подход предполагает, что онтологии столь же множественны, как и теории, что находит выражение, например, в неклассической физике и компьютерных науках.

### 1.2.2 Понятие метаонтологии

Аристотель считал, что первые сущности могут обладать качествами, противоположными качествам других сущностей, но сами они не могут быть противоположностями других первых сущностей [86, XVI 1 1087b]. Это положение следует из общего определения первой сущности, как того, что не сказывается ни о чём другом [87, 5 2a 10-15], поскольку противоположные качества могут содержаться только в предикатах. Говоря о «противоположностях», Аристотель скорее имеет в виду соизмеримые признаки: большее – меньшее, старое – молодое и т.п. Т.е., подлинные индивиды как они есть, без определений, соотносятся друг с другом только на том основании, что они различны – иначе говоря, множественны. Таким образом, подлинные онтологические индивиды как таковые не образуют иерархий, поскольку последние предполагают именно различия в соизмеримых признаках. Возможные отношения между простыми элементами могут быть только сетевыми, т. е. связывающими индивида с индивидом, а не с видом, родом, типом, классом и т.п.

В новоевропейской философии в наиболее продуманном и разработанном виде номиналистическая онтология представлена в «Логико-философском трактате» Витгенштейна. В рамках этой концепции объекты определяются тем, что

называется их «внутренними» свойствами и отношениями, которые, по свидетельству Мура, Витгенштейн позже предпочитал называть грамматическими отношениями [88, с. 87], т.е. отношениями, составляющими некую квази-априорную структуру мира. Так, то, что некоторое яблоко красное, является его «внешним» свойством, а то, что оно, как и любая точка любой наблюдаемой поверхности, имеет некоторый цвет, относится к «внутренним» свойствам. Таким же образом, внутренним свойством музыкального тона является то, что он имеет некоторую высоту, тогда как то, что он даёт ноту «си», — внешнее свойство. Нетрудно заметить, что внешние свойства объектов мира являются случайными, тогда как внутренние можно было бы назвать категориальными, хотя сам Витгенштейн предпочитает говорить о свойствах структуры [67, § 4.122].

Внутренние свойства и отношения объектов не могут быть высказаны в предложениях языка, но «показываются» во внутренних свойствах и отношениях составных частей предложения. С точки зрения этой онтологии, внутренними свойствами объектов, образующих общество, должны быть признаны автономия и свобода воли, поскольку, не имея таких свойств, объекты образовывали бы природные (физические в широком смысле) системы, изучаемые естественными науками. Однако, когда мы пытаемся явно выразить в языке это обстоятельство («человек наделён свободой воли»), мы не сообщаем ничего значимого, как если бы мы сказали: «Всякий наблюдаемый объект наделён цветом». Молодой Витгенштейн сказал бы, что мы высказываем нечто, не имеющее смысла, поскольку пытаемся сказать то, что может лишь быть показано. Поздний Витгенштейн сказал бы, что наше предложение имеет «грамматический», а не эмпирический, характер, поскольку говорит скорее об устройстве языка, чем о его предметах.

Свойства элементарных объектов должны быть сводимы к отношениям, иначе мы не можем мыслить их как элементарные. Атомарная теория Демокрита не была хорошей теорией, именно потому что он был вынужден постулировать разные свойства атомов, чтобы объяснить различные качества вещей. Витгенштейн же говорит, что даже если любой факт состоит из бесконечного множества атомарных фактов, а любой атомарный факт состоит из бесконечного числа объектов, тем не менее должны быть элементарные факты и объекты [67, § 4.2211]. Т. е., речь идёт о логической, а не естественной, необходимости. А коль скоро так, то элементарные объекты и весь спектр возможных

отношений между ними образуют *метаонтологический* уровень теории. Это то, что делает различные онтологии возможными.

В «Трактате» этот уровень описывается, частности, в афоризмах 2.02—2.0272. Объекты просты, поскольку они образуют субстанцию мира. Если бы у мира не было субстанции, истинность одного предложения зависела бы от истинности другого и картину мира невозможно было бы сформировать: ни истинную, ни ложную. Субстанция определяется конфигурацией объектов и может задавать только форму мира, но не материальные свойства. Любой воображаемый мир имеет общим с реальным именно эту форму. Субстанция существует независимо от фактов. Только благодаря наличию объектов у мира имеется фиксированная форма. Объекты неизменны, они — как бы единицы существования, конфигурации их, напротив, изменчивы. Единственно корректным логическим символом объекта является переменная.

Связь семантики с онтологией обсуждается, в частности, в [89]. Е.Д. Смирнова строит интересную концепцию «категориальных сеток», которые, по её мнению, участвуют в порождении возможных миров и которые не пропускают через свой фильтр высказывания вроде «Цезарь есть простое число». И хотя она специально оговаривает, что речь идет «не о разграничении “внутренних” и “внешних”, проявляемых при вхождении в атомарные факты, качеств объектов, как у Витгенштейна, а именно типах данностей объектов» (там же), отсылая к Мейнонгу за пояснениями термина «тип данности», мне всё же представляется на основе её же примера, что речь идёт именно о процедуре задания онтологии через определение объектов и их внутренних свойств — т.е., именно о том, что подразумевалось в «Трактате». Цезарь не может быть простым или каким-либо иным числом не по чисто логическим причинам — логически невозможным является противоречие, которого здесь не усматривается — и не по соображениям какой-либо предметной теории — например, физики, — а по причинам категориального порядка: Цезарь как объект нашей онтологии не обладает внутренними свойствами числа, потому что мы ему их не приписали. Мы могли бы, но полученная в результате онтология была бы не операциональна.

Таким образом, метаонтология — это философская квази-теория, трактующая возможность объектов и субстанции как формы данности мира. Она пытается сказать нечто о «первых сущностях», если пользоваться терминологией Аристотеля. *Конкретные (предметные) онтологии* — это области интерпретации научных теорий, показывающие, какие именно объекты, с ка-

кими свойствами и в каких отношениях полагаются существующими в рамках данной теории. Поскольку в таких онтологиях объекты обретают качественные определения и наборы возможных признаков, они, конечно, представляют собой «вторые сущности». Если воспользоваться популярным в аналитической метафизике различием типа и экземпляра (*type/token distinction*), которое восходит к работам Чарльза Пирса, объекты научных онтологий, безусловно, относятся к типам. Отсюда следует, что высказывание «Единорогов не существует» имеет эмпирический характер и означает буквально следующее: множество экземпляров соответствующего типа является пустым. Аналогичные отрицательные суждения относительно эфира или флогистона являются, в терминах Витгенштейна, «грамматическими» и высказывают нечто о соответствующих типах как объектах конкретной научной онтологии. Они суть предмет творческого выбора теоретиков и не требуют эмпирических подтверждений — за исключением того обстоятельства, что с этим выбором теория должна работать лучше.

То есть, с точки зрения метаонтологии, всё, что мы можем сказать о мире, относится к «фактам» – конфигурациям объектов. О самих объектах как «первых сущностях» мы не можем утверждать ничего, кроме их множественности. А множественность позволяет свести свойства к отношениям и избежать догматического приписывания свойств миру, оставив задачу их объяснения эмпирическим исследованиям. Согласно Витгенштейну, неверно говорить: «Существуют объекты», подобно тому как мы говорим: «Существуют книги» [67, § 4.1272]. Возможность первых, как и их множественность, характеризует наш способ концептуализации мира, а утверждение существования вторых — это уже результат его наблюдения: типу «книга» соответствуют реальные экземпляры.

Но что значит «существовать» в рамках номиналистического взгляда? Мы не только не можем спрашивать, существуют ли объекты. Мы также не можем спрашивать, существуют ли свойства, поскольку это всё равно что спрашивать, существуют ли переменные или системы координат. Объекты находятся в логическом пространстве возможных положений вещей [67, § 2.013], но это «нахождение» не равнозначно тому, что мы подразумеваем под существованием. Существуют (или не существуют) определённые положения вещей, например, животное, обладающее всеми признаками лошади, но с одним рогом; или некто, являющийся королём Франции в наше время. Именно поэтому

бессмысленно задаваться вопросом, существует ли мир как совокупность всех объектов, поскольку объекты – это лишь точки в логическом пространстве. До и безотносительно приписывания разным объектам каких-либо признаков или отношений для них характерно лишь то, что они различны [67, § 2.0233]. Именно множественность объектов, независимая от их свойств и отношений, есть то, что позволяет создавать картины мира, есть его субстанция [67, §§ 2.021, 2.0211, 2.0231].

### 1.2.3 Сведение свойств к отношениям влечёт сетевой подход

Я бы хотел особенно выделить концепцию логического пространства, намеченную, но не развитую в Трактате. Для физического пространства существенно наличие бесконечного числа точек, в которых могут находиться физические объекты. При этом — по крайней мере, в классической физике, — различные (не тождественные) объекты не могут находиться в одной и той же точке. Витгенштейн предлагает пространственную метафору и для логики — объекты, идентичные по своим свойствам, могут, тем не менее, быть различными объектами и самой возможностью своего существования задавать «пространственность» логической семантики. При этом нахождение объекта в логическом пространстве не влечёт — в отличие от нахождения его в пространстве физическом — его актуального существования.

Последнее обстоятельство составляет ещё один сдвиг в онтологическом видении, предложенный Витгенштейном, но отмеченный лишь некоторыми комментаторами, — а именно изменение представления о том, чему должно приписывать подлинное существование. Аристотель искал первые сущности именно как подлинно существующие, в отличие от вторых и всех последующих. Именно поэтому мы обнаруживаем двойственность подхода Аристотеля в определении субстанции мира: в качестве таковой он указывает то на онтологические индивиды, то на их форму, поскольку, в рамках его метафизики всё, что существует, характеризуется формой и материей.

В онтологической концепции Витгенштейна множественные объекты в логическом пространстве задают множество их возможных отношений, которое, как очевидно, не совпадает с множеством актуальных отношений. Отношения

объектов — «положения вещей» — которые мы застаём в актуальном мире, суть то, что образует область существующего. Термины «нынешний король Франции», как и «единорог», обозначают не объекты, а положения вещей. И именно поэтому мы можем осмысленно говорить об их (не)существовании. Эта метаонтология, в отличие от аристотелевской, абсолютно совместима с последовательно проведённым эмпиризмом и не требует метафизических «теорий» для собственного оправдания. Кроме того, эта концепция совместима с идеей множественности онтологий, которые могут соответствовать различным исследовательским целям. Так, например, в рамках социологической или экономической теории отдельный человек может рассматриваться в качестве онтологического индивида — который теоретически примитивен, далее не разложим, и все свойства которого производны от отношений, — тогда как в рамках психологии или медицины такая настройка теоретической оптики не будет полезна.

Подлинной формой утверждения о существовании чего бы то ни было является следующая:

(1)  $\exists x_1 \dots x_n (R(x_1 \dots x_n))$ , где  $R$  — некоторое отношение,

к которой сводится (не в дедуктивном смысле) любая формула вида  $\exists x (P(x))$ , поскольку, если объект обладает одноместным свойством, это указывает на его не-элементарность. И если такой объект существует, это значит, что за его свойство ответственна определённая конфигурация составляющих его элементарных объектов. Никакой определённый объект не может быть элементарным в рамках метаонтологии.

А это, в свою очередь, означает, что преимущество будет иметь та теория (в том числе социальная), язык которой способен репрезентировать объекты одного несложного типа и простые отношения между ними, поскольку онтология такой теории, будучи, насколько это возможно «низкоуровневой», ближе всего стоящей к философской метаонтологии, способна предложить наиболее простые формализации, объясняющие наиболее сложные явления. Очевидно, что отношение сложности объясняемых явлений к простоте объяснительных средств теории характеризуют её эффективность. Онтология предполагаемой научной теории, описывающая один тип объектов и сводящая свойства к отношениям, фактически описывает свои предметные области как сети. Такой способ описания позволяет использовать формальные инструменты, применяемые для исследования сетевых и сетеподобных комплексных объектов: теорию

графов, алгебру отношений и т. п. Далее я показываю, что такого рода сетевые онтологии релевантны как в когнитивной науке, так и в социальном познании.

#### 1.2.4 Возможные контраргументы и ответы на них

Изложенная здесь позиция может вызвать — и действительно вызывает — возражения. Далее я привожу реально прозвучавшие и предполагаемые мною возражения вместе с моими разъяснениями и контраргументами.

1.2.4.1. Все известные до сих пор попытки математизации социальных наук не приносили ожидаемых результатов.

И это неудивительно, поскольку не было предложено удовлетворительной онтологии. С другой стороны, имеются положительные примеры: например, в экономике и некоторых разделах социологии.

1.2.4.2. Неясно, какое место в предлагаемом понимании онтологии занимают идеальные объекты. Они не существуют, но высказывания о них не бессмысленны.

Напомню, что только в философии речь идёт о «первых сущностях», существующих независимо от каких бы то ни было «сказываний» о них. В научных онтологиях объекты задаются как типы, экземпляры которых обнаруживаются в экспериментальной ситуации. Т.е., следуя Аристотелю, как «вторые сущности». Вполне возможно высказывать нечто осмысленное, и даже истинное, о типе, экземпляры которого в реальности не встречаются: «Идеальный газ состоит из упругих сфер», «Единорог бел», «Гарри Поттер носит очки».

1.2.4.3. Положение о множественности онтологий в неклассической физике и компьютерных науках небесспорно.

По поводу физики я могу только отослать к статье [90], где в популярной форме излагаются проблемы, связанные с несовместимостью онтологий СТО и квантовой механики — именно в таких терминах. Что касается компьютерной науки, где «онтологии» как тема появились сравнительно



недавно, наиболее удачным их определением можно считать следующее: «Онтология есть формальная явная спецификация разделяемой [людьми и приложениями] концептуализации» [91]. Имеется в виду, конечно, концептуализация какого-либо «домена» — предметной области.

1.2.4.4. Введение «метаонтологии» скорее запутывает вопрос, чем проясняет его.

Метаонтологию легче всего понять как «онтологию онтологии». Если предметная онтология задаёт определённые объекты как типы, как правило, через перечисление их внутренних свойств («физическое тело имеет массу») и предоставляет исследователю возможность обнаружить их экземпляры на эмпирическом «поле», то метаонтология показывает, как вообще возможны объекты и какие-либо отношения между ними.

1.2.4.5. Возражение, которое я выдвигаю сам против себя: каким образом то, что я описываю как метаонтологию — множественность элементарных объектов, сведение свойств к отношениям — вдруг оказывается «сетевой онтологией»?

Конечно, между ними нет тождества: первая целиком и полностью принадлежит философии, вторую я хочу предложить предметным теориям: в частности, когнитивной науке и социальным дисциплинам. Сетевая онтология так же надстраивается над метаонтологией, как и обычная предметная. Чтобы надстроить предметную онтологию над метаонтологией, в общем случае нужно задать определенные типы объектов с определёнными внутренними свойствами и отношениями и т.д. В случае с сетевой онтологией задаётся лишь один тип объектов и один (или очень немного) типов отношений — связей (направленные, симметричные и т.п.). Вся остальная работа делается формальным аппаратом. Преимущества сетевой онтологии перед предметной я вижу, во-первых, в сравнительной простоте, во-вторых, в более широкой применимости, и в-третьих, в достаточной степени мультидисциплинарной реализованности: она сегодня присутствует в науках о вычислениях (теория сложных сетей, нейрокомпьютинг, мультиагентные системы, «искусственная жизнь» и «искусственное общество»), в социальных науках (теория социальных сетей, социология когнитивных социальных сетей, «сетевое общество») и в когнитивных науках (коннекционизм и «глубинное обучение»).

Как сказано в предисловии к сборнику «Сложные сети», выпущенному издательством Springer в 2004 г., исследования реальных крупномасштабных сложных сетей «открыли много направлений будущих исследований. Они продемонстрировали, что сложные сети — это интеллектуально глубокая и зрелая область, имеющая отношение ко многим научным дисциплинам, включая физику, биологию, инженерию и социальные науки, далеко лежащим за пределами традиционных областей математики и информатики. Теперь исследовательский фронт обращается к динамике сетей» [92, р. V–VI]. Дело осталось за теми разделами и дисциплинами, которые пока ещё не «обращены».

\* \* \*

Итак, свойства элементарных объектов должны быть сводимы к отношениям, иначе мы не можем мыслить их как элементарные. Атомарная теория Демокрита не была хорошей теорией, именно потому что он был вынужден постулировать разные свойства атомов, чтобы объяснить разнокачественность вещей. Элементарные объекты и весь спектр возможных отношений между ними образуют, если можно так сказать, метаонтологический уровень теории. Это то, что делает онтологию возможной.

Всё, что мы можем сказать о мире в рамках метаонтологии, относится к «фактам» – конфигурациям объектов. О самих объектах как таковых мы не можем утверждать ничего, кроме их множественности. А множественность нужна именно для того, чтобы свести свойства к отношениям и избежать догматического приписывания свойств миру, оставив задачу их объяснения эмпирическим исследованиям. Существуют (или не существуют) определённые положения вещей. Именно поэтому бессмысленно задаваться вопросом, существует ли мир как совокупность всех объектов, поскольку объекты – это лишь точки в логическом пространстве. Концепция логического пространства играет важную роль в нашем метаонтологическом исследовании. Нахождение объекта в логическом пространстве не влечёт — в отличие от нахождения его в пространстве физическом — его актуального существования. Преимущество этой онтологии состоит в том, что она абсолютно совместима с последовательно проведённым эмпиризмом и не требует метафизических презумпций для собственного оправдания. Ещё одно преимущество состоит в том, что эта концепция совместима с идеей множественности онтологий, которые могут соответствовать различным исследовательским целям.

Это означает, что преимущество будет иметь та теория (в том числе социальная), язык которой способен обращаться к элементарным объектам и отношениям между ними, некоторые из которых составляют исходные онтологические конфигурации данной теории. Такая теория, естественно, не может быть только качественной — она должна содержать логические или математические формализмы, способные описывать структуры.

Преимущество последовательно проведённого номинализма перед реализмом состоит, во-первых, в том, что его онтология индифферентна по отношению к любой «физике»: вопрос, как мир на самом деле устроен — компетенция эмпирических наук. Напротив, придерживаясь реализма, мы должны предложить метафизическую теорию, «объясняющую», почему существуют именно такие сущности. И эта теория, конечно, не будет свободна от догматизма.

Во-вторых, стоя на позициях реализма, мы должны признать тождественными объекты, наделённые абсолютно одинаковым набором свойств, включая их положение в пространстве и времени. Это не противоречит нашей картине макромира: здесь точные копии вещей или клоны существ всё же не могут иметь одни и те же пространственные координаты в один и тот же момент времени. Но в микромире, мире квантовых взаимодействий, мы не можем опираться даже на закон тождества: например, бозон может оказаться неограниченным количеством тождественных частиц, находящихся в одном квантовом состоянии.

Предлагая некоторую онтологию, я не обязательно должен верить, что мир устроен именно таким образом. Я рассматриваю данную концептуальную схему как основу для создания междисциплинарной методологии, которая поможет наилучшим образом выстроить целостные и взаимопроникающие теоретические концепции познания, общества и культуры.

### 1.2.5 Вывод из раздела 1.2

Наиболее перспективная онтология для социальных и гуманитарных наук структурно ближе всего к **философской номиналистической метаонтологии** типа метаонтологии «Трактата» Витгенштейна, отчасти перекликающейся с аристотелевой концепцией первых и вторых сущностей. Она предполагает существование **элементарных объектов, на множестве которых опреде-**

ляются отношения, типы которых конечны и немногочисленны. При этом условии основная миссия в том, что касается научного объяснения, ложится на математический или иной формальный аппарат — т. е., на собственно теорию.

### 1.3 Философия сознания и её значение для когнитивных наук

#### 1.3.1 Классификация позиций в философии сознания.

##### Функционализм

Основные направления в современной философии сознания подробно описаны в литературе, в том числе русскоязычной (см., напр., [5]). Здесь я хотел бы сосредоточиться на концептуальных препятствиях на пути научных прорывов в этой области, обнаруживаемых вместе или в различных сочетаниях в большинстве известных философских школ. Расхожие концепции сознания и восприимчивые к ним философские теории характеризуются молчаливо принимаемыми позициями — онтологическими презумпциями, — которые можно обозначить как семантический реализм, некритический эссенциализм и онтологический субстанциализм. Общим признаком перечисленных позиций можно считать *гипостазирование* (см. с. 306) терминов «народной психологии», которыми бы обозначаем различные аспекты когнитивной деятельности, и которые в гипостазированном виде обнаруживаются в некоторых философских доктринах.

**Семантический реализм** состоит в том, что слово «сознание» понимается как обозначающее некую единую онтологическую сущность, которая распадается на конкретные проявления — волю, мышление, чувственность, эмоции и т. п.

**Некритический эссенциализм** состоит в представлении, согласно которому у этой сущности есть общие всем её разновидностям свойства: например, идеальность, интенциональность или субъективность — невыразимость в объективных терминах.

**Онтологический субстанциализм** за всеми явлениями, обладающие свойствами сознания, усматривает одно и то же субстанциальное основание – например, мозг.

Попробуем разобрать эти позиции последовательно. Есть основания полагать, что словом «сознание» мы обозначаем в принципе разнородные феномены, объединённые разве что их предполагаемым «нефизическим» характером, т. е. общим отрицательным свойством. В соответствии с августинианским пониманием языка, как его представляет Витгенштейн [66, §1], всё, что называет какое-либо слово, обязано быть сущностью или её атрибутом. Поэтому, если существует термин «сознание», с помощью которого мы для какой-то функциональной надобности объединяем такие явления, как *qualia*, эмоции или логическое мышление, то существует и реальный референт этого термина — некая родовая сущность, сознание как таковое. Заключённая здесь методологическая опасность эта состоит в возможности — неоднократно реализовавшейся — некритического перенесения приемлемого объяснения одного явления, полагаемого нами «психическим», на все другие: например, когда мы считаем, что нечто из того, что можно сказать об ощущении, справедливо для мышления и других элементов «субъективной реальности».

Однако сам по себе факт наличия родового термина для нескольких видов ничего не говорит о соответствующей онтологии. Я вижу два возможных альтернативных основания для его наличия: общность некоторых свойств видов, объединяемых в данный род, – назовём это *катафатическим основанием*, используя богословскую аналогию, – эта возможность будет рассмотрена чуть ниже; или, напротив, неприсущность данным видам свойств всех других вещей: например, материальности, протяженности, пространственной локализации или наблюдаемости – это, соответственно, будет называться *апофатическим основанием*. И если у нас нет объективных оснований, чтобы утверждать, что мышление, эмоции или ощущения суть проявления одной и той же сущности – а я таких оснований не усматриваю, – то решение вопроса о семантическом реализме в понимании слова «сознание» будет зависеть от ответа на вопрос о некритическом эссенциализме.

Последний, как мы помним, состоит в представлении об общности свойств различных проявлений сознания. В качестве общих свойств ментальных явлений усматривается их идеальность, интенциональность или – как в случае с *qualia* – субъективная невыразимость. Но достаточно просто показать, что

интенциональность, под которой понимается «направленность на предмет» или свойство ментальных явлений быть «о чём-то», не является их всеобщим свойством: боль или эмоции вполне могут быть неинтенциональными. Невыразимость не имеет никакого отношения к рациональному мышлению, поскольку и его процесс, и его результаты вполне выразимы и общезначимы.

Чуть сложнее обстоит дело с идеальностью, поскольку за этим термином стоит длительная история философских туманностей и двусмысленностей. Но если мы проанализируем подборку философско-психологических определений этого термина из различных словарей и энциклопедий, доступную в Интернете [93], то увидим вариации из трёх позиций: (1) идеальное как противоположность материальному, (2) идеальное как бытие предмета отражённым в психическом мире субъекта (вариант: во всеобщих формах деятельности и культуры) и (3) идеальное как результат процесса идеализации — идеальный объект, который не может быть представлен в опыте.

Что касается (2), то нетрудно увидеть, что в этой интерпретации речь идёт об «интенциональном предмете», и что, следовательно, так понимаемая идеальность идентична интенциональности, всеобщность которой для различных проявлений сознания была показана чуть выше. Если в случае (3) для идеального объекта существенна его ненаходимость в опыте, то это, конечно, есть разновидность апофатического определения. Ну и в ещё большей мере очевидна апофатичность определения (1).

Таким образом, мы можем утверждать, что нет ни одного свойства, которое было бы обще всем явлениям сознания. А те, что могли бы претендовать на всеобщность, — идеальность в смыслах (1) и (3) — имеют явно апофатический характер. То есть, этим аргументом мы опровергаем не только эссенциализм, но и реализм в онтологии сознания: соответствующий родовой термин объединяет некоторые виды апофатически, то есть на том лишь основании, что их свойства не позволяют отнести их к другим родам, и не обозначает при этом никакой онтологической сущности.

Теперь о субстанциализме. Иллюзия единства сознания вызывает желание найти его единый субстрат или единую функциональную основу. Значительный прогресс, достигнутый нейронаукой в последнее время, поддерживает это искушение искать и находить в сложных нейроструктурах мозга объяснения всем нефизическим феноменам человеческой жизни.

Наиболее ясное философское выражение этой позиции содержится в так называемой «теории тождества типов» (*type identity theory*), в которой напрямую отождествляются типы — в отличие от единичных проявлений — ментальных событий с типами мозговых процессов [94]. Здесь необходимо заметить, что эмпирических подтверждений этой теории нет до сих пор, как нет и концептуальных оснований локализовать сознание в мозге. Однако есть множество контраргументов на концептуальном уровне: например, тезис о множественности физических реализаций ментальных состояний.

В этом месте нам необходим логический ход, который избавит объяснение сознания от всех трёх «идолов» языка. Это должен быть переход в духе галилеевского методологического редукционизма: не требовать от природы ничего, кроме «величины, фигуры, количества и ... движений». То есть переход от что-теории к как-теории, от субстанциализма к функционализму, где каждое проявление сознания идентифицируется по его функциональной роли.

В случае признания аргументов, приведённых выше, наиболее естественной для философии сознания выглядит позиция функционализма — который, по мнению Д. В. Иванова, являет собой «метафизику без онтологии» [95]. Согласно экскурсу в историю функционализма, как его даёт Джегуон Ким [96], направление началось с публикации в 1967 г. статьи Хилари Патнема «Психологические предикаты» [97].

Патнем, как многие полагают, опроверг уже упомянутую теорию тождества типов (которую Ким интерпретирует как убеждение, что любой мир, который идентичен нашему в его физических свойствах, идентичен ему полностью), а также общее представление о тождестве мозга и сознания. Взамен его статья дала начало функционализму, который понимается Кимом как обобщённая и более усложнённая версия бихевиоризма. Важные различия двух направлений сводятся к следующему:

- там, где бихевиорист увидит только поведение (например, крики и стоны при боли), функционалист будет утверждать наличие некоторого внутреннего состояния, которое вызывает это поведение;
- там, где бихевиорист будет видеть только отношение стимула и реакции, функционалист увидит отношение данного ментального состояния к другим.

Функционалист признаёт изначальный онтологический статус за ментальными состояниями, равно как и их собственное место в причинно-следственной

структуре мира. Все индивидуальные экземпляры общих типов ментальных состояний в качестве принципа своей индивидуации имеют следующий: то, что есть между ними общего, состоит в определённой каузальной роли, характерной для этого типа.

Итак, термин «функционализм» я буду использовать для обозначения семейства концепций, пытающихся найти объяснение ментальным событиям при помощи выявления их функциональных зависимостей от ряда других – ментальных и нементальных – событий, отвлекаясь от их возможных онтологических экспликаций.

Противостоящее ему семейство концепций – от традиционных материализма, идеализма, субстанциального дуализма (картезианства) до более современных физикалистского редукционизма, теории тождества типов<sup>4</sup> и др. – общим для которых является признание необходимой зависимости сознания и его свойств от его же субстрата-носителя, как бы последний ни понимался, равно как и неразрывной онтологической связи между ними, я в рамках настоящего текста предполагаю условно называть *субстанциализмом*. Попробуем представить некий общий функционалистский подход к мышлению.

Мышление – это то, что включает в себя мысли, состоит из мыслей. А мысль это то, что может быть истинным или ложным. Далее я попытаюсь показать, что всякий акт мысли подразумевает некую альтернативу: могло бы быть не так, как есть. Если мы построим автомат, который будет последовательно сканировать реальность и сопоставлять ей знаковые выражения пусть даже совершенного, божественного языка, это не будет мышлением, поскольку знаки будут связаны с реальностью причинно-следственным отношением.

Кант, как известно, утверждал: должно быть возможно, чтобы представление «я мыслю» сопровождало все мои представления [99, с. 126]. Я склонен предположить, что немецкое выражение *'ich denke'* может быть переведено не только в соответствии с «высоким штилем» философии как картезианское «я мыслю», но и вполне обыденно — как «я думаю, что...». Представим себе высказывание, например «Снег бел». Введём «[», «]» в качестве наглядного обозначения «трансцендентальности». Тогда, по Канту, наше выражение принимает следующий вид: «[Я думаю, что] снег бел». Но что мешает нам сказать это явно? «Я думаю, что снег бел». Рассел назвал бы это «пропозициональным

<sup>4</sup>Нужно отметить, что дискуссия о совместимости функционализма с физикализмом и теорией тождества типов имеет долгую историю в аналитической философии сознания — см., напр., [98].



отношением» [100, р. 227]. В любом случае, это вполне законное высказывание. Но тогда, в соответствии с Кантом, получаем: «[Я думаю, что] я думаю, что снег бел». Этого, наверное, достаточно, чтобы увидеть путь в дурную бесконечность: я думаю, что я думаю, что я думаю...

Но зададим вопрос: что значит «я думаю»? Что на самом деле мы хотим сказать этим текстом не трансцендентально, а вполне обыденно? «Сегодня дождь будет?» – «Я думаю, что да». Расшифруем: «Все указывает на то, что будет, но может статься и по-другому». И тогда мы просто можем заменить кантианскую формулу на более корректную: «Снег бел [хотя и обратное возможно]». В этом случае, *'ich denke'* оказывается лишь признаком относимости определенной мысли к «сфере *возможного* опыта».

Очевидно также, что так переосмысленное трансцендентальное единство апперцепции лишается своей «объединяющей силы» – мы уже, не греша непоследовательностью, не сможем утверждать вслед за Кантом, что единство Я является трансцендентальным условием единства предмета, поскольку кажется, что само Я у нас куда-то «выпало».

На самом деле, оно просто демистифицировалось. В нем не осталось ничего ни трансцендентного, ни трансцендентального — осталась лишь особая позиция говорящего, с которой ему открывается принципиальная альтернативность мира фактов. «Я утверждаю, что это так [хотя могло быть иначе]». Эта позиция на самом деле имеет смысл только в коммуникации с носителями иных *эпистемических миров* — множеств дескриптивных высказываний, считаемых истинными, ложными или неопределёнными.

Тогда определение мышления должно звучать следующим образом:

(Df1) Мышление есть коммуникативное поведение в условиях принципиальной альтернативности мира фактов.

Возможно возражение: если мышление имеет место только в мире фактов, то логики и математики — носители очевидно не-фактуального знания — не мыслят.

Подчас их собственные высказывания могут быть интерпретированы в платоническом духе: как если бы все истины, например, математики непосредственно усматривались мысленным взором в основных аксиомах. И тогда, конечно, «не думай, а смотри». Но сам автор этого афоризма писал: то, что у

нас есть правило конструирования натурального ряда чисел, не означает, что он нам уже в каком-то смысле дан целиком [66, §281].

То же для логики: если есть истинные посылки, из которых следует некий вывод, это не значит, что он уже тем самым сделан. Субъективная ограниченность знания создает интенциональные контексты. В глазах «абсолютного субъекта» холм Гиссарлык и есть местоположение Трои. Но Шлиман искал именно местоположение Трои, а не холм Гиссарлык. И поскольку никто не может претендовать на обладание всем множеством аргументов, имеет смысл сказать: «Я считаю (думаю), что  $q$ , поскольку верно, что  $(p \Rightarrow q \wedge p)$ ».

Есть говорящая английская идиома *'to be in a position to do smth'*: *'I'm in a position to argue that the snow is white'*. Может быть, имеет смысл говорить об *эпистемической топологии коммуникационного пространства*, структурированного уникальными наборами истинных, ложных и сомнительных утверждений, в котором каждый субъект и участник коммуникации занимает определённую *позицию* в определённый момент и прокладывает собственную траекторию во времени<sup>5</sup>. Поэтому уточнённое определение мышления должно звучать так:

(Df2) Мышление есть коммуникативное поведение в условиях принципиальной альтернативности мира фактов и ограниченности субъективного знания.

«Поведение», потому что на самом деле мы не «просто мыслим», а действуем с определенной целью. «Коммуникативное», потому что действие это имеет адресатом собеседника, внешнего или внутреннего. Именно ему я сообщаю, что, имея «А или не-А», я все-таки утверждаю, что «А».

Мышление решительно вклинивается в естественную взаимосвязь объектов, фактов и событий: «Я утверждаю, что это так, потому что это могло бы быть по-другому». В этом его субъективность, которая есть не что иное, как способность видеть возможные миры из уникальной перспективы субъекта, и которая уравнивается его объективностью как необходимой функции социально-коммуникативной системы.

<sup>5</sup>Об использовании топологических инструментов в эпистемических логиках см. [101]. В свою очередь, эпистемические логики успешно применяются в проектировании мультиагентных систем, так или иначе моделирующих социальную коммуникацию — см. параграф 3.2.3.

Опыт аналитической философии говорит об эвристической ценности исследования возможных употреблений слов, связанных с интересующими нас понятиями. Эта ценность объясняется тем, что сложившаяся лингвистическая практика выявляет концептуальные очертания исследуемых предметов значительно лучше, «рельефнее», чем любые теоретические рефлексии, отделяя «грамматическое» (в витгенштейновском понимании) и категориальное от привнесённого из опыта. «Не думай, но смотри!» [66, §60].

Возможные употребления слова «думать» предполагают понимание мышления не только как диспозиции (состояния), но и как своего рода деятельности. Мышление в первом смысле я ранее попытался определить как позицию в коммуникации. Язык, имеющий коммуникационную природу и предназначение, создаёт основу для эпистемической модальности самой возможностью отрицания того, что утверждается: «Снег бел[, но иное также возможно]». При этом высказывание «Я думаю, что  $p$ , но и  $\neg p$  возможно» – имеет смысл. Тогда как «Я думаю, что  $p$ , но  $\neg p$  невозможно» – вызывает ощущение странности: «я думаю» здесь кажется излишним. В отличие, однако, от многих других примеров с пропозициональными отношениями, этот эффект не зависит от грамматического лица: он проявляется в той же конструкции с «он» вместо «я»: «Он думает, что  $p$ , но  $\neg p$  невозможно». Ранее, на с. 18, мы рассмотрели предложения, которые имеют вид эмпирических высказываний, хотя, на самом деле, выполняют «грамматическую» роль. Например,

(p1) «Мир существовал задолго до моего рождения».

Задумаемся: ближе ли оно по своей функциональной роли в языке к высказываниям типа

(p2) « $A$  тождественно  $A$ »,

или к высказываниям типа

(p3) «Минимальное расстояние от Земли до Марса составляет 55,76 млн. км»?

Ни у кого не вызовет сомнения естественность употребления «Я думаю, что (p3)» в определённых ситуациях (урок в школе и т. п.). Однако употребление «Я думаю, что (p2)» и «Я думаю, что (p1)» — по крайней мере, вне философских дискуссий — в почти одинаковой мере оставляет чувство лингвистической неловкости и излишества. Как будто бы наследуемый нами язык

исподволь делит универсум на две категориальные части: несомненное и проблематичное. И то состояние (диспозиция), которое выражается словами «думать», «считать», «полагать», представляет собой как бы шаг на прокладываемом пути в области проблематичного, тогда как такой же шаг в царстве несомненного выглядит странным: как будто кто-то пытается осторожно ступить на тротуаре, по которому ежедневно проходят толпы. И, с другой стороны, состояние, выражаемое словом «знать», состоит в том, что, по мнению говорящего, путь на просторах проблематичного уже проложен.

Язык выполняет роль оптики, сквозь которую мы смотрим на мир. Приближает ли эта оптика к нам то, что есть, или, напротив, искажает его, — вопрос метафизический. Здесь же уместно было бы прояснить моё понимание трансцендентализма — не важно, кантовского или лингвистического — на простом примере. Мы с детства помним, что Гудвин Великий и Ужасный требовал от всех гостей Изумрудного города надевать зелёные очки, что и делало его город изумрудным. Общее правило трансцендентальной установки я сформулировал бы так: если мы видим некоторое свойство как присущее абсолютно всем вещам мира, то оно, скорее всего, является не свойством вещей или мира, а свойством наших «очков». У Канта в роли очков выступают априорные свойства восприятия и мышления, а у Витгенштейна — «глубинная грамматика» языка.

Теперь о мышлении как процессе или деятельности. Молчаливая деятельность по решению задач — вот что в общем представлении являет собой мышление, протяжённое во времени. И если мы зададимся вопросом, локализовано ли оно в пространстве, то подавляющее большинство из нас укажет на голову (хотя это зависит от качества его результатов). На нашем внутреннем феноменальном экране как бы проплывают образы и понятия, с которыми мы пытаемся совершать операции — к этой картине сводятся общепринятые онтологические презумпции относительно мышления.

Если мышление «идеально», то оно должно совершаться вне времени, сразу. Я думаю, что определённый путь может указать здесь психологическая концепция внутреннего диалога, восходящая к теории интериоризации внешней речи, разработанной Выготским. Мыслящий субъект умеет управлять коммуникацией, в том числе и её эпистемическими модальностями: знание, полагание и т.п., не обязательно создавая их целостные репрезентации. Подлинным «драйвером» мышления как процесса является напряжение коммуникационной ситуации, созданной наличием другого, собеседника. Каждый из собеседников

обладает собственным эпистемическим миром, структурированным категориями несомненного и проблематичного, которое, в свою очередь, делится на известное, установленное (то, что я «знаю») и неизвестное. Области несомненного у собеседников должны более или менее совпадать, чтобы коммуникация была возможной. А вот области установленного и неизвестного могут у них различаться. И тогда, например, в ситуации ученика и учителя область известного, имеющаяся у ученика, объединяется с таковой учителя, а в ситуации равноправной дискуссии, напротив, происходит их взаимное вычитание, что автоматически делает неизвестным то, что ранее одному из собеседников казалось установленным, порождая необходимость новых доказательств и обоснований.

Когда, в соответствии с теорией Выготского, происходит интериоризация внешнего диалога, и тот, кто мыслит, предположительно проигрывает все те же управляющие действия — возможно, в усечённом виде — симулируя наличие другого собеседника. Но, поскольку на проговаривание — даже редуцированное — нужно время, мышление в этом смысле слова оказывается процессом, растянутом во времени, несмотря на то, что в своём «идеальном» виде, изучаемом логикой, все мыслительные операции должны были бы совершаться одновременно, вне временных измерений.

Кто-то может возразить, что время необходимо мозгу по той же причине, по которой оно необходимо компьютеру: и тот, и другой совершают свои вычислительные операции в виде электрических, то есть, физических, локализованных в пространстве и времени, процессов. Несомненно, этот фактор также играет роль. И я думаю, что бессознательные процессы протекают сравнительно быстро именно потому, что требуют времени только на электрические взаимодействия. Что же касается мышления в собственно человеческом смысле слова, в его основе лежит рефлексивное представление о том, что мой эпистемический мир может оказаться не единственным, а главное — не единственно верным: представление, приходящее, безусловно, с опытом коммуникации (соответствующие данные возрастной психологии см. на с. 285). И тогда актуализируется стратегия внутреннего диалога, при которой наш внутренний компьютер должен не просто извлечь из памяти данные, а ещё и «решить», какие данные нужны и как они должны быть обработаны. Эта стратегия исходит из того, что возможны другие эпистемические миры, то есть, собеседники, у которых область проблематичного по-другому структурирована. Я бы сказал, что мышление как процесс диалогично в двух отношениях: во-пер-

вых потому что реально существуют некие другие, которые создают ситуации, требующие мыслительных операций; а во-вторых, потому что во внутренней речи задействуются паттерны, управляющие диалогом.

Как видим, между «мышлением» как диспозицией и «мышлением» как процессом семантически мало общего. Это ещё одно подтверждение верности нашего понимания гетерогенности тех феноменов, которые мы, по некоей причуде языка и народной психологии, объединяем родовым термином «сознание» и соподчинёнными ему (в нашем представлении) видовыми терминами, такими как «мышление». На мой взгляд, путь к научному пониманию сознания лежит через признание того обстоятельства, что сознание как единая онтологическая сущность не является данностью — ни логической, ни эмпирической — а, следовательно, каждый феномен, который мы до сих пор подвёрстываем под это понятие, может и должен быть исследован в своей собственной фактологии и своей собственной концептуальной структуре.

### **1.3.2 Коммуникативный функционализм. Пропозициональные отношения как коммуникационные модальности**

Итак, как мы увидели, экстраполяция функционалистских принципов на интерсубъективную сферу, обнажает структурно-логический каркас проблемы сознания и позволяет тем самым увидеть решение в его концептуальной простоте: слова «знание», «мышление», «сомнение» и т. п. обозначают не более и не менее, чем взаимно определяемые позиции акторов в логической структуре той или иной коммуникационной ситуации — в её эпистемической топологии. Мозг, компьютеры и другие естественные и искусственные устройства необходимы акторам для правилосообразного функционирования, но не достаточны для того, чтобы знать, мыслить или сомневаться. Такую интерпретацию можно было бы обозначить как *коммуникативный функционализм*. Эта позиция позволяет, в частности, развести в стороны «сознание изнутри», его феноменальный аспект, и то, что мы понимаем под «субъектом познания». «Знание», «сомнение», «мнение» имеют отношение прежде всего к информационным процессам внутри систем коммуникации и далеко не в первую очередь — к процессам внутри мозга. Их можно считать *коммуникационными модальностями*.

В доязыковых средах — например, в сообществах животных, — коммуникативный функционализм, который состоит в *возможности дедукации всего, что мы связываем с мышлением, из модальной структуры коммуникации*, оказывается скорректирован в пользу усиления биологического начала — потребности. Это позволяет сформулировать гипотезу, если можно так сказать, *прагматического функционализма*, выводящую сознательное начало из превращения потребности в конструктивно заданное множество возможных значений и проекции последнего на «чувственный экран» в виде значащего чувственного образа. Эта схема претендует на роль пролегоменов к философии доязыкового сознания и мышления. Со стороны когнитивной науки она находит подтверждение в теории предиктивного процессинга (см. параграф 2.2.5) и минимизации свободной энергии, подробнее о которой см. на с. 145.

Языковая коммуникация в этом случае может вступать в игру как весьма важный, но в принципе не необходимый элемент, как несколько более эффективная среда для трансляции значений. Язык как средство коммуникации, метафорически выражаясь, как бы усиливает проективную силу значащего образа, продолжая лучи проекции за пределы чувственного экрана, трансцендируя всё множество полученных значений и конституируя мир объектов в соответствии со структурой изначальной потребности и собственной концептуальной структурой.

Сформулированная гипотеза может иметь значение для искусственного интеллекта. Очевидно, что т. н. «чувственный опыт» — это биологическая форма доступа к данным. В случае с машиной, наделённой сознанием, форма доступа к данным может быть самой разной, в зависимости от потребности создателя: доступ к уже сформированной базе данных, имитация биологических органов чувств, ультразвуковое или инфракрасное сканирование предметной области или что-либо ещё. Главное состоит в необходимости воспроизведения блок-схемы: потребность  $\rightarrow$  множество возможных значений  $\rightarrow$  значащий чувственный образ ( $\leftarrow$  язык)  $\Rightarrow$  онтология (*Umwelt*).

### 1.3.3 Невыразимость субъективного и перспектива первого лица

Главный, на мой взгляд, вызов для концепции коммуникативного функционализма представляют проблемы qualia и «я», понимаемого как внутреннее самоощущение. Они представляют трудность, именно потому что, будучи концептуально определены как нечто невыразимое и непередаваемое, оказываются непрозрачными для самой идеи функционального воспроизведения. В то же время, во многих концепциях qualia имплицитно рассматриваются как строительные кирпичики сознания, которые своей субъективной невыразимостью делают его — сознание — в целом субъективно-невыразимым ощущением *каково-это-бытности*<sup>6</sup>. Далее я постараюсь внести некоторую ясность в эти вопросы.

В рамках концепции «позднего» Витгенштейна, в языковых играх qualia иррелевантны — конструирование значения происходит без их участия. Можно ли каким-либо образом вывести qualia из нашего понимания коммуникации? Для этого необходимо задуматься, в чём состоит разница между (а) «чувствовать красное» и (б) «понимать, что такое «красное»». Возможно ли (а) без (б)? Какое знание добавляет к своему компендиуму начитанная Мэри, впервые в жизни выйдя из чёрно-белой комнаты и впервые увидев пожарный гидрант или спелый помидор, если она знала заранее, что люди наделяют их одним и тем же свойством, которое ей до сих пор было «субъективно» недоступно [103]? Например, мы учим ребёнка: сложи сюда всё красное, а сюда — всё круглое. Почему мы считаем, что круглость, в отличие от красности, — объективное качество? У нас есть язык (концепции) для его экспликации (в диалоге): мы можем сказать, что круглый — это равноудалённый от центра, квадрат — это равносторонний прямоугольник и т. п. Для «красного» просто нет языка экспликации, но есть практики различения. Отсюда — представление о личном, закрытом для других контейнере, где содержатся qualia.

Поясню на витгенштейновском примере. В «Философских исследованиях» подробно разбирается проблема приватности ощущений, расхожее понимание которой Витгенштейн обозначает фразой «Только я могу чувствовать мою боль» [66, §253]. С представлением о приватности qualia связан также некри-

<sup>6</sup>Калька с англ. 'what-is-it-likeness' — существительного, образованного от названия знаменитой статьи Томаса Нагеля [102].



тически принимаемый тезис, что значением слова «синий» является моё (или чьё угодно) приватное невыразимое ощущение синего цвета. Витгенштейн с помощью известной аргументации показывает, что это не так, что различие цветов есть одна из языковых игр, понимаемых им как социальные институты. Иными словами, произнося «синий», мы называем не собственное невыразимое ощущение, а социально санкционированную практику отличия синих объектов от зелёных, жёлтых и любых других. И, следовательно, значение слова «синий» никак не зависит от субъективной определённости ощущения этого цвета в «сознании».

Рассмотренные нами сомнения в отношении qualia способны вызвать смятение у тех, кто отводит опыту решающую роль в познании. Однако уже Канту было ясно, что опыт не сводится к чувственным переживаниям: он состоит в том, что «нечто обстоит так-то и так-то», а для этой идеи необходим расчётливый синтез. Опыт выражается в фактуальном предложении, истинность которого может быть установлена наблюдением.

Кроме того, против отождествления опыта и чувственности работает и следующий аргумент: один и тот же факт – например, «Эта комната – узкая» – может соответствовать различным чувственным комплексам у слепоглохого и зрячеслышащего человека.

Данное предложение, выражая «суждение», располагается по эту сторону qualia-экрана, а по ту сторону мы предполагаем наличие «факта» как отношения между «объектами», логически идентичного отношению между именами в предложении. Как может выглядеть семантика этого предложения? Отношение между объектами – какими? Очевидный ответ: более узкой и более широкой комнатами. Вполне законная версия, если мы, например, осматриваем квартиру. А если мы пытаемся внести рояль? Можно сказать, что и здесь имеется в виду соотношение размеров комнаты и рояля, только по-другому определённое. Ну а если мы просто выражаем свою эстетическую неудовлетворённость геометрией комнаты? Наша естественная установка побуждает нас искать референты одноместных предикатов непосредственно в чувственности («ощущение узости»). Но мы уже знаем, что искать «значения» там бесполезно. Уж не говоря о том, что во многих ситуациях высказывание «Здесь узко» будет полностью синонимично обсуждаемому. Но какое отношение и между какими объектами оно выражает?

Ответ даёт автор «Философских исследований» [66]. Вы хотите знать, каково значение предложения «Эта комната – узкая»? А как вы объясняете это значение ребёнку? Ваши движения и взаимодействие с маленьким слушателем и будут составлять это значение в языковой игре «объяснение значения знака ребёнку». В языковой игре «меблировка комнаты» ваши действия будут другими, соответственно, и значение этого предложения будет немного отличаться. С этого момента мы можем говорить о, если так можно сказать, «динамической семантике» предложений. Если языковая игра (ЯИ) состоит из лингвистических и экстралингвистических компонентов, то значением первых являются те изменения, которые их употребление вносит во вторые. А семейство всех ЯИ и составляет наш «жизненный мир». Оно же составляет и цемент всей социальной реальности, так как нетрудно показать, что социальные отношения обеспечиваются систематическим обменом знаковыми сообщениями.

Примитивный язык состоит из сигналов и команд. Эволюционируя, он дифференцирует внутри себя имена, глаголы и другие грамматические категории. И то значение, которое в примитивной ЯИ может быть выражено возгласом «Здесь узко!», в более сложных передаётся дескриптивным выражением «Эта комната – узкая». Но «дескриптивным» его делает наше «естественное» убеждение в том, что если комната узкая, значит она существует. И не просто существует, а описывается дескрипцией «комната». Так грамматика языка создаёт для нас онтологию «мира».

На мой взгляд, можно говорить о двух принципиальных установках в философии сознания: феноменологической, в которой принимающее свободные решения «я» есть непосредственная данность, и редукционистской, в которой ментальные состояния, как более сложные, являют собой функциональные или эмерджентные свойства не-ментальных элементов.

Так, со второй точки зрения, *qualia* можно представить как – предположительно, развитый в результате эволюции – чувственный интерфейс между центром управления поведением и полем деятельности организма. Локк называл их «вторичными качествами» вещей, не принадлежащими самим вещам. Беркли вполне логично возражал, что представление о «вещи» здесь не необходимо, а следовательно, излишне. Юм и Кант постепенно формируют убеждение в том, что за чувственным «экраном» что-то существует, но это что-то и образы «экрана» не обязаны быть друг на друга похожи. А поскольку «опыт» отождествлялся ими с этим кино, которое скрытый ноумен показывает субъ-

екту, следовал вывод, что он — опыт — не может быть источником знания о «вещи-в-себе».

Функциональную структуру феноменального сознания можно рассмотреть на примере схемы «сенсор — сигнал — реакция», действие которой подчинено тому, что мы называем поведенческой целесообразностью, то есть, вообще говоря, имеет некое прагматическое оправдание. Взятая в отдельности, эта цепочка пока не предполагает необходимости qualia. Вряд ли фотоэлемент в метро чувствует что-либо.

Но для этой схемы существенно, что она не линейна, каждый её элемент и вся она в целом находится во взаимосвязи со множеством других сенсоров, стимулов и реакций, и прагматически оправданный результат требует интеграции всех этих элементарных взаимодействий в единую систему обеспечения целесообразного поведения организма. Поэтому возникает потребность в разрыве непосредственных причинно-следственных связей элементарных стимулов и реакций и ввода некоторых из них во временный режим ожидания — подобно требованию предварительной систематизации информации по всему предприятию для принятия сложных управленческих решений.

Чем сложнее управляемая ситуация, тем требуется большая степень разрыва — или временной приостановки действия — причинно-следственных цепочек и отстранения «взгляда» управляющих центров от отдельных деталей. Возникает механизм репрезентации целостных структур данных, которые должны быть оценены, и по которым должны быть приняты управленческие решения. «Изнутри» это выглядит как повышение «сознательности», при которой в большей степени задействуются память и внимание. Ощущения и чувственные образы — первые ступени восхождения к сознательности, комплексы данных, представленные решающей инстанции в виде целостных репрезентаций. С появлением социальной коммуникации в эту же схему включается язык. Целостное «схватывание» происходит, потому что элементарные операции не становятся предметом внимания.

Таким образом, ощущения, чувственные образы, представления, концептуальное содержание и лингвистические смыслы — это всё различные степени сложности сетевых механизмов управления индивидами и группами. В [104] достаточно убедительно было показано, что любая чувственная картина мира — у нас или у животных — подобна не пейзажу, а скорее компьютерному интерфейсу. Она эволюционно сформирована таким образом, чтобы максимально

облегчить деятельность живого организма, а вовсе не затем, чтобы одарить его возможностью непосредственного созерцания объективной истины. Разница эта очевидна любому пользователю компьютера: никому ведь не придёт в голову считать, что графический интерфейс пользователя является точным «отражением» компьютерных внутренностей, или хотя бы даже логики программы.

Однако, при всей субъективности наших ощущений, дело может обстоять таким образом, что чувственная картина мира воспроизводит реальность структурно, через внутренние отношения элементов, которые изоморфны объективным отношениям онтологических сущностей. В этом направлении движется часть философских исследований, возрождающих, например, концепцию иконических знаков Чарльза Пирса [105]. Нужно только иметь в виду, что предполагаемая интенциональность чувственных образов и характер репрезентации, ими демонстрируемый, составляют «лёгкую» проблему сознания. «Трудная» проблема — каким образом качественная феноменология сознания может быть встроена в физикалистский язык научных описаний — ответами на «лёгкие» вопросы не решается.

### 1.3.4 Проблема «я». Может ли «я» умереть?

Ещё один комплекс вопросов, составляющих «трудную» проблему сознания — это человеческое «я», известное также как «самость» («self») или апперцепция. В соответствии с классической установкой в философии, «я» играет важную роль в структурировании сознания и в познании, осуществляя некий синтез впечатлений или значений. На мой взгляд, в основе такой позиции лежит эссенциалистский подход к языку: убеждение в том, что всякое слово называет некоторую сущность. Каждый из нас называет себя «я», а поэтому — «у каждого есть своё Я». Философский «я»-дискурс, таким образом, представляет собой теорию, описывающую в третьем грамматическом лице значения терминов первого лица.

В то же время, есть подлинная загадка человеческого «я». Это тот вопрос, который задаёт себе (или окружающим) каждый ребёнок в какой-то момент: почему я — это я? Что будет, когда я умру? Однако едва ли в настоящее время просматривается какая-либо перспектива позитивного решения этих головоло-

мок, поэтому данная территория пока уверенно удерживается философией. Вместе с тем, достойным рассмотрения выглядит то, что я назвал бы *парадоксом философской эгологии (ПФЭ)*:

ПФЭ: Я не могу быть бессмертным как биологическое существо, как вещь этого мира, но я не могу быть смертным как системообразующий принцип этого мира.

Парадокс основан на (1) допущении сущностного единства «я» и (2) понимании его как «контейнера» субъективных переживаний, которые (3) упорядочены во времени и неизменны памяти.

Может ли «я» умереть? Для выявления возможности ответа на этот вопрос выделим два аспекта, в которых может употребляться «я»: *коммуникативный* и *квалитативный*. В первом из них ПФЭ решается тривиально и нефилософски: «я» — это прагматический оператор языка, означающий «тот, кто это говорит». Поэтому никаким «системообразующим принципом» этот оператор не является, а биологическое существо в должное время умирает, как ему и положено.

Во втором аспекте ПФЭ не имеет места: представления о границах жизни — о дате рождения и дате смерти — целиком принадлежат коммуникативному аспекту. Но «коммуникативное Я» умереть не может, так же как не может перестать существовать «здесь» и «теперь».

Рассмотрим два ряда событий и фактов — субъективный (С) и объективный (О). Первый соотнесён с моими переживаниями и памятью, второй имеет языковую, коммуникативную, теоретическую природу. К первому относится, например, внешний облик моих родителей в моей памяти, ко второму — дата моего рождения. Событие моего рождения (СМР) — событие О-ряда, которое не имеет места в С-ряду. Близкие к нему по времени события С-ряда, содержащиеся в моей памяти, по мере приближения к СМР становятся всё менее чётко воспринимаемыми и всё более зависимыми от событий О-ряда: сообщений других людей, документальных записей и т. п.

Помимо этого, мы сталкиваемся с временной однонаправленностью сознания: память хранит события прошлого, но не будущего. Онтология прошлого предстаёт как соотнесённость событий О-ряда с содержанием памяти. Онтология будущего, в свою очередь, — как проекция возможных событий О-ряда на предполагаемые факты С-ряда. Прошлое дано в том или ином виде, будущее

только предполагается, у него нет актуального С-ряда. Т. о., СМР не сопоставлено ни одно впечатление С-ряда. Но то же можно сказать о событии моей смерти (СМС). Я предполагаю, что в О-ряду СМС будет соответствовать определённая дата. И я могу предположить, что в этот момент времени ему будут сопоставлены какие-то переживания С-ряда. Но в каждый настоящий момент СМС является таким же О-событием без сопоставленных ему С-впечатлений, как и СМР. Причём первое — исключительно концептуальная конструкция: события будущего предполагаются, но не переживаются. Чистая же «теоретичность» СМР — эмпирическая загадка: почему я ничего не помню об этом?

Однако несомненно то, что ни одной из О-границ моей жизни (ни СМР, ни СМС) не соответствует ни одно событие С-ряда.

Итак, я имею два ряда содержаний моего сознания: С-ряд, состоящий из «здесь и теперь» и памяти, но не содержащий никаких представлений о собственных границах; и О-ряд, заставляющий меня полагать, что я в определённый момент родился и в определённый момент умру, но всецело зависящий от культуры и обучения. На какой из них нужно положиться?

В «Загадке человеческого Я» [106] Феликс Михайлов пишет, что наглядная разгадка — это слепоглухонемые дети, успешно социализированные и получившие хорошее образование. В каком-то смысле это верно, поскольку показывает, что подлинным значением термина «я» не является ничто из относящегося к С-ряду: ни его события, ни сам он в целом. Людей, не обладающих «нормальным» С-рядом, можно научить правильно пользоваться словом «я». Думаю, что этому же можно научить и компьютер в исключительно коммуникационном аспекте.

А в качественном? Очевидно, что компьютер можно научить хранить в памяти изображения и совершать какие-то действия, если эти изображения узнаваемо похожи на актуально воспринимаемые образы, чтобы, например, моделировать поведение ребёнка, который тянется к матери, узнав её. Но можно ли научить компьютер сохранять в памяти «образы» его включения и отключения?

Опираясь на свой С-ряд, компьютер никогда не сможет определить, что значит для него «быть выключенным». Но из этого не следует, что он будет включён вечно.

Итак, можно утверждать, что в ПФЭ смешиваются два концепта: ( $y_1$ ) «я» как особая позиция говорящего в коммуникативной системе и ( $y_2$ ) приватные

ощущения и переживания (проприоцепция и т.п.).  $Я_1$  не может умереть, поскольку не является определённым существом или вещью, а только способом языковой автореференции говорящего. А носитель приватного содержания ( $я_2$ ) умереть вполне может и, скорее всего, это сделает, но не может в каждый данный момент отразить этот факт в своих впечатлениях С-ряда.

Кроме того, высказывание «я умру» по смыслу эквивалентно высказыванию «моё тело умрёт». Чего нельзя сказать, например, о высказывании «я знаю»: всё же это не моё тело знает, что я родился в определённый день определённого месяца и года. Физикалист, однако, скажет, что именно тело это знает, поскольку мозг – часть тела. Контраргумент может быть следующим: «я знаю», «он знает» — не внутренние состояния тел, а их позиции в системе коммуникации как акторов. Моё тело, включая мозг, хранит процедуры, благодаря которым я могу осмысленно включиться в межчеловеческую коммуникацию, где другие могут сказать обо мне: «Он знает, что  $p$ ».

Ещё раз: «я» (в смысле  $я_1$ ) не может умереть, как не может пройти «сейчас» и нельзя покинуть «здесь». Смерть становится проблематичной, когда « $я_1$ » смешивают с « $я_2$ »: психологическим «контейнером» событий С-ряда. Первое не может умереть по чисто концептуальным основаниям, а второе не способно на качественную репрезентацию события собственной смерти.

Рассмотрим теперь трансценденталистские притязания на философскую исключительность «я». Введём для этого различие трансцендентального  $Я_t$  и психологического  $Я_p$ . Когда говорят, что  $Я_t$  есть конститутивный принцип мира, условие единства опыта и т. п., смешивают по ложной аналогии  $я_2$  (в предыдущей части анализа) — «только я могу иметь мои ощущения» — с  $я_1$ , придавая последнему свойства исключительного собственника семантики осмысленных высказываний о мире, каковым оно не является.

Само время является одной из опорных конструкций О-ряда: если С-переживаниям не сопоставлены О-события, локализованные на временной шкале, то я, скорее всего, не смогу определить, какие из впечатлений, хранимых моей памятью, получены «раньше», а какие — «позже». Я могу только отличить С-впечатления «здесь и теперь» от более или менее отчётливых содержаний памяти.

Можно было бы сказать, что ПФЭ не имеет места, поскольку С-ряд, переживаемый  $я_2$ , не образует мир (мир упорядочивается в коммуникации), а  $я_1$  не является никаким системообразующим принципом. Но на чём же основана

неистребимость концепции «бессмертия души»? В её основе — представление о привилегированном положении моего тела как единственного наблюдательного пункта моего я<sub>2</sub>: если тело умирает, то прекращается киносеанс, а кто же тогда встанет и выйдет из зала?

Ответ может состоять в следующем. Концепция «жизни» принадлежит я<sub>1</sub>, поскольку, как уже было сказано, только на временной шкале О-ряда возможны «до» и «после». С-ряд, принадлежащий я<sub>2</sub>, локализован на этой шкале только во взаимодействии с культурой, а сам по себе не имеет временных характеристик, и поэтому не имеет смысла говорить о его «прекращении».

Мой интенционально предполагаемый собеседник говорит: «Я не могу представить себе свою смерть и то, что будет после неё». — А могу ли я представить себе своё рождение и то, что было до него? Стоит лишь абстрагироваться от представления об однонаправленности времени, и мистерии становятся обыденностью. Если душа должна пережить тело, потому что невозможно представить себе полное отсутствие представлений, то почему нас не смущает полное отсутствие представлений о её дотелесном существовании, да и о моменте воплощения?

«До» и «после» — навигационные инструменты, которыми «душа» (я<sub>2</sub>) изначально и не владеет. Только в системах координат культуры оказывается, что всё то в С-ряду, что не локализовано в «здесь и сейчас», представляет собой или содержание снов, или память о событиях «прошлого». Поэтому, «что будет после моей смерти?» — это не тот вопрос, ответ на который стоит искать в тайниках моей души. Качественное «я» умереть может, но не может иметь этого опыта, также как не имеет опыта собственного рождения. Я знаю, что я родился, но не помню этого. Я знаю, что умру, но... Если «душа» может иметь посмертные впечатления, почему у неё нет пренатальных? Загадка коммуникативного «я» тривиальна и решается эмпирически. Загадка качественного «я» относится к сфере мистического, невыразимого, и средствами известных нам науки и философии не решается.

Объектное, эссенциалистское понимание «я» может предполагать следующие варианты:

- (1) «я» как термин, относящийся к физическому телу, что, по видимости, происходит в высказываниях типа «Я болею» или «Я потолстел»;



- (2) «я» как термин, обозначающий эмпирическое содержание памяти и сознания, например, «В школе меня не любили», «Я плохо запоминаю цифры»; и, наконец,
- (3) декартовское *ego cogito*, а также кантианское понимание «я» как априорного представления, необходимо сопровождающего любой акт мышления.

Рассмотрим первый случай. Предположим, случилось страшное, и я прибавил в весе. Тогда высказывание «Игорь Михайлов поправился» (в плохом смысле слова) будет истинно как в устах другого человека, так и в моих устах, в последнем случае лишь немного странным по способу выражения. Тогда как «Я поправился» будет истинно только в моих устах. Это ясно указывает на то, что условия истинности высказывания с «я» лежат в сфере прагматики, то есть, свойств конкретной коммуникационной ситуации, а не только в сфере семантики, как в случае и «именным» высказыванием. Если «я» и «Игорь Михайлов» не являются синонимами, а демонстрируют взаимозаменяемость только при определённых прагматических условиях, то логично было бы рассматривать роль термина «я» в языке в качестве своего рода прагматического оператора. Можно говорить о таких стилистически обусловленных синонимах этого термина, как «ваш покорный слуга», «автор этих строк», а также английское 'this man' (хотя под другим углом зрения и при другой прагматике их можно рассматривать как дескрипции).

Нетрудно заметить, что те же аргументы имеют силу и для варианта (2). То есть, вне зависимости от того, о каких эмпирических свойствах личности — физических или интенциональных — мы говорим, «я» может заменить другой термин, если и только если заменяемый термин является именем или дескрипцией говорящего. Теперь о декартовском, кантианском, гуссерлианском — или вообще трансцендентальном, — понимании «я». Не вдаваясь в тонкости, все три подхода можно объединить на одном общем основании: имплицитное присутствие мыслящего «я», лишённого каких бы то ни было психологических определённостей, полагается в всех трёх традициях — картезианской, кантианской и феноменологической — совершенно необходимым для того, чтобы «чистое» мышление было возможным. Во всех трёх традициях трансцендентальное «я» получается путём освобождения эмпирического «я» от несущественных свойств, с тем чтобы оставить то единственное, которое выражает его сущность: мышление у Декарта, внутренняя связь суждений у Канта

и направленность на предмет у Гуссерля. В свою очередь, освобождая все три концепции от несущественных различий между ними, думаю, было бы корректно сказать, что — Гуссерль выразил суть точнее всех — сущность чистого «я» состоит в его «о-чём-то-бытности»: я не могу придумать лучшего перевода слова 'aboutness', придуманного Сёрлом.

В этом пункте возникает законное подозрение: не используется ли здесь слово «я» вне его категориальной принадлежности, а именно — как существительное, как синоним слова «сознание»? Если это не так, то необходимо показать, что сознание раскрывает свою «о-чём-то-бытность» — то есть, интенциональность, — только в перспективе первого лица, только в моей собственной интроспекции. Этот подход в какой-то мере роднит Декарта с Гуссерлем, тогда как у Канта он не столь очевиден: именно поэтому концепция трансцендентального единства апперцепции выглядит у него немного *ad hoc*.

Попробуем использовать естественную асимметрию грамматических лиц: эффект, проявляющийся в том, что, если высказывание «Он не знает, что  $2 \times 2 = 4$ » имеет смысл, то высказывание «Я не знаю, что  $2 \times 2 = 4$ » его не имеет. Но «знание» всё же выражает некую эпистемическую — или, как я уверен, коммуникационную, — позицию, тесно связанную с истинностью подчинённого предложения. Мы же, вслед за тремя великими философами, пытаемся увидеть сущность чистого «я» в более простом и абстрактном свойстве — свойстве «быть о чём-то», что необязательно должно быть связано с истинностью или реальностью этого чего-то. Для этой цели лучше подходит интенциональный предикат «думать» («полагать») — что, по видимому, и имел в виду Декарт.

«Я не думаю, что расстояние от Земли до Луны составляет 384 400 км» — «Ты не думаешь, что расстояние от Земли до Луны составляет 384 400 км» — «Он (она) не думает...» Как, видим, есть вполне представимые ситуации, в которых это предложение в любой из личных грамматических форм окажется осмысленным. А следовательно, использование личного местоимения в качестве философской категории представляет собой некий атавистический след эпохи интроспекционизма и, в общем и целом, не имеет смысла.

Наш беглый грамматический анализ показывает, что перспектива первого лица имеет значение для некоторых, но не для всех интенциональных актов. А следовательно, «трансцендентальное» присутствие «я» во всех проявлениях сознания не требуется, что автоматически лишает его статуса трансцендентальности.

Иначе говоря, наша техника состоит в следующем. Трансцендентальность предполагает аподиктичность — то есть, всеобщность и необходимость. Но если присутствие чего бы то ни было в сознании всеобще и необходимо, то отрицание этого присутствия должно быть немыслимо (как существование физического мира вне пространства и времени). И тогда достаточно показать мыслимость отрицания некоторых интенциональных предикатов в первом лице, чтобы лишить чистое «я» статуса всеобщности и необходимости, а следовательно — трансцендентальности. Но, поскольку мы это продемонстрировали, то теперь мы вправе утверждать, что трансцендентальное «я» — это философская фикция.

Философы витгенштейнианской школы всегда исходили из того, что философские — в собственном смысле слова — проблемы обусловлены неправильным пониманием категориальной структуры и функционирования нашего естественного языка. Символы, которые мы используем, ничего не говорят об их категориальной принадлежности. Отсюда возникает иллюзия, что слова «стол», «я», «человек», «мышление» функционируют одинаково, то есть обозначают некоторые объективно существующие сущности, и задача исследователя — их определить и классифицировать.

На самом же деле — в некотором приближении — мы можем говорить о слова-константах, словах-переменных и словах-операторах. И, несмотря на наше интуитивное стремление отнести все имена нарицательные ко второй категории, некоторые из них, а также многие, если не все, местоимения, относятся к третьей. Так, слово «я» вовсе не обозначает никакой сущности или состояния (вроде трансцендентального единства апперцепции), а является оператором аутореференции — «тот, кто это сейчас говорит», подобно оператору 'this' в некоторых языках программирования. Аналогично, слово «человек» может быть осмыслено как оператор родовой аутореференции — «я и любой, кто в принципе может понять то, что я сейчас говорю». Отличие оператора от переменной состоит в том, что его значение относится не к семантике, а к прагматике — к свойствам конкретной коммуникационной ситуации. Поэтому, что такое «человек» определяется параметрами такой ситуации и объединяется отношениями «семейного сходства»: каждый контекст употребления слова «человек» чем-то похож на некоторый другой, но нет ничего такого, чем они все были бы похожи друг на друга. «Человек» прагматически определяет коммуникационный универсум: это тот, кто в принципе может участвовать в коммуникации с помощью данного языка.

### 1.3.5 Сознание и осознание

Понимание степени осознанности как степени интегрированности нейронных возбуждений в различных отделах мозга подтверждается интересными эмпирическими исследованиями, например в [107]. Дискуссия вокруг этих терминов развернулась, в частности, в 2006–2007 гг. в журнале *Journal of Consciousness Studies*, после публикации статьи Кристиана де Куинси [108]. В статье утверждалось, что смысл слова «сознание» (*consciousness*) относится скорее к онтологии, тогда как в термине «осознание» (*awareness*) выражается психологический аспект. С точки зрения де Куинси, можно говорить о философском и психологическом преломлении понятия сознания. В первом отношении этот термин обозначает «базовую, сырую способность к чувствительности, ощущениям, опыту, субъективности, само-деятельности (*self-agency*), интенции или знанию любого рода» [108, р. 9]. В психологическом контексте под ним подразумевается «состояние осознанности (*awareness*), противопоставляемое “бессознательному” (*unconscious*), — например, состояние бодрствования и бдительности в отличие от состояния сна и сновидений». Вместе с тем, по мнению автора, нам едва ли удастся добиться концептуальной ясности в исследовании сознания, и виной тому рекурсивный характер этого понятия: нам нужно сознание, для того, чтобы размышлять и говорить о сознании.

Сразу отметим проблематичность некоторых положений де Куинси. Явления, перечисленные в его «онтологическом» определении сознания являются, по крайней мере, неочевидно однородными и даже не обязательно образуют «лестницу» или иерархию, структурированную от простого к сложному.

Чувствительность сопровождает естественную реактивность живых организмов; ощущения как будто несут некоторую информацию о внешнем мире, однако их информативность раскрывается в некоторых априорных концептуальных схемах — например, в схеме цветности или плотности. Опыт, хотя и в значительной степени основывается на чувственных данных, тем не менее, выражается в форме высказываний, которые не всегда подразумевают чувственное содержание. Понятие субъективности не более определённо, чем понятие сознания, и выражает, скорее всего, широкий спектр явлений, которые в общем и целом свидетельствуют об автономности познающего и действующего субъекта. То, что де Куинси называет само-деятельностью, имеет отношение к

тому, что со времён Канта обозначается как спонтанность — не важно, рассудка или действия. Здесь, безусловно, есть некоторое концептуальное пересечение с субъективностью, хотя в имплицитном определении последней присутствует ещё и аллюзия на возможность человеческой ошибки — в духе *errare humanum est*. В контексте настоящего исследования субъективность была определена как способность видеть возможные миры — иными словами, способность к контрфактуальному мышлению (с. 50). Интенциональность в основном понимается как свойство сознания быть «о чём-то», при этом не утихают дискуссии о (не)универсальности этого свойства для всего того, что мы объёмлем термином «сознание».

«Знание», в свою очередь, слишком обширная тема, чтобы её можно было бы осмысленно раскрыть в рамках нашего неспецифичного для неё исследования. Традиция анализа этого понятия и стоящих за ним явлений простирается в диапазоне от классической философии до современной философии науки и эпистемической логики, от социальной антропологии до когнитивной психологии и компьютерных наук. Каждая из упомянутых и многие из не упомянутых здесь областей может представить свои собственные определения знания, иногда весьма разнородные, поэтому вопрос об онтологическом статусе всех проявлений знания и об их отношении друг к другу не менее сложен, чем аналогичный вопрос относительно сознания.

Таким образом, относительно явлений, перечисленных в определении сознания, данном де Куинси, мы не можем с уверенностью утверждать ни то, что они находятся в онтологическом родстве, ни то, что они образуют какую-либо иерархическую лестницу или систему, ни даже то, что они в какой-либо мере логически предполагают друг друга. «Единство сознания» выглядит неоправданно сильным теоретическим допущением, основанным на традициях обыденного словоупотребления.

Ещё более важно и интересно, что эти соображения подрывают его тезис о «рекурсивности» сознания как помехе на пути его изучения. С учётом наших рассуждений о вероятной онтологической гетерогенности явлений сознания (или, по крайней мере, недостаточно обоснованной их гомогенности), можно сказать, что этот тезис был бы справедлив, если бы ощущения изучались ощущениями, субъективность — субъективностью и т.д., что очевидно не так. Я думаю, что эту же критику можно обратить и против популярного в феноменологии тезиса о «рефлексивности» сознания, и против поверхностного

понимания специфики социальных и гуманитарных наук, в которых субъект и объект якобы совпадают.

Автору возражает Роберт Арп [109], который исходит из различения обсуждаемых терминов по другому основанию: *awareness* он понимает как реактивную способность всех живых организмов — и в его случае это слово уже нельзя перевести на русский как «осознание», — тогда как сознание есть, по его мнению, результат «комплексного взаимодействия мультимодальных ассоциативных областей мозга животного, по крайней мере, височно-теменной и лимбической коры, и относится к способности воспринимать себя как существо, тождественное в прошлом, настоящем и будущем, включая рефлексию себя как существа, реагирующего (*aware*) на окружающую его среду» [109, р. 102–103].

Интересно, что относительно обоих обсуждаемых понятий Арп вводит «принцип реостата»: *awareness* в его понимании, как и сознание, не находятся в одном из двух возможных состояний — «выключено»/«включено», — но демонстрируют множество степеней включённости или присутствия — здесь он противопоставляет метафору реостата метафоре простого выключателя. Из этого принципа естественным образом следует, что «степень, в которой животное способно инкорпорировать распознавание себя как видимого в свете прошлого, настоящего и будущего, есть в строгом смысле степень в которой это существо можно рассматривать как скорее частично или скорее полностью сознательное. Чем больше способно животное воспринимать себя в сложных отношениях с другими «я» и объектами относительно прошлых, настоящих и будущих событий, ситуаций и сценариев, тем более сознательно это животное».

«Принцип реостата» сам по себе довольно интересен, поскольку позволяет преодолевать метафизические иллюзии относительно логически непроходимых линий или строгих дизъюнкций там, где предполагать их нет никакой необходимости. Однако, на мой взгляд, эффективность этого принципа сводится Арпом к нулю, поскольку он применяет его к неудачной онтологической схеме. Оставим пока его небесспорное понимание *awareness*, хотя не совсем понятно назначение термина, если его значение сводится к изначальной реактивности живых организмов, для которой общеупотребимый термин уже есть. Главная проблема, на мой взгляд, заключена в его понимании сознания.

В этом понимании можно выделить два тезиса:

- (1) сознание как осознание (распознавание) себя как сущности, тождественной в прошлом, настоящем и будущем, и
- (2) сознание как осознание (распознавание) себя в сложных отношениях с другими сознаниями и объектами.

В обоих случаях сознание фактически определяется как осознание себя в качестве осознающего существа: в (1) относительно временных модальностей, т.е. в диахроническом аспекте, а в (2) — относительно актуального окружения, т.е. в синхроническом аспекте.

Но подумаем, что значит «осознавать себя», например, в прошлом, настоящем и будущем. Предположим, у меня есть представление, что человек, относительно которого я помню, как в 1984 году в первой поточной аудитории Первого гуманитарного корпуса МГУ ему вручали диплом о высшем образовании, это я и есть. И зададимся вопросом: а чем это «самосознательное» представление отличается от простого фрагмента моей памяти — «я помню, как мне вручали диплом»? С точки зрения Арпа — и не только его одного — моё ретроспективное осознание представляет собой конъюнкцию двух высказываний: «я помню вручение диплома кому-то, виденное мною из перспективы первого лица» и «этим кем-то был я». Если бы истинным для меня был только первый член конъюнкции, то мне следовало бы отказаться в обладании сознанием.

Очевидно, что в тот день дипломы вручали ещё многим людям, и, возможно, у меня есть какие-то воспоминания и о них, однако понятно, что вручение диплома именно мне я помню принципиально по-другому. Возможна ли противоположная ситуация: «я помню вручение диплома кому-то, виденное мною из перспективы первого лица» и «это был не я»? Думаю, читатель легко согласится со мной, что если речь не идёт о душевном расстройстве или о фантастическом сюжете, такая ситуация грамматически исключена. Следовательно, «осознание себя в прошлом» эквивалентно просто хорошо работающей памяти.

Что же в настоящем? Верно, что я сейчас, находясь в полном сознании, пишу этот текст. Насколько важно для моего сознательного состояния осознание того факта, что пишу его именно я? Если в этот момент на мой подоконник села ворона, насколько важно, что села именно она и именно на него? С моей точки зрения, логически эти ситуации эквивалентны: если и только если существуют прагматические обстоятельства, в которых подобные уточнения имеют

смысл, то они имеют смысл. Во всяком другом случае моё утверждение «я пишу философский текст и знаю, что это делаю именно я» будет эквивалентно утверждению «я пишу философский текст и пока ещё в своём уме». Эти же рассуждения справедливы и для тезиса (2), где речь идёт о синхроническом (структурном) осознании себя.

Теперь, каким образом можно «осознавать себя в будущем»? Я знаю, что послезавтра у меня присутственный день в Институте философии. По мысли Арпа, я должен пойти туда как тот же самый человек, который сейчас пишет этот текст. Если моя память будет работать нормально, то никаких проблем с этой самоидентификацией не будет. Однако, если какие-нибудь злоеущие конкуренты сотрут мою память, а текст передадут другому лицу, то я могу перестать осознавать себя как автора этого текста, равно и как человека, которому в 1984 году вручали диплом. То есть, диахроническое единство осознания себя есть ни что иное, как хорошо работающая память.

Но всё же добавим некоторые соображения относительно тезиса (2). Речь, как мы помним, идёт о «рефлексии себя как существа, реагирующего (aware) на окружающую его среду». Если некто переходит дорогу в состоянии зомби, он или она может идти, не оглядываясь и не останавливаясь, не предпринимая других мер предосторожности, подвергая себя тем самым известному риску. Если, напротив, это лицо «включает сознание», оно выполняет все необходимые действия, по которым распознаётся сознательное существо. Значит ли это, что подключается именно осознание себя? На мой взгляд, естественнее предполагать, что подключается всё та же память, способность к обучению, представления о возможных сценариях событий, причинно-следственных связях и весь тот самый сложный комплекс психических и нейрофизиологических механизмов, наличие которых у кого-либо выдаёт в нём высокоорганизованное существо. То есть это всё та же способность правильно реагировать на внешние раздражители, которые Арп — непонятно, почему — называет awareness. И непонятно также, что принципиально меняется в субъекте и его поведении, когда к awareness добавляется ещё и рефлексия самого себя.

Таким образом, мы должны признать онтологию, предложенную Арпом, неработающей. К сожалению, она обесценивает и поддерживаемый мной «принцип реостата». Дело в том, что реостат как реальное устройство и как теоретическая идеализация предполагает принципиальную измеримость управляемых им параметров: меняя сопротивление на определённую величину



и зная напряжение, вы всегда вычислите силу тока на выходе. Если же мы провозглашаем, что в животном столько сознания, насколько оно осознаёт себя в различных контекстах, то — в каких единицах и какими приборами мы измерим это осознание? Опуская пока подробное обсуждение этой темы, скажу, по моему глубокому убеждению, философия тогда делает полезную работу, когда она помогает науке сформировать онтологию с измеряемыми свойствами и отношениями объектов.

На мой взгляд, схема, предложенная Арпом, демонстрирует то свойство классического философствования, которое Витгенштейн называл неосмысленностью (*Unsinnigkeit*). Философ предлагает концептуальные конструкции, которые, выглядя глубокими и величественными, ничего, на самом деле, не меняют ни концептуально, ни фактуально, ни методологически.

Сравнивая подходы обоих участников дискуссии, я бы отдал — с некоторыми оговорками — предпочтение определениям де Куинси, но в то же время позаимствовал бы «принцип реостата» у Арпа. Наиболее естественным было бы солидаризироваться с де Куинси в том, что сознание следует понимать как философско-онтологическую категорию, указывающую на наличие определённых способностей, а осознание — как психологический термин, связанный со способностью субъекта дать отчёт о своих психических состояниях. И добавить соображение Арпа, что наличие или отсутствие как сознания, так и осознания, описывается не булевой логикой «да»/«нет», а «принципом реостата». Однако и здесь есть некоторые тонкости, на которые следует обратить внимание.

Таким образом, сознание мы можем определить как способность к самостоятельной регуляции собственного поведения, а осознание — как состояние «подсвеченности» памятью и вниманием. Состояние, которое может менять свою локализацию и «охват» в зависимости от психического здоровья субъекта и других обстоятельств. Поэтому в русском языке мы должны различать то, что мы делаем бессознательно — например, дышим или ходим, — и то, что мы делаем неосознанно — например, рисуем, сидя на совещании. В первом случае имеется в виду неучастие сознания в соответствующих действиях. Во втором — отсутствие состояния осознанности.

Философам есть смысл заимствовать у психологии термин «осознание» и использовать его категориально — ввести с его помощью секторальность или зональность сознания и знания.

Выше я говорил о том, что философия принесла бы реальную пользу наукам о сознании, если бы предложила онтологию, основанную на объектах с измеряемыми свойствами. Так вот, теория, в которой степени осознанности соответствует степень интегрированности нейросети мозга, отвечает этому требованию: эту интегрированность можно измерить с помощью такого количественного понятия, как плотность (в теории графов) — отношения числа имеющихся сетевых связей (рёбер) к их максимально возможному количеству. С этим методологическим преимуществом, дело остаётся за эмпирическими подтверждениями связи сознательных состояний со сравнительно большей плотностью нейрональной сети, которые в некотором количестве уже имеются [110—112].

### 1.3.6 Интенциональность. Коммуникативная интенциональность

Говоря о сознании, невозможно обойти вниманием различные концепции интенциональности, играющие важную роль в современной философии и имеющие длительную историю становления в рамках феноменологической и аналитической философских традиций.

Современные дискуссии о природе интенциональности во многом восходят к этапной книге Франца Brentano «Психология с эмпирической точки зрения», где он в частности писал:

Каждый ментальный феномен характеризуется тем, что схоластики Средневековья называли интенциональным (или ментальным) несуществованием объекта — тем, что мы могли бы назвать, хотя и не вполне однозначно, ссылкой на содержание, направлением на объект (что здесь не следует путать с ситуацией, когда нечто подразумевается) или имманентной объективностью. Каждое ментальное явление включает в себе объект, хотя и не одинаковым образом. В представлении нечто представляется, в суждении нечто утверждается или отрицается, в любви нечто любимо, в ненависти — ненавидимо, в желании нечто желается, и так далее.

Это интенциональное несуществование характерно исключительно для ментальных феноменов. Никакое физическое явление не демонстрирует ничего подобного. Мы можем, таким образом, определить ментальные феномены, как явления, которые содержат объект внутри себя интенциональным образом [113, p. 88-89].

Идеи, впервые сформулированные австрийским психологом, фактически обозначили путь радикального отхода от понимания ментальных феноменов в духе сенсуализма XVIII в. в сторону поиска логических критериев того, что может быть в принципе отнесено к сознанию. Как сформулировал это Чизолм, метафизический смысл концепции Brentano «заключается в том, что интенциональные феномены имеют некоторые логические, или структурные, особенности, которые не присущи тому, что не относится к области психологического» [114, с. 40-46]. Этот новаторский подход фактически предопределил философское развитие его ученика Эдмунда Гуссерля, который считал, что существенные свойства интенциональности не зависят от существования или несуществования предмета, на который направлен интенциональный акт. Для прояснения этой позиции Гуссерль вводит два понятия: ноэма и эпохе (то есть, вывод за скобки – синоним феноменологической редукции). С помощью ноэмы Гуссерль получает доступ к внутренней структуре актов сознания. Феноменологическая редукция призвана помочь добраться до их сущности, оставив в стороне все наивные представления о различии между реальными и фиктивными сущностями [115].

В приведённой выше цитате Brentano намечает программу исследований, включающую три тезиса:

1. Ментальные состояния – любовь, желание, полагание, надежда и др. – представляют собой интенциональные феномены, чья интенциональность состоит в направленности на внешние предметы.
2. Предмет, на который направлено сознание, обладает качеством, которое Brentano называет интенциональным несуществованием.
3. Интенциональность является сущностным признаком ментального: все и только ментальные состояния обнаруживают интенциональность.

Третий тезис Brentano породил серьёзные логико-философские дискуссии относительно интенциональных объектов: есть ли такие вещи, которые не существуют? Большая часть философов аналитической традиции склонна отвечать

на этот вопрос отрицательно, но есть меньшинство, признающее реальность интенциональных объектов.

В номиналистической картине мира элементы реального неинтенционального отношения должны существовать как конкретные вещи в пространстве и времени. Если Клеопатра целует Цезаря, то они оба должны быть реальными сущностями, чтобы это отношение могло иметь место. Но если Клеопатра любит Цезаря, это требование перестаёт быть обязательным. Ведь вместо Цезаря можно любить Гарри Поттера или любую другую фиктивную сущность без вреда для этого интенционального отношения. Brentano даже предлагал считать по этой причине интенциональное состояние квазиреляционным.

Наиболее распространённым среди философов убеждением является уверенность в том, что, во-первых, интенциональные отношения должны быть поняты путём их сравнения с неинтенциональными отношениями; а во-вторых, интенциональные отношения, направленные на фиктивные сущности, должны быть поняты на фоне интенциональных же отношений, но направленных на реальные сущности в пространстве и времени.

Так, предложения:

- (1) «Аристотель был учителем Александра Македонского» и
- (2) «Стагирит был учителем Александра Македонского»

взаимозависимы в своих истинностных значениях на основании тождества значений терминов «Аристотель» и «Стагирит». То есть, если истинно (1), то истинно и (2), и наоборот. Однако если мы введём интенциональное отношение:

3) «Пётр знает, что Аристотель был учителем Александра Македонского» и

4) «Пётр знает, что Стагирит был учителем Александра Македонского», то эта взаимозависимость не сохранится, поскольку Пётр может не знать, что Аристотеля также называли Стагиритом. Более того, возникает ещё одна логическая аномалия: суждение тождества, например:

(5) «Аристотель и есть Стагирит»,

аналитическое по своей сути, может быть одновременно истинным и информативным, то есть вести себя как синтетическое.

Фреге предложил решение обеих проблем, которое оказало определяющее влияние на последующее развитие всего комплекса дисциплин, имеющих дело с

сознанием и языком [116, с. 220—246]. Это решение основано на его знаменитом различении значения (*Bedeutung*) и смысла (*Sinn*). Смысл представляет собой способ презентации значения, и если для имени определяющим является его значение, то предложения, содержащие единичные термины, идентифицируются именно своим смыслом. Следовательно, в терминах Фреге, высказывания (3) и (4) после союза «что» содержат разные предложения.

Другая проблема интенциональности связана с возможностью несуществования её объекта. Однако подобная проблема была идентифицирована и за пределами интенциональных отношений, а именно – в сфере истинных высказываний о несуществовании единичных объектов. Проблема связана с имплицитной логической интуицией, что любое высказывание о единичном объекте может иметь смысл и, следовательно, претендовать на истинность только если его объект реально существует. Тогда высказывание

(6) «Пегас не существует»

или не может говорить о Пегасе, или не может быть истинным.

Рассел в своей знаменитой статье 1905 г. «О обозначении» [71] пришёл к выводу, что на самом деле большинство имен собственных в естественных языках по своим логическим функциям являются не подлинными «логическими» собственными именами, но «замаскированными» или сокращенными определенными дескрипциями. Знаменитая расселовская формула, восстанавливающая подлинную логическую форму псевдо-единичных высказываний, содержащих псевдо-собственные имена:

(7)  $\exists x[Fx \wedge \forall y(Fy \Rightarrow y = x) \wedge Gx]$

описывает любое такое высказывание, которое истинно тогда и только тогда, когда существует один и только один элемент, являющийся одновременно F и G. Это предложение будет ложным как в случае отсутствия такого элемента, так и в случае его существования в количестве более одного.

Таким образом, согласно Расселу, решение состоит в том, чтобы принять вторую часть дилеммы: речь в этом предложении идет не о Пегасе. Высказывание (6) истинно, так как не существует индивида, который являлся бы крылатым конём. Тем самым, Рассел показал, каким образом, в терминах Brentano, чтобы мысль была истинной, носитель интенционального отношения должен существовать, а его объект – не обязательно.

В философии сознания и языка 1960-х и 1970-х годов имело место интересное отклонение от фрегеанско-расселовской парадигмы в объяснении

интенциональности, которое получило название «теория прямой референции». Если Фреге считал, что в качестве семантики единичного высказывания выступает не конкретный его референт, но абстрактный смысл, а Рассел, в свою очередь, полагал, что высказывания, выглядящие как высказывания об индивидах, в большинстве своём являются квантифицированными пропозициями, то, например, Сол Крипке отметил важное различие в поведении собственного имени и определённой дескрипции с тем же значением, выражающей несущественное свойство референта, в модальных контекстах [117]. Пример из недавнего прошлого:

(8) «Дональд Трамп мог бы быть демократом».

(9) «Президент США, избранный в 2016 году, мог бы быть демократом».

Говоря в терминах семантики возможных миров, (8) отсылает к возможным мирам, в которых Трамп был бы членом демократической партии. Что касается (9), то здесь интерпретация может быть двоякой: 9a) то же, что и (8);

(9b) эквивалентно высказыванию «Президентом США в 2016 году могла бы быть избрана Хилари Клинтон».

Понятно, что в (9a) и (9b) речь идёт о разных множествах возможных миров. На основании этого различия Крипке вводит понятие «жёсткого десигнатора» (ЖД) – термина, сохраняющего своё значение во всех возможных мирах. Таковым может быть подлинное имя собственное – например «Барак Обама» – которое во всех возможных мирах обозначает одного и того же индивида, или определённая дескрипция, отсылающая к существенному свойству – например «корень квадратный из 81» – значение которой неизменно во всех возможных мирах. Различие двух последних, согласно Крипке, состоит в том, что имя собственное является ЖД *de jure*, а определённая дескрипция – *de facto*.

Теория прямой референции подчеркивает глубокое различие между интенциональностью единичных и общих пропозиций. Согласно этой теории, функция таких языковых средств, как собственные имена, индексикалы и местоимения, состоит в том, чтобы представлять индивида в предложении или в мысли. Как сказал Дэвид Каплан, индивид попадает «ловушку» единичного предложения [118]. Теория прямой референции противостоит фрегеанско-расселовской тенденции к минимизации влияния индивидов – референтов высказываний на индивидуализацию единичных пропозиций.

Однако спор между теорией прямой референции и концепциями, восходящими к Фреге и Расселу основан на точке зрения, которую можно было

бы назвать ортодоксальной: согласно этой точке зрения, мы можем говорить только о существующих объектах, то есть об онтологических индивидах в пространстве и времени. Однако в исходной концепции Brentano заложена возможность альтернативной парадигмы, в соответствии с которой в качестве интенциональных могут выступать несуществующие или абстрактные объекты.

Теория интенциональных объектов прямо связана со вторым тезисом Brentano, говорящем об их «интенциональном несуществовании». Рассмотрим схему вывода, основанную на правиле экзистенциального обобщения для неинтенциональных отношений, интенциональных отношений к существующим объектам и интенциональных отношений к несуществующим объектам:

- (10a) «Клеопатра поцеловала Цезаря».
- (10b) «Клеопатра поцеловала нечто».
- (11a) «Клеопатра любила Цезаря».
- (11b) «Клеопатра любила нечто».
- (12a) «Древние греки поклонялись Зевсу».
- (12b) «Древние греки поклонялись чему-то».

Во всех этих парах высказывание с индексом 'b' следует из высказывания с индексом 'a' по правилу экзистенциального обобщения. Проблематичной оказывается пара (12), поскольку экзистенциальный квантор, полученный в результате вывода, принято читать «существуют такие x, что...».

- (12c) « $\exists x$  (Древние греки поклонялись x)»

Но то, чему поклонялись древние греки (Зевс), по общему мнению, не существует.

Мейнонг, ученик Brentano, в 1904 г. высказал мнение, что такие объекты, как Зевс, обнаруживают свойства, которые им приписывают, но не обнаруживают свойства существования [119]. Рассел считал такую теорию интенциональных объектов онтологически неприемлемой, поскольку она принимает объекты, не согласующиеся с законами науки [71]. Парсонс, исследуя гипотезы Мейнонга и его ученика Эрнста Малли, предложил теорию несуществующих объектов, которая основана на предположении, что существование представляет собой особый вид свойств [120]. Эта теория использует квантор " $\exists$ " который не предполагает существования. Чтобы выразить существование, в теории используется предикат «E!». Таким образом, утверждение, что есть несуществующие объекты, может быть представлено в теории Парсонса без противоречия, с помощью логической формулы

(13)  $\exists x(\neg E!x)$

Кроме того, Парсонс различает «ядерные» (nuclear) и «внеядерные» (extranuclear) свойства. К первым относятся обычные неинтенциональные свойства, и только они способствуют индивидуализации объектов. К внеядерным свойствам относятся интенциональные, модальные свойства и свойство существования. Вооружившись этим различием, Парсонсу удалось избежать возражений Рассела против наивной теории интенциональных объектов, предложенной Мейнонгом.

Несколько иной вариант такой теории предложил ученик Мейнонга Эрнст Малли: вымышленные и мифические объекты, а также такие объекты, как круглый квадрат, не демонстрируют свойства, приписываемые им, но «детерминируются» ими [121]. Нет никакого противоречия в том, что некий объект детерминирован свойствами «быть конём» и «быть крылатым», но при этом ни один из существующих индивидов не проявляет этих свойств.

Неомаллианец Залта для разработки теории абстрактных объектов использует два вида предикации – экземплификация (что соответствует «проявлению» свойств) и кодирование (что соответствует понятию детерминации у Малли) [122]. На самом деле, согласно Залта, абстрактные объекты вполне могут существовать. Тем не менее, можно сказать, что некоторые абстрактные объекты (например, Шерлок Холмс и Зевс), «не существуют» в том смысле, что ни один индивид не экземплифицирует свойства, которые они кодируют.

Теории, предложенные Парсонсом и Залтой, имеют целью дать ответы на серьёзные онтологические вопросы, поставленные Куайном в известной работе «О том, что есть» [68]. Тем не менее, по сей день многие современные философы не спешат солидаризироваться с этими теориями, оставаясь в рамках ортодоксальной парадигмы, поскольку их, по видимости, трудно примирить с онтологией современного естествознания, согласно которой в мире существуют только конкретные объекты в пространстве и времени.

Но в рамках ортодоксальной парадигмы имела место и другая попытка подойти к решению головоломок интенционального небытия, которая состояла в том, чтобы попытаться прояснить онтологические трудности на более высоком, семантическом уровне, или, по словам Куайна, «вынести обсуждение в область, где обеим сторонам легче договориться об объектах (то есть, о словах)» [123]. Семантическое восхождение позволяет подняться от разговоров о вещах к разговорам о разговорах о вещах, т. е. к словам. В современной анали-



тической философии Родерик Чизолм первым приступил к формулированию «рабочего критерия, с помощью которого мы можем отделить предложения, которые являются интенциональными, или употребляются интенционально в определенном языке, от предложениями, которые таковыми не являются» [124, р. 298]. Идея заключается в том, чтобы изучать предложения, которые сообщают об интенциональности, а не саму интенциональность. Так возникает идея связи интенциональности с интенциональностью.

В логической семантике интенциональность противоположна экстенциональности. Для последней характерны две особенности. Во-первых, если языковой контекст экстенционален, два кореферентных термина можно заметить один на другой с сохранением истинностного значения:

(14) «Утренняя звезда (Morgenstern) сияет».

(15) «Вечерняя звезда (Abendstern) сияет».

Если (14) истинно, то истинно и (15) на основании

(16) *Morgenstern*  $\equiv$  *Abendstern*  $\equiv$  *Венера*.

Во-вторых, закон экзистенциального обобщения применяется или к (14), или к (15), чтобы получить

(17) « $\exists x(x$  светит)».

В интенциональных контекстах всё не так:

(18) «Пётр считает, что утренняя звезда (Morgenstern) сияет».

(19) «Пётр считает, что вечерняя звезда (Abendstern) сияет».

Истинность (18) не всегда влечет истинность (19). Вместе с тем, лишается общезначимости и правило экзистенциального обобщения, так как, если Павел считает, что у ангелов есть крылья, из этого ещё не следует, что существуют такие вещи, относительно которых Павел считает, что у них есть крылья. Чизолм выделил три критерия интенциональности в её связи с интенциональностью:

1. Предложение сообщает об интенциональном феномене, если оно содержит особый термин, который предполагает отнесение к некоторому объекту, но это отнесение таково, что ни сам термин, ни его отрицание не означает, что предполагаемый референт этого особого термина существует или не существует. Первый критерий равнозначен признанию того, что если предложение, содержащее единичный термин, сообщает об интенциональном явлении, то оно не удовлетворяет закону экзистенциального обобщения (из  $Fa$  невозможно вывести  $\exists x(Fx)$ ).

2. Истинное предложение сообщает об интенциональном феномене, если оно содержит единичный термин «а», и если замена «а» на кореференциальный термин «b» приводит к изменению истинностного значения предложения. То есть, иными словами, для такого предложения не выполняется правило взаимозаменяемости тождественных терминов.
3. Наконец, если сложное предложение, содержащее придаточное «что»-предложение, сообщает об интенциональном феномене, то ни оно само, ни его отрицание не влечет истинность придаточного «что»-предложения.

Чизолм утверждал, что описания интенциональных или психологических явлений не могут быть сведены к описанию поведения. Интенциональность как свойство отчётов об интенциональных феноменах была призвана показать, что описания и объяснения интенциональных явлений не могут быть сведены к описаниям неинтенциональных явлений. Чизолм пришёл к этому заключению, чтобы показать правильность второго тезиса Brentano: интенциональность является существенным признаком ментального. Ради этого он *de facto* отождествил интенциональность с интенциональностью.

Вслед за Чизолмом и Куайном лингвистический подход к интенциональности был принят, например, Деннетом [125].

Однако против такого отождествления можно выдвинуть, по крайней мере, два возражения. Во-первых, знание как очевидно интенциональный феномен, в отличие от полагания, не удовлетворяет третьему критерию Чизолма: из истинности любого предложения формы «А знает, что *p*» следует истинность *p*. А, во-вторых, если мы и видение рассмотрим с точки зрения интенциональности, то выяснится, что оно не удовлетворяет ни одному из трёх критериев. Так, высказывание

(20) «Мой дед видел, что Сталин курил трубку»

не удовлетворяет первому критерию, поскольку из его истинности следует, что Сталин существовал. Оно не удовлетворяет второму критерию, поскольку из его истинности следует, в том числе

(21) «Мой дед видел, что Коба курил трубку»,

вне зависимости от того, был ли мой дед в курсе партийных кличек вождя. Наконец, оно не удовлетворяет третьему критерию, поскольку из (20), как и из (21), следует

(22) «Сталин курил трубку».

Таким образом, если видение – интенциональный феномен, то не все интенциональные феномены с необходимостью порождают интенциональные контексты. С другой стороны, можно указать на интенциональные контексты, порождаемые высказываниями, не содержащими отчётов об интенциональных феноменах, – например, модальными высказываниями, высказываниями, выражающими естественные законы и т. п. Иначе говоря, области интенциональности и интенциональности не совпадают, и, следовательно, логико-семантические техники анализа интенциональных контекстов не могут существенно помочь в лингво-философском исследовании интенциональности.

В той же работе, где Куайн солидаризируется с тезисом Чизолма о неустранимости интенционального словаря, он формулирует важную для современной философии сознания дилемму, в рамках которой предстоит или принять «неустранимость интенциональных идиом и важность особой науки об интенции» и отказаться от физикалистской онтологии, или принять физикализм и отречься от «необоснованных» интенциональных идиом и «пустой» науки об интенции [123].

Наряду со скепсисом в адрес возможного отождествления интенциональности с интенциональностью, можно подвергнуть сомнению и тезис Brentano, что только ментальные явления обладают интенциональностью, заметив, например, что свойство интенционального несуществования обнаруживают предложения естественных языков. Они имеют значение и, следовательно, как и ментальные состояния, они направлены на объекты, отличные от себя самих, некоторые из которых могут не существовать в пространстве и времени. Однако предложения естественных языков не являются ментальными феноменами.

Один из популярных ответов на это возражение состоит в том, чтобы приписать предложениям вторичный, зависимый или деградированный интенциональный статус. С этой точки зрения, предложения естественного языка сами по себе не имеют значения, если оно не присвоено им людьми, которые используют их, чтобы выразить свои мысли и передавать их другим. Высказывания заимствуют свою «производную» интенциональность у «исходной» (или «примитивной») интенциональности человеческого сознания, которые люди используют в своих целях. Если принять концепцию Фодора, предполагающую существование «языка мысли», который состоит из ментальных символов с синтаксическими и семантическими свойствами, то, возможно семантические

свойства этих символов и являются основными носителями «исходной» интенциональности.

Вопрос о том, могут ли какие-либо не-ментальные феномены демонстрировать исходную интенциональность (интенциональность первого порядка), возвращает нас к дилемме Куайна: если верен второй тезис Brentano, то нужно выбирать между ним и физикалистской онтологией. Так называемые «элиминативные материалисты» [57] решительно выбирают вторую часть дилеммы Куайна и отрицают реальность человеческих убеждений и желаний. Но, как следствие, они сталкиваются с проблемой существования человеческих артефактов как физических объектов, чьё существование зависит от воли и желания их создателей. Такие философы, как Деннет, которые отрицают различия между оригинальными и производными интенциональностями, придерживаются «инструменталистской» позиции. По их мнению, интенциональный язык не в состоянии описать или объяснить никакое реальное явление. Тем не менее, в отсутствие детальных знаний о физических законах, которые регулируют поведение физических систем, интенциональный язык выполняет полезную функцию прогнозирования поведения систем. Среди философов, придерживающихся физикалистской онтологии, лишь немногие приняли прямолинейное отрицание реальности убеждений и желаний в духе элиминативного материализма. Но и немногие из них с лёгкостью могли бы ответить на недоуменный вопрос, вызванный инструменталистской позицией: как с помощью интенциональных идиом можно делать полезные предсказания, если они не описывают и не объясняют ничего реального?

Значительное число философов-физикалистов склоняются к примирению существования интенциональности с физикалистской онтологией. Задача состоит в том, чтобы показать, по словам Деннета, что нет «непреодолимой пропасти между ментальным и физическим» [125], и что можно придерживаться физикализма и интенционального реализма одновременно. Для любого состояния обладать интенциональностью значит обладать семантическими свойствами. Следовательно, задача состоит в том, чтобы показать, как чисто физическая система может обнаруживать интенциональные состояния. Вопрос для физикалиста стоит так: любой ли не-ментальный феномен обнаруживает интенциональность?

В последние двадцать лет в аналитической философии было намечено несколько путей решения этой проблемы, которые получили общее название

«натурализация интенциональности», с целью показать, что Brentano был неправ, утверждая, что только ментальные феномены могут обнаруживать интенциональность. Одна из стратегий этого направления была предложена Дретске [126] и состояла в теоретико-информационном предположении, что любое физическое устройство, которое несет информацию, на самом деле обладает определенной степенью интенциональности. По сути, это расширение концепции «естественного значения» Грайса [127]: если «огонь» означает огонь лингвистически, то дым означает огонь естественным образом. Отпечатки пальцев несут в себе информацию о личности человека. Высота столбика ртути в термометре несет информацию о температуре. Положение иглы в гальванометре несет в себе информацию о движении электрического тока. Компас несет в себе информацию о местонахождении Северного полюса. Во всех этих случаях свойство физического устройства номологически коррелирует с некоторым физическим свойством среды обитания.

Поскольку то, что белые медведи живут на Северном полюсе, не является законом, компас не несёт информацию об их месте жительства, несмотря на то, что он несёт информацию о Северном полюсе. Если это так, то отчеты об указаниях компаса соответствуют одному из критериев интенциональности Чизолма, а именно кореферентные термины в таких отчетах не являются взаимозаменяемыми с сохранением истинности. С другой стороны, если существует закон, согласно которому изменения температуры коррелируют с изменениями атмосферного давления, то, показывая температуру, высота столбика ртути в термометре будет также указывать на атмосферное давление. Следовательно, интенциональность, порождаемая отчётами об информации от физических приборов, очевидно «слабее», чем интенциональность, порождаемая отчётами об интенциональных состояниях, поскольку первая не порождается в случаях закономерной связи явлений.

Вторым заметным предложением по решению неурегулированных проблем теоретико-информационной парадигмы был телеосемантический подход Рут Милликен [128]. Основная идея Милликен состоит в том, что открытое Brentano отношение интенционального несуществования обнаруживается в биологических функциях.

Любая биологическая цель или предназначение могут оказаться недостижимыми. Например, если функция сердца млекопитающего состоит в том, чтобы качать кровь, то оно может с этим и не справиться. Конечно, биологические

функции не обнаруживают интенциональности в смысле Brentano – они не демонстрируют свойства «быть о чём-либо». Милликен, однако, утверждает не то, что выполнение функции достаточно для интенциональности, но то, что оно необходимо. Возможно, физическое устройство не может быть «о чём-то» или представлять что-либо, пока не существует возможности, что оно может представить свой объект в искажённом свете. Предположим, для устройства представить что-либо в ложном свете значит просто работать неправильно. Но если бы у него не было никаких функций, оно бы и не могло ошибаться при их выполнении.

Если это так, то ничто не может быть репрезентативной системой, то есть иметь содержание или обнаруживать интенциональность, если оно не обладает тем, что Милликен называет «надлежащей» функцией. В соответствии с её телеосемантической теорией, «проект», создаваемый естественным отбором, является основным источником функции, которая, в свою очередь, является источником содержания или интенциональности.

Парадигматическим для этой теории интенциональности является неинтенциональный процесс естественного отбора, открытый Дарвиным. В ходе эволюции не существует интенционального объекта, который отвечал бы за отбор. Таким образом, эта теория стремится убить двух зайцев одним выстрелом. С одной стороны, она выводит интенциональность сознания из интенциональности биологических феноменов. С другой стороны, она показывает, что нормативность ментальных состояний отчасти проявляется уже на уровне биологических функций. Однако такая позиция вызывает оживлённые дискуссии в современной философии сознания. Такие философы, как Дэвидсон, Крипке, Макдауэлл и Патнем представили серьезные основания для скептицизма. «Пропозициональные установки» – восприятия, убеждения, желания, намерения – обладают интенциональностью определённого рода: они должны демонстрировать как бы обоюдостороннее соответствие сознания и мира. Чтобы прояснить эту двойственность, Элизабет Энском [129, р. 56][р. 56] предлагает рассмотреть пример со «списком покупок». Для покупателя это просто набор инструкций, который не должен быть пересмотрен в свете того, что на самом деле лежит у него в корзине. Для детектива, целью которого является определение того, что этот покупатель обычно покупает, ситуация иная. Если имеет место несоответствие между содержанием продуктовой корзины и списком, то виной тому покупатель, а не список. В этом случае детектив должен исправить свой список.

Опираясь на рассуждение Эנסком, Сёрл утверждает, что существуют два противоположных направления соответствия: для речевого акта это направление «от-слова-к-миру», для убеждений и восприятий это «от-сознания-к-миру» [130]. Функция утверждения – констатировать факт или фактическое состояние дел. Кроме того, функция убеждения и восприятия – соответствовать факту. В отличие от утверждений, приказы демонстрируют направление соответствия «от-мира-к-слову». В отличие от убеждений и представлений, направление соответствия для желаний и намерений – «от-мира-к-сознанию». И здесь самое время задаться вопросом: насколько прав Brentano и вся феноменологическая традиция в том, что все ментальные состояния обнаруживают интенциональность, если отношение между субъектом и объектом может быть разнонаправленным, а у некоторых феноменов (например, боль, эмоции) оно вообще не очевидно? Да и вообще, зачем нужно выделять некий общий критерий ментального?

Хомски утверждал, что искать критерий «сознания» или «ментального» так же странно, как искать всеобщий критерий «химического», «оптического» или «электрического» [131, р. 106]. Тот факт, что слово «ментальный» мы употребляем по отношению к таким разным феноменам, как боль и убеждение, что 5 является простым числом, может оказаться просто случайным лингвистическим обстоятельством, поскольку, как выразился Ричард Рорти, «они, кажется, не имеют ничего общего, кроме нашего нежелания называть их физическими» [132, р. 22]. Он делает вывод, что слово «ментальный» является лишь частью академической языковой игры, не имеющей научной объяснительной силы. Немногие современные философы сознания принимают ирреалистический подход Рорти к сознанию, но большинство из них признают, что если сознание существует в действительности, то возникают две проблемы: проблема интенциональности и проблема сознательного феноменального опыта, и решение первой не влечёт решения второй.

Проблему феноменального опыта часто называют «проблемой qualia», поскольку сила субъективных визуальных, аудиальных, болевых и других переживаний обладает интроспективной очевидностью. Задача философии сознания в данном случае – объяснить, по известному выражению Томаса Нагеля, «на что похоже» быть неким существом с феноменальным опытом [102]. Что такое феноменальный характер, феноменология, различные формы человеческого опыта? Мало кто из философов – не важно, физикалистов или нет –

склонен считать, что, подобно интенциональности, не-ментальные сущности могут обнаруживать феноменальный опыт. Не многие станут отрицать, что, несмотря на то, что боль, зрительные ощущения, обонятельные ощущения, слуховые ощущения являются собой очень разный опыт, тем не менее, все они обладают общим свойством, называемым «феноменальным сознанием». С общенаучной и сугубо физикалистской точек зрения, вопрос состоит в том, может ли феноменальное сознание быть объяснено в физических терминах, как результат процессов, происходящих в мозге.

Нед Блок ввёл известное различие между «сознанием доступа» (или Д-сознанием) и феноменальным сознанием (или Ф-сознанием) [133]. Состояние относится к Д-сознанию, если оно доступно для свободного использования в процессах мышления и для прямого рационального контроля действий и речи (то есть, доступно для нескольких когнитивных механизмов). В недавней работе Блок утверждал, что многие результаты когнитивных научных исследований зрительной системы подтверждают это его различие [134, р. 481-538]. Но, если проблема феноменального опыта должна быть ясно отделена от проблемы интенциональности, ключевой вопрос состоит в том, чтобы объяснить, как ментальное состояние может быть Ф-сознательным. Однако очевидно, что для интересующей нас проблемы Д-сознание является единственно релевантным. И при всех различиях интенционализма и анти-интенционализма, экстернализма и интернализма, мы согласимся с тем, что в сложной поведенческой ситуации выбора агент задействует те ресурсы сознания, которые доступны его когнитивным механизмам.

Итак, проблему интенциональности сознания мы могли бы представить как случай применения интенционального языка к сложному миру человеческих (или человекоподобных) действий – применения с целью экономии языковых средств, облегчения коммуникации и компенсации недостатка знаний о «подлинных» причинах и механизмах, определяющих поведение. Очевидно, что применимость этого языка в данном случае не зависит от интроспективной «достоверности» опыта сознательных актов на феноменальном уровне. Основа существования и функционирования концепций интенциональности — или недостаток наших знаний о подлинных причинах событий, или связанный с ними удобный способ «экономии языка»: язык интенциональных предикатов лаконичнее описывает и объясняет поведение агентов, чем условный язык физики.



Но, с другой стороны, легко показать, что вера во всеобщую причинную обусловленность, в свою очередь, также является одним из языков описания, основанным на определённых онтологических презумпциях и имеющим целью облегчение понимания при коммуникации. Поэтому теория интенциональности ни в коей мере не является средством избавления от третьей антиномии Канта.

Развиваемый здесь подход предполагает объединение менталистского функционализма с лингвистическим: такие слова, как «знать», не только не означают ничего, кроме функциональных отношений в мире, но их собственная функция в языке — это не указание на предмет или сущность, а связывание субъекта и объекта с некоторой коммуникативной модальностью. Мы также можем говорить о «коммуникативной интенциональности», что означает, что «ментальное состояние» или высказывание не только направлено на объект, но также подразумевает собеседника, который может даже не существовать <sup>7</sup>.

Я поддерживаю точку зрения Деннета, согласно которой мы используем интенциональную установку, то есть, приписываем некоторым объектам интенциональные качества, в тех случаях, когда такая стратегия объяснения является не менее эффективной, но более экономной, чем физикалистская стратегия: «Интенциональная установка – это такая стратегия интерпретации поведения объекта (человека, животного, артефакта, чего угодно), когда его воспринимают так, как если бы он был рациональным агентом, который при «выборе» «действия» руководствуется своими «верованиями» и «желаниями». Эти термины в кавычках (примененные здесь с натяжкой) позаимствованы из того, что обычно называют «народной психологией», т. е. из повседневного психологического дискурса, в котором мы участвуем, обсуждая психическую жизнь наших братьев-людей. Интенциональная установка – это позиция или точка зрения, которую мы обычно занимаем по отношению к друг другу, так что выбор интенциональной установки по отношению к чему-то иному представляется сознательной антропоморфизацией» [135, с. 33-34].

Это означает, что интенциональные состояния, которые могут быть описаны как функции социальных отношений (коммуникации), должны быть описаны таким образом, а описание остальных должно сместиться на уровень «ниже» – в нейросеть головного мозга. Таким образом, мы получим два типа интенциональных состояний и актов: супервентные <sup>8</sup> по отношению к социальной

<sup>7</sup>Подробнее концепция коммуникативной интенциональности изложена в разделе 3.5.3

<sup>8</sup>См. определение на с. 306.

коммуникации и супервентные по отношению к нейронным связям мозга. При этом для их научного объяснения достаточно в одном случае онтологии, понимающей социальную ткань как системное взаимодействие простых индивидов, а в другом – онтологии, понимающих нейроцеребральную сеть как обеспечивающую трансляцию возбуждений между простыми узлами. Ни в одном из этих случаев не требуется никакой дальнейшей редукции к теориям более низкого уровня.

### 1.3.7 Вывод из раздела 1.3

**Основные сюжеты философии сознания (осознание, интенциональность, перспектива первого лица и квалиа) имеют ограниченное влияние на философию когнитивных наук.** Сознание и осознание должны быть онтологически разведены, и поиск их нейрофизиологических коррелятов не должен направляться никакими метафизическими соображениями о единстве их природы. То же касается проблемы «я», самосознания и перспективы первого лица. Интенциональность должна получить функциональное объяснение с учётом существенной доли коллективной интенциональности в структуре человеческого сознания. Проблема квалиа может быть решена нахождением правильного «перевода» языка феноменального описания в язык научного объяснения. Задача сложна, но нет убедительных философских аргументов в пользу её неразрешимости.

## 1.4 Проблема феноменального опыта

### 1.4.1 Важность проблемы qualia

В музее, известном как Лувр, что находится в городе, известном как Париж, выставлена картина Леонардо да Винчи, известная как «Портрет госпожи Лизы дель Джокондо». Если вам удастся протолкнуться сквозь толпы туристов,

вы увидите фигуру женщины с тёмно-руссыми волосами, в зелёном складчатом платье с жёлтыми рукавами, сидящую на фоне зеленовато-голубовато-розоватого пейзажа. Всё, что составляет качественные характеристики этого, по мнению многих, величайшего произведения изобразительного искусства, навсегда останется у вас в памяти и будет время от времени вставать перед вашим мысленным взором с разной степенью детальности.

Но вот Лувр закрылся на ночь, вы пошли гулять по вечерним Елисейским полям, а холст остался в полутёмном и пустом помещении, вдали от каких бы то ни было глаз. Где в этот момент существует картина с бессмертным образом третьей жены удачливого флорентийского торговца тканями — в музее, в вашей (и не только) зрительной памяти, или нигде?

Позитивно мыслящий человек, привыкший полагаться на здравый смысл, удивится этому вопросу и увидит в нём ещё один повод посмеяться над философами. Наш человеческий Umwelt устроен таким образом, что картины остаются на своих местах в музеях, пока их не украдут или не отдадут на реставрацию. А вот, например, физик или химик уже могут задуматься. Ведь, глядя на Мону Лизу, мы видим не форму органических молекул красок и волокон холста, которые только и существуют объективно, согласно нашей научной картине мира, вполне наследующей локковское различие первичных и вторичных свойств. Мы видим «субъективный» образ, который как раз в музее не остаётся, а в несколько упрощённом виде уносится нами с собой. Леонардо создал некую физико-химическую матрицу, которая при столкновении с нашим когнитивным аппаратом порождает зрительную иллюзию, целиком состоящую из того, что философы называют квалиа. И что, собственно, образует великое произведение искусства — «объективная» матрица или «субъективная» иллюзия?

«Картина» как когнитивно-социальный комплекс имеет сложную многоуровневую онтологическую структуру. Какие-то элементы и уровни этой структуры можно отнести к достаточным условиям, какие-то — к необходимым. Очевидно, что наличие качественных восприятий — например, цвета — относится к необходимым условиям существования картины как картины: без них, в отличие от литературного произведения, произведение живописи немислимо. Следовательно, без когнитивных ресурсов, поставляемых мозгом — например, цветовосприятия — картины нет. Но её также нет и без сложного комплекса коммуникативно интерпретируемых ресурсов, которые можно отнести к социально-когнитивным: миф, общепринятая семантика символов, эмоциональные

паттерны, такие как выражения лица, и многое другое, что порождается не нейронной, а социальной сетью.

Если это было бы не так, то поздно ночью, когда Лувр закрыт, Мона Лиза переставала бы существовать. Но она продолжает существовать как социально-интенциональный объект в коллективной памяти, в коммуникационных потоках, ссылках и т.п. В случае (не дай бог) утраты её можно восстановить в физическом воплощении с той или иной степенью достоверности, подобно тому как были восстановлены разрушенные войной Гданьск и Варшава. Я бы назвал это случаем двойной интенциональности: картина как изображение указывает на некий интенциональный объект, чьё существование предполагается за пределами холста. Но и сама она существует как интенциональный объект в коллективном сознании, в какой-то степени независимый от индивидуальной материи. Например, при реставрации мы не всегда отдаём себе отчёт, да и не всегда интересуемся, какая часть материальных условий (те же холст и краски) была обновлена. Мы воспринимаем этот культурно-исторический объект как тот же самый.

А теперь представим себе ситуацию, когда в результате серии последовательных реставраций, осуществлённых по всем правилам, с сохранением состава материалов, методов и рецептов соответствующей эпохи, физическая основа заменяется полностью. Тогда поздно ночью в Лувре висят уже другие холст и краски. А картина — та же или другая?

Если она та же — а таковым, как мне кажется, будет ответ большинства — тогда возникает вопрос о природе идентичности культурного объекта. Наиболее простой и очевидный ответ на этот вопрос таков: культурный объект, предполагающий его непосредственное чувственное качественное восприятие — например, картина, — идентифицируется как интенциональный объект; т.е. он тот же самый, пока некое сознание — индивидуальное или коллективное — не видит отличий, достаточных для того, чтобы идентифицировать его как другой. И пока носители этой интенции апеллируют к нему в его меняющейся физической природе как к тому же самому. Подобно тому как мы сохраняем родственные и человеческие отношения с людьми, живые клетки которых полностью обновляются за некий период времени.

Но представим теперь, что в процессе серии реставраций специалисты начали вдруг использовать современные синтетические материалы, созданные с применением инновационных химических, нанотехнологий и компьютерного

моделирования. В результате мы получаем изображение, производящее то же самое визуальное и эмоциональное впечатление — сохранены характер и интенсивность оттенков, детали выражения лица, даже текстура холста и красок, видная в увеличительное стекло — всё то, о чём позаботился бы лучший в мире реставратор, но — на совершенно иной физической и технологической основе, невозможной в эпоху Леонардо. Можно даже для смелости воображения добавить технологии голографии и 3D-печати. До скольких процентов растает большинство, ответившее «да» в первом случае?

С другой стороны, подумаем не о живописи, а о фотографии. Произведения этого искусства, имеющие высокую историко-художественную и рыночную ценность, насколько мне известно, могут существовать более чем в одном экземпляре, но количество отпечатков строго ограничено, гарантировано автором или его законными представителями, и любой отпечаток, выходящий за эти количественные пределы, считается незаконной подделкой или легальной копией — в зависимости от обстоятельств обнаружения, — т.е. не идентифицируется как то же самое произведение искусства, хотя физически он может ничем не отличаться от зарегистрированных легальных отпечатков, и, наоборот, легальные могут быть сделаны на бумаге различных типов с применением разных химикатов. Но все легальные, например, двадцать отпечатков — это одна и та же фотография, имеющая примерно одну и ту же аукционную или рыночную цену и распределённая среди двадцати музеев и частных коллекций. И здесь, в отличие от Моны Лизы, нет проблемы ночного выпадения из бытия: в то время как в некотором музее выключен свет, и нет посетителей, в другом месте частный коллекционер с любовью рассматривает ту же самую фотографию, радуясь удачной своей инвестиции.

И совсем третья ситуация представится нам, если мы подумаем о литературных произведениях и их переводах на другие языки. Здесь физическое воплощение — форма письменных знаков, произносимые звуки при чтении вслух — может быть совершенно различным в некоторых случаях, но произведение воспринимается как одно и то же. Более того, зададимся вопросом, почему, например, авторский пересказ писателем Волковым «Волшебника страны Оз» считается отдельным литературным произведением — сказкой «Волшебник Изумрудного города», — а авторизованный перевод Заходером сказки Милна — всё тем же «Винни-Пухом». Почему в музыке в некоторых случаях исполня-

емый фрагмент считается законной вариацией известной мелодии, а в других — отдельным произведением, лишь использующим чужую тему?

И снова сам собой напрашивается ответ: идентификация произведений искусства зависит от социальных институтов, определяющих в каждом особенном случае особые правила идентификации объектов, в коих правилах качественные аспекты чувственного восприятия могут присутствовать в разной мере и играть разные роли. В одном случае достаточно отнесения экземпляра фотографии к множеству зарегистрированных легальных отпечатков, в другом эксперты всматриваются в увеличительные приборы, стараясь с помощью мелких деталей, заметных лишь профессионально тренированному глазу и мозгу, отличить единственный легальный экземпляр от множества подделок, в третьем случае всё решает традиция и устоявшееся мнение экспертного сообщества.

Кроме того, согласно Нельсону Гудману [136, р. 113–115], имеется принципиальное различие между искусствами, которые он назвал аутографическими (как, например, живопись и скульптура), и, согласно его терминологии, искусствами аллографическими (примером могут быть музыка и литература). В первом случае подлинным произведением считается один единственный экземпляр, во втором произведение создаётся в виде некоторого сценария, текста или прототипа, который может существовать в потенциально бесконечном множестве различных реализаций — исполнений, постановок, изданий и т.п. Очевидно, что, если чувственное восприятие и играет где-то решающую роль в идентификации или оценке произведения искусства, то, конечно, когда мы имеем дело с произведениями аутографических искусств. Аллографические произведения, как правило, значительно более схематичны, и качественные особенности чувственных образов там чаще отдаются на откуп производителям реализаций (исполнителям, постановщикам, издателям).

Но почему вообще может возникнуть представление, что картина существует, только пока на неё смотрят, пока длится событие обработки мозгом сигналов, принятых от глаз, и порождения чувственных субъективных — качественных — образов? Потому, будет ответ, что Моной Лизой мы считаем данный конкретный образ женщины с тёмно-русыми волосами, зелёным платьем, определённым овалом лица и перспективой на заднем фоне. И всё это поставляет нам та часть нашего когнитивного аппарата, которая видит мир качествен-

но, в форме субъективных чувственных образов, чьи свойства воспринимаются непосредственно и не могут быть переданы другому, а только пересказаны.

Тогда возникает другая философская проблема: а не видим ли мы разные картины, описывая их одними и теми же словами? Ответ на этот вопрос во многом зависит от того, считаем ли мы, что квалиа — качественные особенности чувственных образов — являются репрезентациями, т. е. представляют какие-либо свойства реальности.

### 1.4.2 Квалиа как репрезентации

Концепция, согласно которой квалиа представляют собой репрезентации (см., напр., [137] — работа, где Чалмерс не только заявляет, что квалиа суть репрезентации, но и сознание как таковое сводит к качественной *what-is-it-likeness*<sup>9</sup>), т.е. свойства интенциональных объектов, «представляющие» свойства объектов реальных, может столкнуться со следующей трудностью. «Сильный» репрезентационализм утверждает, что квалиа суть репрезентации и ничего более, природа и назначение их качественной определённости заключается в способности представлять некое содержание, отличное от них самих. Предположим, что качественный характер слышимых мною тонированных звуков — а именно их высота как чувственно воспринимаемое качество — репрезентирует некоторое свойство реальных объектов — волновых колебаний воздушной среды, а именно их частоту. Когда я слышу такие звуки, упорядоченные определённым образом по законам гармонии, как звучащие одновременно, так и следующие друг за другом, я наслаждаюсь музыкой. Более того, я знаю, что ощущение их гармоничности вызвано действительной математической гармонией частот соответствующих акустических колебаний. И у меня, действительно возникает искушение признать ощущаемую на слух высоту каждой отдельной ноты репрезентацией определённой частоты колебаний воздуха.

Однако я вынужден заметить, что существуют и иные формы репрезентации тех же колебаний — например, их графическое изображение. И, став

<sup>9</sup> «Каково-это-бытность» — не самый изящный русский перевод, но он содержит прямую аллюзию на статью Томаса Нагеля [102] «Каково быть летучей мышью?», где такое понимание сознания было закреплено в сознании читателей благодаря эффектному мысленному эксперименту.

на минуту «сильным» репрезентационалистом, я должен буду предположить, что зрелище наложенных друг на друга графиков колебаний воздуха должно пробудить во мне такое же воодушевляющее ощущение гармонии, какое я испытываю при прослушивании музыки. Но даже убеждённый репрезентационалист, я уверен, признает, что, скорее всего, этого не произойдёт. Если даже я увижу некоторую эстетическую ценность в гармонично сочетающихся графиках, это будет наслаждение совсем другого характера, нежели пир звуков, очаровывающий тонкие натуры.

Этот аргумент, как мне представляется, опровергает взгляд, согласно которому квалиа суть не более, чем репрезентации, и объяснение их качественности сводится исключительно к указанию на эту их роль. Мой пример показывает, что, независимо от того, являются ли квалиа репрезентациями в принципе, в их качественном своеобразии есть нечто, что обладает каузальной силой, не сводимой к их предполагаемой репрезентативной природе, что способно вызывать следствия, недоступные для репрезентаций того же содержания, выполненных средствами другого когнитивного модуля. А, поскольку именно неинтерпретируемое никакими иными способами качественное своеобразие этих состояний и составляет «трудную проблему сознания», и сильный репрезентационализм оказывается опровергнут моим аргументом, оставшийся «на поле» слабый вариант репрезентационализма не приближает нас к её решению. Ведь если, настаивая на том, что квалиа суть репрезентации, мы, тем не менее, признаём за ними некий нерепрезентационный качественный «остаток» — в духе концепции «ментальной краски» Хармана [138] и Блока [139], — некую часть их качественного своеобразия, которая обнаруживается интроспективно и даже способна самостоятельно участвовать в причинно-следственных отношениях, но в то же время не участвует в репрезентациях, то само указание на способность квалиа транслировать какое бы то ни было интенциональное содержание становится излишним в контексте нашей основной философской проблемы.

### 1.4.3 Аргументы от *qualia* против функционализма

По крайней мере, в двух из наиболее известных в аналитической философии мысленных экспериментах — «китайская комната» [140] и «комната Мэри»



[141] — имеется существенный элемент лукавства. Конструкция Сёрла, имеющая целью опровергнуть «сильный искусственный интеллект», строится на том, что условный американец, находящийся внутри комнаты не понимает китайского, но, следуя правилам, выдаёт осмысленные ответы на осмысленные вопросы. Вывод: понимание не функционально, а субстанциально: оно зависит от «каузальных сил» конкретной биологической субстанции и, следовательно, не может быть реализовано в другой. Однако, как мне уже доводилось показывать (см. [142], [70, с. 75–76]) , Сёрл использует иллюзионистский трюк: исподволь переклюкает наше внимание на человека в комнате, который, согласно нашей расхожей картине мира, и должен быть субъектом понимания. Тогда как в конструкции эксперимента человек в комнате — всего лишь деталь понимающей системы и не обязан сам по себе обнаруживать понимание. Подобным же образом, отдельный нейрон моего мозга, который в числе прочих возбуждается, пока я пишу этот текст, не расскажет вам о его целях и содержании.

Но эксперимент Джексона не менее, а пожалуй даже более, лукав. Его цель — опровергнуть физикализм, согласно которому всё, что в мире есть нефизического, супервентно <sup>10</sup> по отношению к физическому, и полный физический дубликат нашего мира был бы его полным дубликатом [143]. Мэри, проведшая полжизни в чёрно-белой комнате, обладает, согласно Джексону, полным знанием физики, а также физиологии цветовосприятия, поскольку — внимание! — она изучила все имеющиеся теории на эту тему. Тем не менее, выходя из комнаты и впервые видя красочный мир, она очевидно получает некоторое новое знание о нём — возражать было бы контринтуитивно. Следовательно, мир не сводится к физическому, в нём есть что-то ещё.

Но нетрудно заметить, что логика Джексона строится, по крайней мере, на одной не высказанной явно презумпции: знание о мире изоморфно самому миру. В частности, здесь предполагается, что физический мир адекватно отражён в физических теориях, а наличие необъяснимых этими теориями феноменов, таких как квалиа, указывает на их принципиально нефизический характер. Роль квалиа в этом эксперименте может быть двойкой: (1) они полагаются полностью репрезентативными и как таковые демонстрируют знание о чём-то, что не описывается физическими теориями и, следовательно, не относится к физическому миру, или (2) мы безразличны к их репрезентативности, но само их наличие опровергает физикализм.

---

<sup>10</sup>См. определение на с. 306.

Рассмотрим вариант (1). Обширная, инновационная и провокативная (в хорошем смысле слова) аргументация в пользу того, что адекватность перцептивных образов не только сомнительна, но и просто невозможна в силу эволюционных причин, содержится в [104]. Но даже не столь амбициозный наш собственный эксперимент с различными репрезентациями звука показал, что чувственные качественные феномены или вовсе не являются репрезентациями или, как минимум, содержат важную нерепрезентативную часть. Следовательно, их нельзя считать изоморфными миру знаниями о нём. Следовательно далее, невозможно утверждать, что, обрета некую новую феноменальность, Мэри узнала что-то о нефизической части мира.

Впрочем, остаётся возможность, предусмотренная вариантом (2): она не узнала, а непосредственно обрела нечто, что не является физическим. Предполагается, что квалиа сами по себе, вне зависимости от их репрезентационных возможностей, являют собой некую нефизическую часть реальности. И вот здесь мы переходим к главной уловке, скрытой в эксперименте Джексона: к понимание физического.

Джексон исходит из того, что факт знакомства Мэри со всеми имеющимися на данный момент физическими и физиологическими теориями гарантирует исчерпывающую представленность физической части мира в её сознании. И тогда любое дополнение к картине — неважно, в виде репрезентации чего-либо неучтённого в ней или в виде его непосредственного присутствия — означает существование нефизического. Можно было бы возразить в очевидном и привычном духе: любая физическая теория исторически ограничена, и если существование человечества продлится достаточно долго, мы с высокой степенью вероятности узнаем много нового и неожиданного о физическом мире. Но нужно уточнить и то обстоятельство, на которое указывал Нед Блок [144, р. 1115]: наиболее существенный прирост научного знания происходит даже не с обнаружением новых фактов, а с изменением концептуальных схем — когда, например, мы вдруг понимаем, что горение, ржавение и метаболизм суть однотипные явления, равно как и свет, звук и круги на воде. В той теоретической картине мира, с которой Мэри была знакома изначально, возможно, отсутствовали концептуальные схемы, которые позволят великим учёным будущего интегрировать квалиа в физическое описание мира.

Т.е., фокуснический трюк Джексона заключается в том, что своим экспериментом он объясняет неизвестное через неизвестное: трудная проблема

сознания состоит в том, что у нас нет хорошей естественнонаучной теории, объясняющей квалиа, и он предлагает эксперимент, которым «доказывает», что физикализм ложен, потому что... у нас нет хорошей естественнонаучной теории, объясняющей квалиа.

Интересное обсуждение «комнаты Мэри» содержится в [5, с. 184–202]. Обобщая всё множество контраргументов против эксперимента Джексона, высказанных в литературе, Дмитрий Иванов объединяет их в три принципиальные группы, суть возражений которых, по его мнению, Джексону удаётся преодолеть. Это, во-первых, мнение, согласно которому Мэри не получила нового знания, увидев цвета. Во-вторых, это интерпретация, согласно которой речь может идти об обретении новой способности, а не о новом фактуальном знании. И, наконец, третья группа возражений сводится к тому, что Мэри лишь освоила новые модусы знания старых фактов. Каким образом Джексону удаётся защититься от этих контраргументов, подробно описывается в цитируемой главе. Иванов заключает, что единственное, чего не удалось обосновать Джексону, это то, что его аргумент знания принципиально отличен от линии аргументации Нагеля [102].

Однако я думаю, что у Джексона не всё так гладко со второй группой аргументов. Он отвечает на них тем, что как теоретическое знание, имевшееся у Мэри в пределах комнаты, так и вновь полученное после освобождения, имеет пропозициональную природу, т. е., выражается истинными предложениями. Отличие заключается в том, что во втором случае это могут быть индексальные предложения типа: «Красный цвет подобен *этому*» [5, с. 191–192]. Таким образом, Мэри не только обрела новую феноменальную способность, но и получила новое знание, касающееся феноменального опыта других людей. Но такое знание можно считать новым только в том случае, если Мэри получила возможность предсказывать факты, которые не могла предсказать ранее. Например, если, не видя красного цвета воочию, она не могла утверждать, что другие люди могут уподобить пожарный гидрант помидору по этому признаку. Но это очевидно не так, по условию эксперимента. Мы помним, что, находясь в комнате, Мэри знает *все* факты относительно электромагнитного излучения и физиологии цветовосприятия. Следовательно, зная, какие спектры излучения отражают поверхности тех или иных предметов, она с не меньшей точностью, чем люди в нормальных условиях, должна предсказать, какие предметы в феноменальном восприятии должны быть окрашены в один и тот же цвет. И, следовательно,

некое индексальное предложение «Красный цвет выглядит как *это*» будет истинным или ложным для неё как до, так и после выхода из комнаты. Более того, если абсолютного дальтоника, поднаторевшего в физике и физиологии цветовосприятия, снабдить индикатором частоты отражённого излучения, его или её поведение будет неотличимо от поведения цветовосприимчивого человека.

И тогда необходимо признать, что Джексон не преодолевает вторую группу контраргументов, а его мысленный эксперимент не создаёт серьёзных проблем ни для физикализма, ни для функционализма.

#### 1.4.4 Интерпретация мысленных экспериментов

Но даже на том уровне теоретического развития, на котором находимся мы вместе с Мэри, можно прояснить некоторые закономерности взаимодействия чувственного и рационального, феноменального и дискурсивного, чтобы окончательно понять, что же произошло с девушкой.

Представим себе, что в комнате, где мы находимся, на стене сидит муха, а на лампе повисла летучая мышь. В какой-то момент открывается дверь, и в комнату входит, например, наш философский оппонент, или неважно кто. Вы слышите звук открывающейся двери; подняв взгляд, вы видите её, двери, изменённое положение и некий новый комплекс чувственно воспринимаемых признаков, который соответствует вошедшему человеку. Предположим, что это же событие замечено мухой. Она — существо с фасеточными глазами и совсем по-иному устроенным мозгом. Мы можем попытаться представить себе, каким образом это событие отразится в её чувственном мире, но, скорее всего, наша фантазия будет слишком антропоморфной. И совсем не поддаётся человеческому воображению чувственный образ того же события, созданный эхолотом летучей мыши — героини хрестоматийной статьи Томаса Нагеля [102]. Таким образом, мы имеем по крайней мере три качественно различных чувственных образа одного и того же события.

Далее, мы ничего не знаем о том, есть ли теоретики-естествоиспытатели среди мух и летучих мышей. Однако среди нас они есть. Каждый из них — физик, химик, физиолог и т. д. — опишет это событие в терминах и в каузальном контексте, соответствующих его дисциплине. В некоторых науках это

событие может оказаться нерелевантным и совсем не получить описания. Но точно так же есть события, хорошо описываемые теми или иными научными теориями, но не релевантные нашему перцептивному аппарату — проще говоря, не замечаемые нами.

То, что мы называем научным объяснением, состоит в установлении некоторой взаимосвязи между чувственным образом события и его же теоретическим описанием. Такая взаимосвязь даёт нам новый контекст, из которого становится ясно, почему данное событие произошло именно таким образом, каковы были его причины, каковы последствия — и этот контекст будет различаться в различных научных теориях. Попробуем разобраться в тех материалах, из которых мы собираем наши чувственные и словесные (или количественные) образы событий. Для создания первых мы используем цвета, «вшитые» в наш когнитивный аппарат, пространственные формы и локализации, также конструируемые известными областями мозга, слуховые ощущения и т.п. Особенности этих «строительных материалов» таковы, что их качественный характер определяется нашим физиологическим устройством и не до конца изученными когнитивными механизмами. В свою очередь, дискурсивные образы событий, поставляемые нашими теориями, состоят из символов, определяемых теми или иными символическими системами с их конвенциональным синтаксисом и такой же семантикой.

Если мы говорим о физической реальности, то, как видим, доступные нам способы её отображения не заимствуют свои «строительные материалы» из неё самой, а обладают ими в определённом смысле *a priori*. Тогда, неожиданно увидев мир в цвете, Мэри не обрела нового знания, а лишь получила в своё распоряжение некоторые новые изобразительные средства — «строительные материалы» образов реальности, притом что у неё уже имелись качественно отличные образы, относимые научной традицией к той же самой реальности. Как если бы мы вдруг получили возможность чувственно воспринимать движения кварков и бозонов, о которых знаем из физических теорий. Это событие, безусловно, важно как расширение возможностей нашего когнитивного аппарата, но оно ничего не говорит об онтологическом составе реальности.

Репрезентационалистская интерпретация (1) мысленного эксперимента Джексона предполагает, что имеющий непосредственный доступ к цветовосприятию человек знает о мире больше, чем не видящий цветов, но знакомый с теорией цветовосприятия. С моей точки зрения, приведённые выше рас-

суждения демонстрируют недостаточную обоснованность этой интерпретации, поскольку репрезентативной способностью обладают не сами качественно своеобразные элементы чувственного образа, а скорее их наблюдаемые различия, которым ставятся в соответствие символические конструкции, полагаемые репрезентациями тех же фактов и событий. Однако, даже если мы встанем на крайне репрезентационалистскую точку зрения и скажем, что качественное своеобразие элементов чувственного образа также представляет собой репрезентацию чего-либо в реальности, то в любом случае из «не-физического» характера репрезентации не следует не-физический характер репрезентируемого. И это при том, что само представление о не-физическом характере средств репрезентации основано на отсутствии объяснения для них в наших текущих физических теориях.

Онтологическая интерпретация (2) истории с Мэри предполагает, что девушке было продемонстрировано нечто, имеющееся в реальности, к чему прежде у неё не было доступа, и это нечто имеет не-физическую природу, поскольку непосредственно не описывается физическими теориями. Однако вспомним на минуту, что у Мэри до переломного события в её жизни всегда был в распоряжении альтернативный способ описания реальности — с помощью физических теорий. Последние построены из материала символических систем, которые, свою очередь, также обладают свойствами, не описываемыми физическими теориями: например, семантикой. Если неожиданное открытие квалиа ставит перед Мэри и перед Джексоном онтологические проблемы, то почему их не ставит факт существования самих физических теорий? В чём принципиальная разница между чувственными и символическими средствами отображения с точки зрения их «физичности»? Пожалуй, только в том, что символические системы мы строим сами и целенаправленно, тогда как строительные материалы чувственных образов подарены нам природой.

#### 1.4.5 (Не)выразимость и (не)репрезентативность квалиа

Старая философская дискуссия о выразимости квалиа ещё более осложняет поиски ответа на вопрос о релевантности репрезентационалистского их понимания. Нат и Панда в [145] пишут, цитируя Витгенштейна, что сказанное

об ощущении в третьем лице — это описание, а то, что говорится о нём в первом лице, есть выражение (expression). Я полагаю, что всё это не настолько строго. На приёме у врача любой из нас может описать свои ощущения в деталях. Более того, детальные описания ощущений с позиции третьего лица достаточно распространены в художественной литературе, поскольку, как представляется, по законам жанра, читатель внутренне отождествляет себя с героем. Иными словами, есть ситуации, в которых имеет смысл описывать — а не просто выражать — ощущения с позиции первого грамматического лица и детализировать «внутренний мир» кого-то, о ком говорится в третьем лице, как будто он принадлежит говорящему. Следовательно, нет непроходимых логических барьеров между языками описания ощущений в первом и третьем лице.

Можно было бы возразить, что ни в том, ни в другом случае мы не можем указать очевидные условия истинности для этих описаний. Я бы сказал, что оба случая соответствуют правилу формулирования опыта от первого лица: «Если мне кажется, что мне больно, то мне больно». Таким образом, ни пациент, ни литературный персонаж не могут ошибаться в своих ощущениях, потому что даже если они каким-то образом окажутся неправы, это не окажет никакого влияния на дальнейшее развитие событий — не сыграет каузальной роли, выражаясь метафизически. А это значит, что никакое репрезентативное содержание — реальное или предполагаемое — не определяет ни природы квалиа, ни их функциональной роли в когнитивных процессах.

Наука, как она развивалась до сих пор, использует язык описания в третьем лице. Следовательно, если нам нужно научное решение проблемы qualia — «невыразимых» феноменальных качественных конститuentов опыта, — философы должны показать, как язык описания феноменального опыта от первого лица может быть переведён в таковой же от третьего лица.

\* \* \*

Одним из источников концептуализации чувственного опыта являются объектные модели, поставляемые мозгом. Они могут быть или априорными условиями чувственности (цвет, размер, место расположения, плотность), или извлекаемы из предыдущего опыта (мы ожидаем увидеть мышь с хвостом, а хомяка без него). Другой источник — это «знания», приходящие вместе с социальной коммуникацией и языком, широко известные как культура. Они могут

быть или «грамматическими» в витгенштейнианском смысле (когда мы, например, ожидаем, что одушевлённые объекты будут вести себя по-иному, нежели неодушевлённые), или чисто эмпирическими, но переданными по социальным и культурным каналам, т.е., с помощью языка. Первый источник роднит нас с животными, второй же, как представляется, отличает нас от них.

Функциональную структуру феноменального сознания можно рассмотреть на примере схемы «сенсор — сигнал — реакция», действие которой подчинено тому, что мы называем поведенческой целесообразностью, то есть, вообще говоря, имеет некое прагматическое оправдание. Однако для этой схемы, как она реализована в человеке, существенно, что она не линейна: каждый её элемент и вся она в целом находится во взаимосвязи со множеством других сенсоров, стимулов и реакций, и прагматически оправданный результат требует интеграции всех этих элементарных взаимодействий в единую систему обеспечения целесообразного поведения организма. Поэтому возникает потребность в разрыве непосредственных причинно-следственных связей элементарных стимулов и реакций и ввода некоторых из них во временный режим ожидания. Возникает механизм «репрезентации» целостных структур данных, которые должны быть оценены, и по которым должны быть приняты управленческие решения. «Изнутри» это выглядит как повышение осознанности, при этом в большей степени задействуются память и внимание. С появлением социальной коммуникации в эту же схему включается язык. «Целостное схватывание» происходит потому, что элементарные операции не становятся предметом внимания.

Таким образом, ощущения, чувственные образы, представления, концептуальное содержание и лингвистические смыслы — это всё различные степени сложности сетевых механизмов управления индивидами и группами.

#### 1.4.6 Вывод из раздела 1.4

В этой части исследования были получены сильные аргументы против сильного репрезентационализма в отношении квалиа — убеждения в том, что их качественное своеобразие представляет некоторое внешнее содержание. Слабый репрезентационализм в отношении квалиа, признающий неинтенциональный



«остаток» в квалиа, не приближает нас к решению «трудной проблемы». Анализ наиболее известных аргументов в пользу нефизического характера квалиа показал их слабость. Общий вывод может быть сформулирован так: **качественным своеобразием обладает репрезентационный механизм, а не сам феноменальный образ.**

## 1.5 Вывод из главы 1

В главе 1 были рассмотрены проблемы философско-методологического характера, определяющие дальнейший ход исследования. На методологическом уровне было определено, что **философское исследование должно адресоваться онтологическому аспекту науки, избегая каких-либо утверждений на её теоретическом и эмпирическом уровнях.** В случае с когнитивными и социальными науками существенную помощь может оказать **номиналистическая метаонтология**, разработанная некоторыми философскими школами и признающая существование множества простых объектов, все свойства которых или сводятся к отношениям, или обеспечивают возможность вступать в отношения. Применительно к интересующим нас группам наук это означает **предпочтение взгляда на их предметы как на сетевые структуры, состоящие из узлов (нейронов) и рёбер (соединений) между ними.** Простота и лаконичность этой онтологии означает, что объяснительная функция почти целиком ложится на собственно теоретический уровень, т. е. на формальные описания законов (правил) и алгоритмов.

Было установлено, что **правильно построенная когнитивная наука, способная к междисциплинарной интеграции с социальным знанием, должна быть очищена от неоправданных философских презумпций.** Онтологический анализ помог установить, что сознание соотносится, но не зависит логически от наличия осознанных состояний. Последние, в свою очередь, не предполагают с необходимостью самоосознание.

Согласно широко распространённому в философии сознания убеждению, объяснить сознание можно, только объяснив интенциональность — способность ментальных состояний быть «о чём-то». В этой главе показано, что, во-первых, **интенциональность как свойство не исчерпывает всех явлений**

**сознания**, так как то, что делает квалиа субъективными и невыразимыми в перспективе третьего лица, не является интенциональным. Во-вторых, имеется **необходимость обратиться к коллективной интенциональности**, т. е. к интенциональным свойствам тех явлений сознания, которые имеют основу в социальном образе жизни людей.

## Глава 2. Концептуальные основы когнитивных наук

Рассмотрев в предыдущей главе общие концептуальные подходы к когнитивным функциям, сформулированные на философско-онтологическом уровне, я перехожу к анализу методологических проблем конкретных наук о сознании и познании — прежде всего, психологии и нейрофизиологии. Вторая глава посвящена методологическим последствиям когнитивной революции середины XX в., в частности, парадигмальным понятиям вычислений и репрезентаций и их современным интерпретациям.

### 2.1 Развитие психологии в направлении вычислений

#### 2.1.1 Кант и проблема научности психологии

Широко известно кантовское изречение о роли математики в науке, которое в расхожей версии звучит так: во всякой науке столько науки, сколько в ней математики. В действительности, в «Метафизических началах естествознания» содержится пространное рассуждение на этот счёт [146, с. 251-252], схему которого можно передать следующим образом. Наука в собственном смысле слова должна включать в себя не только эмпирическую («историческую»), но и «чистую» часть, где, собственно, формулируются её собственные основания и законы. Однако эта чистая часть науки не может строиться на одних только понятиях — в противном случае она будет философией природы. Наука, в отличие от последней, преследует цель познания не вообще объектов как таковых, а определённых вещей природы. А это значит, что соответствующие им понятия должны быть не только помыслены рассудком, но и «конструированы». Будучи знакомы с идеями «Критики чистого разума», мы понимаем, что под конструированием имеется в виду сопоставление понятию пространственно-временной схемы на основе априорных форм чувственности. Априорные формы чувственности — пространство и время — суть, по Канту, условия и предметы математического знания (соответственно, арифметики и геометрии). Отсюда

вывод: если наука не хочет увязнуть в исключительно «историческом» перечислении и описании фактов, но при этом также не хочет остаться бесплотной и бесплодной метафизикой, она обязана математизироваться. Надо отметить, что несмотря на сомнительность и неоднозначность некоторых посылок в этом сложном силлогизме, общий вывод оказался верным в исторической и методологической перспективе.

Однако для нашей темы важнее рассуждение, которое следует за этим. Заметив вскользь, что химия на тот момент ещё не достигла уровня науки в собственном смысле слова, поскольку «пока не найдено поддающегося конструированию понятия для химических воздействий», Кант переходит к психологии:

В еще большей мере, нежели химия, эмпирическое учение о душе должно всегда оставаться далеким от ранга науки о природе в собственном смысле, прежде всего потому, что математика неприменима к явлениям внутреннего чувства и к их законам, если только не пожелают применить к потоку внутренних его изменений закон непрерывности, однако подобное расширение познания относилось бы к тому расширению познания, которое происходит на основе математики в учении о телах, примерно так же, как учение о свойствах прямой линии относится ко всей геометрии в целом. В самом деле, чистое внутреннее созерцание, в котором должны были бы быть конструированы душевные явления, есть время, имеющее всего лишь одно измерение. Но даже в качестве систематического искусства анализа или в качестве экспериментального учения учение о душе не может когда-либо приблизиться к химии, поскольку многообразие внутреннего наблюдения может быть здесь расчленено лишь мысленно и никогда не способно сохраняться в виде обособленных [элементов], вновь соединяемых по усмотрению; еще менее поддается нашим заранее намеченным опытам другой мыслящий субъект, не говоря уже о том, что наблюдение само по себе изменяет и искажает состояние наблюдаемого предмета. Учение о душе никогда не может поэтому стать чем-то большим, чем историческое учение, и — как таковое в меру возможности — систематическое естественное учение о внутреннем чувстве, т.е. естественное описание души, но не наукой о душе, даже не психологическим экспериментальным учением [146, с. 253],.

Итак, психология никогда, согласно Канту, не может стать наукой в собственном смысле слова, поскольку к ней не приложима математика, а наукообразующая роль математики доказана предыдущим силлогизмом. Неприложимость математики к психологии обосновывается следующими обстоятельствами: (1) недостаточность времени в качестве априорной схемы для конструирования психологических понятий (аналогия с учением о прямой в его сравнении с геометрией в целом), (2) невозможность разложения потока внутреннего наблюдения на устойчивые элементы с целью их последующего синтеза, (3) искажение наблюдаемого самим актом наблюдения и (4) невозможность непосредственного наблюдения другого мыслящего субъекта и «заранее намеченных опытов» с ним. Строго говоря, к математике имеют отношение только первые два аргумента. Третий и четвёртый как бы роднят психологию с физикой элементарных частиц и квантовой механикой — поскольку там также наблюдение меняет объект, а наблюдать возможно только его «следы», но не его самого. Но есть важная оговорка о «заранее намеченных опытах» с другим субъектом: это как раз отличает психологию, как её видел Кант, от современной физики. По-видимому, здесь имеется в виду принципиальная непредсказуемость и, следовательно, непригодность для систематического экспериментального исследования какого бы то ни было субъекта, наделённого «свободной волей».

Здесь, конечно, легко обвинить Канта в поверхностности и отсутствии теоретической перспективы: как его скрупулёзный аналитический ум не подсказал ему, что анализировать и измерять можно не данные интроспекции, но реакции, а экспериментальное исследование психики вполне возможно непрямым образом? Однако тот, кто видит «после», всегда успешен в критике того, кто смотрел «до». И дело даже не в зачаточном состоянии психологии, современной Канту. Его «коперникианский переворот» был успешен, но не всеобъемлющ: он показал конструирующую роль субъекта и трансцендентальность многого из того, что до тех пор приписывалось миру, но он не избавил философию от субстанциалистской установки на спекуляции про «что» вместо исследования «как». Отсюда «априорные формы чувственности», отсюда же представление о психологическом исследовании как о наблюдении за потоком «внутреннего чувства». Отсюда же и неразрешимая проблема квалиа, до сих пор ставящая в тупик аналитическую философию сознания: если в восприятии субъекта вещь обрабатывается нефизическими качествами, то эти качества или не существуют реально

— но тогда картина мира дальтоника не должна ничем отличаться от картины мира обычного человека, — или существуют реально — но тогда физическое единство мира под угрозой.

Проблема здесь состоит во взгляде на психические явления как на то, что может «существовать» в том же смысле, в котором существует стол или электрон. При этом видении то, как я вижу зелёный цвет или как я чувствую страх, становится объектом моего самонаблюдения, а следовательно, объектом в онтологическом смысле. «Как» исподволь превращается в «что».

Справедливости ради надо сказать, что именно интроспекционистская психология начала разрабатывать строгие эмпирические методы (см. об этом [147, с. 40–52]), но вполне закономерным образом сдала позиции бихевиоризму, поскольку постепенный уход от субстанциализма характеризует путь развития науки в целом.

### **2.1.2 Качественно-субстанциальный и количественно-функциональный подходы**

На основе этого исторического экскурса можно предположить наличие двух принципиальных подходов к описанию и объяснению мира. Тот, который можно было бы назвать качественно-субстанциальным (КСП), практикуется в обыденном познании и в метафизике. Напротив, количественно-функциональный (КФП) используется в естественных науках. КСП во многом основан на естественном языке. Мне уже неоднократно случалось защищать взгляд, согласно которому естественный человеческий язык представляет собой интерфейс между двумя типами сетевых вычислительных систем: нейронной сети мозга и социальной сети, в которую включён его носитель (см. [70, с. 175–182], а также [148]). В отличие от распределённых параллельных вычислений в обеих сетях, язык имеет серийную (последовательную) структуру. Он является единственным и критически важным медиумом, посредством которого нейронные вычисления распределяются по социальной сети, в результате чего их общая эффективность повышается. Поэтому и мозг, и общество по необходимости подстраиваются под его архитектуру, эмулируя последовательные вычисления в виде «рационального мышления» индивидов, а также логики, риторики и

диалектики, с античных времён обеспечивающих существование важных социальных институтов (судебная тяжба, защита диссертации и т. п.).

Картина мира, создаваемая КСП, отражает структурную модель языка. Зависит ли эта картина от грамматических особенностей национальных языков или от более общих параметров знаковых систем, до сих пор является предметом дискуссий. Напротив, КФП не строит никаких картин. Более того, его построения часто требуют интерпретации на объектных схемах КСП — подобная интерпретация чаще всего имеется в виду, когда говорят о «физическом смысле» какой-то формулы, и подобные же интерпретации обсуждались ранее в настоящей работе как онтологические презумпции и научные онтологии в разделах 1.1.2 и 1.1.3. Но этот подход оказывается весьма инструментальным, когда речь идёт о превращении знаний в технологии. Т.е., он улавливает что-то важное относительно мира, и в этом состоит его главное преимущество. Он не даёт нам удовлетворения от понимания того, что же нам противостоит, но, по крайней мере, позволяет нам правильно и эффективно обращаться с этой вещью-в-себе.

В основе КСП лежат:

- (1) принцип тождества;
- (2) принцип непротиворечия;
- (3) различие объектов, свойств и отношений;
- (4) различие терминов по степени общности.

В основе КФП лежат:

- (1) сознательно упрощённая онтология, представляющая собой своего рода проекцию наблюдаемой реальности в логическое пространство с одним или ограниченным набором свойств;
- (2) модель, связывающая объекты, свойства и отношения этой онтологии с количественными параметрами;
- (3) система функций, выражающая основные закономерности исследуемой реальности.

### 2.1.3 Объектная и функциональная онтологии

Такие, казалось бы, совершенно разные мыслители, как Платон и Беркли, в одном отношении были солидарны: объекты — это иллюзия, существуют лишь свойства. Только у Платона свойства существуют как идеи в истинном мире, а объекты возникают как морок, как случайные и несовершенные блики, экземпляфикации идей при столкновении их с материей. Для Беркли свойства существуют как ощущения, а объекты суть необязательные дополнения к ним, поставляемые «народной» метафизикой. Однако любая онтология имеет дело с объектами, свойствами и отношениями, решая всякий раз, что мы должны полагать в качестве первых, вторых и третьих сущностей.

Но почему она заключена в эти концептуальные рамки? Конечно, нельзя сбрасывать со счетов грамматику используемого языка — но грамматику не поверхностную, с падежами и суффиксами, а глубинную, на уровне категорий. Грамматика не возникает спонтанно. Её можно рассматривать как вычислительную структуру серийного интерфейса, связывающего воедино параллельные вычисления в голове и в сообществе, имея в виду, что это трио вместе — две сети и связывающий их интерфейс — и каждый его элемент в отдельности суть продукты длительной и слепой эволюции. По-видимому, объектно-атрибутивная структура лучше соответствует потребностям коммуникации: «сущностями» оказываются поведенчески значимые конструкторы.

Функционализм может опираться на альтернативную онтологическую схему, в рамках которой всё, что есть, — это функции, соотносящие некоторые параметры, и множество элементарных абстрактных объектов, по которым пробегают переменные. Такая онтология менее интуитивна, поскольку не соответствует внутренней структуре естественного языка, но более научно продуктивна, если можно так сказать. В качестве объясняющей аналогии можно взять графические компьютерные приложения векторного типа, где изображение определяется не взаимным расположением окрашенных пикселей, как в растровых программах, а функциями, описывающими прямые и кривые линии форм, а также площади, заполняемые определённым цветом.

Нуждается ли такая онтология в объектах? В некоторой степени да, но она более индифферентна в отношении онтологических презумпций: множественны ли объекты, обладают ли они свойствами и т. п. Она может быть применима как



к плюралистическим объектным схемам типа Демокрита или Лейбница, так и к монистическим, представленным, например, у Спинозы и Гегеля. В рамках этого видения мир состоит из типовых функций, каждая из которых может быть реализована множеством разных способов. На мой взгляд, только такой подход может сохранить предмет психологии как нечто «суверенное», не сводимое к предметам нейрофизиологии, с одной стороны, и философии сознания, с другой. Наука, основанная на функциональном подходе, нуждается не в метафизике, выясняющей, «что за этим стоит», а в наиболее адекватном и точном языке описания изучаемых функций, на котором в равной степени однозначно и недвусмысленно можно было бы выражать как теоретические допущения, так и данные экспериментов и наблюдений.

Встав на эту позицию, попробуем кратко определить общие подходы к наиболее важным психологическим «предметам».

#### 2.1.4 Коммуникативное понимание знания

Самое простое и очевидное определение когниции из встреченных мною звучит как «способность помнить, думать и рассуждать» [149]. На заре когнитивной науки эти способности рассматривались в рамках «компьютерной метафоры»<sup>1</sup>, которая из-за своей исторической ограниченности навязывала серийную модель вычислений в качестве объяснительного принципа. Фонноймановская компьютерная архитектура, лежащая в основе большинства современных компьютеров (в том числе того, на котором я пишу этот текст), подсказала такие понятия когнитивной психологии, как «рабочая память», «долговременная память», «процессор» и т. п., поскольку серийные вычисления предполагают последовательную обработку строчек символов, которые откуда-то берутся, чем-то обрабатываются, и результаты куда-то записываются. Нетрудно понять, что в основе этого представления лежат общие принципы

<sup>1</sup>В данном случае само употребление слова «метафора» можно рассматривать как метафорическое. В большинстве своём пионеры когнитивной науки верили, что их предмет не метафорически описывается, а реально является вычислительным по своей природе. В каком-то отношении любая научная теория метафорична: реалисты верят, что истинная теория описывает мир как он есть; антиреалисты считают, что нам просто повезло, если какая-то теория почему бы то ни было работает.

человеческой вербальной коммуникации, которая увязывает деятельность мозга и социальные связи в единое вычислительное целое, но вовсе не обязательно отражает закономерности их специфической деятельности в формах своей собственной. Интерфейс и не должен быть похож на то, что он связывает. Иначе говоря, то, как мы анализируем, например, мышление, определяется языком как моделью, а не знаниями о подлинной нейроцеребральной деятельности. Я не исключаю, что попытки обнаружить в нейронных процессах непосредственные физиологические корреляты логико-лингвистических категорий — таких как знание — не были до сих пор успешны просто в силу неверной постановки задачи. Когда появились теории и технологии, описывающие и реализующие распределённые вычисления, основанные на них когнитивные модели продемонстрировали заметно большую реалистичность (см. об этом [70, с. 130–136]). Это говорит о том, что наивный репрезентационализм первой волны когнитологов и некоторой части философов, возможно, требует существенных уточнений. Не исключено, что мы никогда не найдём прямых физиологических коррелятов психологических категорий, поскольку их — коррелятов — просто не существует. Более адекватным представляется видение, в соответствии с которым мозг обеспечивает вычисления на более низком «программном» уровне, нежели тот, который описывается классическими когнитивными моделями.

Среда для нейронных вычислений и репрезентаций включает изменения напряжения в дендритах, нейронные спайки (потенциалы действия), нейротрансмиттеры и гормоны. Информация предположительно закодирована (репрезентирована) частотой нейронных спайков и точным их временем<sup>2</sup>. Серии спайков, которыми обмениваются нейроны внутри сети в режиме реального времени, очевидно субстрато-независимы и реализуют некоторые алгоритмы, подпадая, тем самым, под определение естественных вычислений [150; 151]. Значения удельной плотности выделяемых нейротрансмиттеров и гормонов, которые также могут рассматриваться как входящие данные, изменяются континуально, как и спайки. Но абсолютные значения последних менее функционально важны, чем их наличие или отсутствие. Таким образом, мозг может быть описан как смешанное дискретно-континуальное вычислительное устройство.

---

<sup>2</sup>Хотя экспрессия генов и мозговые ритмы, по мнению многих, также играют существенную роль.

Модель процессора, адекватного когнитивной реальности, должна включать в себя и мозг, и моторно-двигательный аппарат, и социальную сеть (и здесь рождается целый букет новых дисциплин: социальная когнитивная нейронаука, когнитивная социология, психология совместной деятельности и др. — см. об этом [152]). Состояние сети в каждый момент описывается векторной величиной, интегрирующей значения весов всех межнейронных связей. Уильям Рэмзи в старой, но основополагающей статье [153] показал, что нет необходимости рассматривать каждый такой вектор в качестве репрезентации в психологическом — или классическом когнитивистском — смысле этого слова. Скорее его стоит понимать как преходящее состояние сложной вычислительной системы, включающее в себя в том числе многочисленные взаимные репрезентации вычислительных операций, осуществляемых в различных отделах мозга. Но было бы проявлением крайне некритического научного «реализма» искать репрезентацию Моны Лизы в голове человека, стоящего напротив неё в Лувре. Точно так же вряд ли может быть точно локализована репрезентация какого-либо фрагмента знания:  $2 \times 2 = 4$ ,  $E = mc^2$ , «Шлиман открыл местоположение Трои» и т. п. Сложные визуальные восприятия и семантические связи основываются на нетривиальных комбинациях многих разнородных функций, включая мелкую моторику глаз, обработку данных в специализированных отделах мозга, участие памяти, обучение и т. п. Это многоуровневая система, где каждый низший уровень имеет формы своего «представительства» на более высоком, и наоборот.

Высказываются предположения, основанные на большом массиве эмпирических исследований, что в основе того, что мы называем интеллектом, лежит мелкомасштабная топология (*small-world topology*) нейро-церебральной сети и способность локальных ансамблей нейронов к быстрой реконфигурации [154]. В [111] эмпирически показывается, что интеллект определяется слабыми связями между модулями или регионами мозга. В свою очередь, знаменитый социолог Марк Грановеттер считал, что общество также определяется слабыми связями [155]. Можно предположить, что слабые связи и образуют другой уровень вычислений, на что указывает степень модулярности сети, согласно [111]: если узлы сбиваются в группы, между элементами которых связи более многочисленны и плотнее распределены, чем между группами, то сеть характеризуется высокой модулярностью. Это очевидно, если в каждом модуле имеется «хаб» — узел, из которого исходит слабое по весу ребро к хабу другого модуля. Другой

уровень возникает, если все узлы одного модуля слабо (через хаб) связаны с узлами другого модуля. Тогда отношение управления может быть понято как изменение весов рёбер между хабами.

Но если мы встречаем существо, по виду и поведению которого мы можем заключить, что оно обладает когнитивным аппаратом, аналогичным нашему, это существо ведёт себя как обитатель общего с нами эпистемического мира, в котором некоторое  $p$  истинно, то мы склонны сказать: он/она/оно знает, что  $p$ . Постоянство знаний обеспечивается на уровне социального процессинга<sup>3</sup>. Таким образом, в предлагаемой концепции я сочетаю философско-психологический функционализм с лингвистическим: такие слова, как «знать», не только не обозначают ничего в мира — в т. ч. в нервной системе, — кроме функциональных отношений, но и их собственная функция в языке — не обозначение внешних сущностей, а функция связывания субъекта с некоторой коммуникативной модальностью. Можно также говорить о «коммуникативной интенциональности»: ментальное состояние или суждение не только направлено на объект, но также подразумевает собеседника, который, как и объект, может даже реально не существовать.

### 2.1.5 Вычислительный образ науки

19 декабря 2019 г. в Институте философии РАН на круглом столе «Вычисления в науках о мозге и сознании: атавизм, метафора или эвристика?» в рамках регулярного семинара «Междисциплинарные проблемы когнитивных наук» [156] состоялась дискуссия о природных вычислениях — одно из мероприятий, на которых были апробированы основные положения настоящей диссертации. По результатам обсуждения были написаны и предложены к публикации в «Вопросах философии» три статьи: моя [157], А. В. Родина [158] и В. И. Шалака [159].

Я отстаивал и продолжаю отстаивать идею возможности обнаружения вычислительных процессов в самой природе. В. И. Шалак считает, что в природе существуют алгоритмические процессы, но их нельзя считать вычислительными. Он представил интересный сравнительный анализ номологического и

<sup>3</sup>См. определение на с. 306.

алгоритмического способов научного объяснения. А. В. Родин полагает, что вычисления и алгоритмы — слишком сильная метафора для естественных процессов, и в их изучении нужно ограничиться старым добрым естествознанием. В его статье, готовой к публикации, также содержится интересная аргументация против неоправданного, по его мнению, компьютеризации в естественных науках.

Согласно научно-реалистическим представлениям, мы живём в мире, где правят законы, открываемые традиционными науками. В этом бесконечном однородном пространстве камень, лишённый опоры, всегда падает под действием силы тяжести, гидроксид натрия, встретившись с соляной кислотой, всегда образует поваренную соль и воду, а товар всегда дорожает, если спрос на него превысил предложение. По широко известному мнению Карла Гемпеля [160], законы, открываемые науками, участвуют в научном объяснении в качестве первой посылки условно-категорического силлогизма:

- (1) А всегда влечёт Б (*закон*).
- (2) Имеет место А (*известный факт*), следовательно
- (3) Необходимо имеет место Б (*объясняемый факт*).

Всеобщая форма закона, таким образом, оказывается моделью, порождающей то, что мы называем объяснениями, и обеспечивающей предсказуемость мира — реальную или воображаемую, — которая вызывает чувство комфорта и безопасности у существ, вынужденных искать способы адаптации к не слишком дружественной среде. Если бы эта модель была безошибочной и действительно универсально применимой, развитие науки шло бы по кумулятивной схеме: путём накопления истинных знаний в рамках одной теории в каждой научной дисциплине. Однако взглянем на объяснение Аристотелем причины землетрясений:

Но поскольку ясно, как было ранее сказано, что испарения должны возникать и из влажного, и из сухого, то и землетрясения являются необходимым следствием существования этих испарений. Сама по себе земля сухая, но из-за дождей в ней содержится много влаги, так что, когда под действием солнца и собственного огня она нагревается, как снаружи, так и в недрах земли образуется много пневмы

, а эта [пневма] в одних случаях сплошным [поток]ом] вся вытекает наружу, в других — вся [направляется] внутрь, а иной раз делится [надвое]. Поскольку же *иначе быть не может* (курсив мой — И. М.), нам следует, пожалуй, рассмотреть теперь, какое тело более всех других способно быть источником движения [161, с. 500-501].

Пневма, как мы знаем из других его работ, — это то, кипением чего является огонь. Она опосредует взаимодействие души и тела, она же всегда стремится вверх к родственному ей эфиру. В рамках аристотелевского видения «иначе быть не может» — предложенная им концептуальная схема искренне мыслится им как единственно возможная. Почему же мы с ним не согласны? Позволю себе ещё одну длинную, но многое проясняющую цитату, но уже из нашего современника — Неда Блока:

У философа-досократика не было бы никакой возможности понять, как тепло может быть видом движения или как свет может быть разновидностью вибрации. Почему? Потому что у философа-досократика не было подходящих концепций движения, а именно понятия кинетической энергии и ее роли, или [понятия] вибрации, т. е., понятий, связанных с волновой теорией света, которые позволили бы понять, как такие разные понятия могут выражать одно и то же явление. <...> У нас часто есть несколько понятий одного и того же. Понятие света и понятие электромагнитного излучения [в диапазоне] 400–700 нм представляют одно и то же явление. Философу-досократика не хватает понятия света и соответствующей [ему] концепции вибрации (которая требует целой теории). Досократика не достаёт не только теоретического определения, но и понимания того, какие вещи [концептуально] сгруппированы с научной точки зрения [144, с. 1114-1115].

Точное замечание Блока, конечно, применимо не только к досократикам, но и к Аристотелю, и к нашим современным теориям, наивности которых будут удивляться учёные будущего, если человечество до него доживёт. Последнее соображение указывает на неточность формулировки в последнем предложении цитаты. Аристотелева концепция землетрясений группирует понятия тоже «с научной точки зрения» — точки зрения той науки, которую он создавал. Ведь

для того, чтобы увидеть, что «иначе быть может», человечество должно было накопить не только опыт открытых и воспринятых фактов, но и опыт, показывающий возможность других объяснений там, где до тех пор единственное объяснение выглядело безальтернативным.

Но что создаёт эту постоянную возможность другого взгляда? Мы привычно полагаем, что законы традиционной науки появляются как обобщения известных фактов, которые, будучи подтверждены серией экспериментов в разных условиях, становятся неотъемлемой частью картины единого и однородного мира, предсказуемого в своих основных проявлениях. И тогда удачно составленной номологической картине мира достаточно лишь пополняться новыми фактами и изредка — новыми законами, не отменяющими старые. Но реальная история науки выглядит не так.

Заметим, что в качественном рассуждении Аристотеля роль закона (по силлогистической схеме Гемпеля) выполняет универсальное положение, гласящее, что при нагревании влаги должно выделяться большое количество пневмы. И это номологическое высказывание вряд ли можно считать обобщением опыта — откуда у нас может быть опыт наблюдения ненаблюдаемой по определению субстанции? Но откуда тогда уверенность, что «иначе быть не может»?

Мой ответ уже был изложен в некоторых публикациях [64; 74]. Наука — не важно, античная, средневековая или современная, — помимо фактуальных и теоретических высказываний, содержит высказывания о типах объектов, их свойствах и отношениях. Комплекс таких высказываний в каждой научной дисциплине можно назвать *научной онтологией*<sup>4</sup> (далее — «онтология»). Онтология описывает объектную модель, на которой интерпретируются теоретические и фактуальные высказывания. В отличие от последних, её высказывания не корректно оценивать на истинность/ложность — они по природе своей конвенциональны. Именно этого не понимали Аристотель и большинство мыслителей до и после него: высказывание «Пневма есть субстанция сухая и горячая» относится к онтологии и является частью предложенной объектной модели, тогда как высказывание «При нагревании влаги выделяется большое количество пневмы» есть уже высказывание собственно теоретическое и как таковое может быть истинным или ложным — но только в рамках предложенной объектной модели. А высказывание «Пневмы не существует» означает лишь то,

<sup>4</sup>Подробнее см. 1.1.3.

что у нас появились более экономные, стройные и рациональные теории, которые «лучше» (в каком-то смысле) объясняют те же и вновь открытые факты, не прибегая при этом к постулированию ненаблюдаемых сущностей<sup>5</sup>.

В качественных теориях, типа аристотелевской, различия собственно теории и онтологии не столь очевидны, как в математизированном естествознании Нового времени. Но даже в рамках последнего далеко не сразу удалось избавиться от унаследованного от греков способа объяснения: если имеет место явление, природа которого неочевидна, естественно предположить существование ненаблюдаемой сущности — носителя свойств, обнаруживаемых в этом явлении. Так, эфир оставался объяснительным принципом в физике и космологии вплоть до начала XX века. По иронии, при том, что специальная теория относительности объясняет результат эксперимента Майкельсона–Морли, не прибегая к постулированию эфира, сам Эйнштейн не только не исключал его существования полностью, но заявлял, что эфир существует, просто не надо понимать его как нечто, состоящее из существующих во времени элементов [162]. Не менее интересные истории связаны с теплородом и флогистонем, о чём подробнее см. [163—166].

Таким образом, история науки знает как случаи, когда противоречащие друг другу теории основывались на одной и той же онтологии, так и противоположные случаи — когда одни и те же теоретические положения считались истинными в рамках разных онтологий (например, как включающих теплород, так и обходящихся без него). Мы примерно знаем, как науке удалось избавиться от некоторых онтологических излишеств: например, от теплорода — объяснив макросостояние термодинамической системы определённой статистикой её микросостояний; или от эфира — предположив, что само пространство обладает физическими свойствами (именно в этом смысле Эйнштейн считал эфир существующим). Прибегая к философским обобщениям, можно было бы сказать, что онтологии упрощаются и избавляются от слишком уж произвольно домысленных сущностей за счёт вытеснения субстанциального подхода функциональным. Субстанциальный подход за каждым объективно регистрируемым свойством заставляет предполагать его носителя. Функциональный

<sup>5</sup>Возможна ситуация, как указывает А. В. Родин, когда учёный пользуется онтологией, включающей несуществующие, по его мнению, объекты (например, эпициклы орбит), просто потому что пока нет хорошей замены. Но даже такое предвосхищение не свидетельствует о его/её способности проникать в суть вещей помимо научной абдукции.



подход допускает, что по крайней мере некоторые из свойств могут возникать как эффекты взаимодействия сущностей, ни одна из которых в отдельности данным свойством не обладает. Так, ни одна из молекул в отдельности не является ни холодной, ни тёплой. Но совокупный статистический результат их движений определяет тепловые свойства состоящего из них тела.

Однако даже сегодняшняя функционализация научных онтологий пока не помогла объяснить существование усложняющихся систем в мире, где, казалось бы, действует второе начало термодинамики, и всё должно развиваться в направлении нарастания энтропии<sup>6</sup> Как получается, что некоторые материальные системы усложняются, перестраивают свои отношения со средой, получают способность к росту, размножению, передаче наследственной информации, обретают интенциональные свойства, объединяются в сложноорганизованные группы, надстраивают уровни управления и предписывают сами себе правила поведения, не следующие ни из физических, ни из биологических законов?

Мы, например, знаем, что наличие асимметричных атомов углерода создаёт условия для формирования стереоизомеров — органических веществ, чьи физико-химические свойства зависят от пространственных структур их молекул не в меньшей степени, чем от атомного состава последних. Но что делает неизбежным появление таких молекул, на которых основано всё царство живой природы? Насколько я знаю, пока нет удовлетворительной теории, которая могла бы предложить номологическое объяснение этого перехода, хотя есть гипотезы, отсылающие к специфическим условиям, имевшимся в момент формирования Земли или даже Вселенной.

Я позволю себе предположить, что и здесь, как в случае с досократиками, Аристотелем и многими другими оставленными позади этапами в истории науки, мы имеем дело не с недостатком эмпирических знаний, а с ограничениями, налагаемыми нашими текущими концептуальными схемами.

Оставаясь в рамках нынешних стандартов научности, мы, конечно же, хотели бы избежать какой бы то ни было телеологии в объяснении возрастающей сложности некоторых систем. Если мы не склонны возвращать аристотелевскую энтелехию — а подобные решения как раз относятся к обла-

---

<sup>6</sup>Я отдаю себе отчёт в том, что очень лапидарно и поверхностно затрагиваю сложную и в высшей степени дискуссионную тему. Но дальнейшее рассуждение покажет, что изучаемая здесь проблема реальна вне зависимости от тех или иных интерпретаций теоретических вопросов физики.

сти онтологических обязательств (*commitments*), по Куайну — то мы должны объяснить эту тенденцию, исходя из текущих условий. Проблема в том, что текущий набор известных законов природы не позволяет сделать это номологически. Тогда мы можем или продолжить искать пока не обнаруженные законы — что, как представляется, и попытались сделать основатели синергетического проекта — или изменить принцип объяснения.

Объяснение — специфически человеческая процедура, сильно детерминированная психическим строем вида *Homo sapiens*. Причины, по которым нам что-либо становится ясным, что ранее таковым не было, многообразны и не всегда ясны сами по себе. Возможно, наша позитивная реакция на обретенное понимание представляет собой, как и во многих подобных случаях, биологическое вознаграждение за создание новой модели предсказаний возможных изменений важных для жизни обстоятельств. Предсказание возможно там, где имеет место повторяемость. Можно предположить, что на родовом уровне объяснение представляет собой доказательство повторяемости объясняемого. Тогда усматриваются по крайней мере два вида демонстрации повторяемости. Один из них сводится к обнаружению общего правила (закона), которое предопределяет неперенное появление интересующего нас явления в определённых обстоятельствах. Как работает этот вид объяснения, мы вкратце обсудили выше.

Но человеку можно объяснить нечто и другим способом — а именно, показать, как это можно сделать. Для того, чтобы такая демонстрация служила объяснением, она должна гарантировать, что определённая последовательность действий в определённых условиях гарантированно вызовет к жизни наш *explanandum*. Поскольку здесь мы говорим об объяснениях в естественных науках, такой способ объяснения предполагает нахождение правдоподобного схематичного описания того, как природа производит интересующее нас явление.

Схема, описывающая типичную конечную последовательность шагов, на последнем из которых достигается некоторый определённый результат, называется алгоритмом. Рецепт приготовления хека с овощами, правила умножения в столбик, репликация РНК — это всё алгоритмы, вне зависимости от того, кто или что их реализует. Так получилось исторически, что наиболее точное формальное описание алгоритмического процесса было дано в знаменитой статье Алана Тьюринга, посвящённой математической проблеме вычислимых

чисел [167]. Поэтому «алгоритм» прочно ассоциируется с «вычислением» в математическом смысле<sup>7</sup>, хотя, например, проезд перекрёстка на автомобиле в соответствии с правилами дорожного движения является алгоритмическим действием, несмотря на то что никакие математические объекты в ходе его не вычисляются. И в природе, и в культуре существует множество алгоритмических процессов, которые не являются математическими вычислениями в привычном понимании и не используют их.

Учитывая это обстоятельство, возникает соблазн остановиться на убеждении, согласно которому все вычислительные процессы являются алгоритмическими, но не наоборот. И многие, если не большинство, на нём и останавливаются. Чтобы утвердиться в правильности такого видения, нужно ясно показать видовые отличия, фиксируемые понятием вычисления, которые соответствующим образом ограничивают объём родового понятия алгоритма. Соответственно, если окажется, что таковые отличия рационально не определимы, то оба понятия придётся считать эквивалентными.

Интуитивно мы понимаем вычисление как процесс, связанный с решением математических (или, в крайнем случае, логических) задач. Поскольку математика и логика существуют только в человеческой культуре, то, согласно этой традиционной установке, вычислять может только человек или специально созданное им искусственное устройство, физические процессы которого организованы по человеческим алгоритмическим правилам таким образом, что их результат может быть интерпретирован человеком как результат вычислений. Я вижу два принципиальных свойства так понятых вычислений: их *символизм* и *телеологию*. Первый предполагает, что по крайней мере предмет вычислительных операций присутствует в процессе будучи выражен в символах — физических формах, обладающих семантикой, т. е. представляющих некое содержание, отличное от них самих. Вторая предполагает, что вычисление определяется не наличными причинами, как ординарный физический процесс, а целью, содержащейся в сознании вычислителя. При этом принципиально важно, что оба свойства не должны быть интерпретируемы в терминах [какой-либо теории] алгоритмов, поскольку в противном случае они теряют свою специфичность как отличительные признаки вычислений.

---

<sup>7</sup>На самом деле, эта ассоциация состоялась ещё в Средние века благодаря персидскому математику Аль-Хорезми, имя которого в латинской транскрипции превратилось в «алгоритм».

Итак, предположим, что вычисление в привычном математическом смысле оперирует символами, которые вычислитель должен понимать и интерпретировать, и это обстоятельство *prima facie* отличает его от не-вычислительных алгоритмических процессов. Семиотика Ч. С. Пирса содержит три трихотомии, характеризующие знак: соответственно, знак сам по себе, в его отношении к обозначаемому объекту и в его отношении к интерпретатору. В каждой из трихотомий (если опустить собственную тяжеловесную терминологию Пирса) знаки подразделяются на обозначающие через качество, через уникальное существование и через правило, т. е. конвенционально. Собственно символ есть третий член второй трихотомии, т. е. «знак, относящийся к объекту, который он обозначает в силу закона, обычно ассоциации общих идей, которая действует, чтобы заставить символ интерпретироваться как относящийся к этому объекту. Таким образом, он сам является общим типом или законом, то есть — легисайном» [168, с. 102].

Приняв эту концептуальную систему, мы должны далее исходить из того, что вычисления в привычном смысле, в отличие от прочих алгоритмических процессов, оперируют со своими предметами — символами — как с типами, а не уникальными единичными сущностями («репликами», в терминологии Пирса), и, кроме того, семантика этих символов связана с ними в силу «закона», т. е. соглашения интерпретаторов.

Полемизируя с идеей натурализации вычислений, схожую мысль высказывает Андрей Родин [158], опираясь на известное *различие типа и токена* (*type/token distinction*), также восходящее к Пирсу. По его мнению также, символы в вычислениях присутствуют и обрабатываются как типы, а не как токены — последнее характерно для не-вычислительных процессов.

Буква «А» как таковая является всегда тождественным себе типом, тогда как каждое её единичное воплощение — *токен* — может отличаться регистром (заглавная/строчная), шрифтом, особенностями почерка в рукописном написании, цветом чернил, положением на странице и т. п. И алгоритмы, которые мы интуитивно рассматриваем как вычислительные, должны — согласно этой позиции — оперировать с символами как с типами. Это обстоятельство концептуально связано с их семантическими свойствами — значение символа не должно зависеть от особенностей его физического воплощения. Если мы принимаем этот аргумент в комплексе (символ как тип плюс семантика), мы обязаны признать нетождественность алгоритмических и вычислительных процессов.

Однако этот аргумент работает только в том случае, если сами эти видовые особенности вычислений имеют неалгоритмическую природу.

Люди старшего поколения помнят, что в прошлом веке, когда нам нужно было отправить бумажное письмо обычной почтой, нам настоятельно советовали заполнить специальный пунктирный шаблон цифрами почтового индекса определённой геометрической формы, образцы которой для каждой цифры были в качестве подсказки изображены рядом на конверте. Предполагалось, что в почтовых отделениях наши конверты должны были автоматически сортироваться с помощью специального сканера, и это должно было значительно ускорять их доставку<sup>8</sup>. Очевидно, что технологии тех лет могли гарантировать правильное распознавание символа только при условии его специальной нормализации, т. е., целенаправленного устранения каких бы то ни было различий типа и токена (или «реплики»). Возможно возражение: эта машина не была вычислительной в строгом смысле. Я не уверен: она должна была уметь совершать по крайней мере одну вычислительную операцию: различать значение цифр в зависимости от регистра. Первые три цифры индекса кодировали регион, вторые — номер почтового отделения, и сортировка должна была осуществляться иерархически. И, фактически, через операции с нормализованными, но всё же — токенами.

Возьмём для примера современные компьютеры, в вычислительном статусе которых никто не усомнится. Согласно общему мнению, язык, понятный компьютеру, — это двоичное исчисление, использующее только символы «1» и «0». Однако в реальном «железе» эти символы хранятся и обрабатываются в виде определённых значений электрического напряжения в ячейках памяти [169, р. 356]. Для человека эти значения также являются типами, но для железного устройства это всего лишь конкретные свойства конкретных его частей, в зависимости от которых открываются те или иные логические вентили. Таким образом, различие типа/токена имеет место скорее в интерпретации вычислительного процесса, а не в его осуществлении, будь то сортировка писем, работа компьютера или — может быть, даже — электрохимические процессы в мозге.

Вместе с тем, это различие обретает важность в такой несомненно вычислительной задаче как распознавание письменного текста. Именно здесь

---

<sup>8</sup>Очевидцы утверждают, что в реальности эти сканеры чаще всего стояли сломанные или вообще отсутствовали, поэтому сортировка всё равно осуществлялась вручную. В этом примере-токене хорошо отражается суть советской жизни как типа.

компьютер должен суметь абстрагироваться от конкретных физических особенностей реплик символов и точно идентифицировать их тип, независимо от написания. Насколько мне известно, эта задача оказалась решаемой только с помощью компьютеров, реализующих статистические алгоритмы — искусственные нейросети или другие системы машинного обучения — т. е., машины, специально обученные этой странной для них задаче. Рискну предположить, что различие типа/токена имеет значение только для человека как биокомпьютера, использующего крайне несовершенные и энергетически неэффективные интерфейсы сетевого общения — зрение, слух, руки и гортань с языком. Малая пригодность этих устройств, созданных эволюцией с другими целями<sup>9</sup>, для реализации эффективных сетевых протоколов создаёт ситуацию, когда единицы информации порождаются с сильными отклонениями от своей нормальной формы, и получатель сообщения должен быть специально обучен энергоёмкому искусству её восстановления. Которое — восстановление — также есть алгоритмический процесс.

Вкратце этот аргумент можно резюмировать следующим образом. У нас есть утверждение, что вычисление оперирует с единицами обрабатываемой информации как с типами, а не токенами, и именно поэтому они являются символами в соответствии с определением Пирса. Я привожу пример сканера конвертов, который очевидно является вычислительным устройством, поскольку различает регистры цифр на конверте и сортирует их иерархически. Но его успешная работа может быть нарушена из-за малейшего отклонения физической формы символа от заданной нормы, т. е. из-за любого индивидуального отличия. Если некий тип включает все особенности объекта как типические, то само понятие типа теряет смысл. Например, если у вас есть устройство распознавания лиц на входе в золотохранилище, и оно позитивно реагирует только на исчерпывающий набор отличительных черт определённого лица, то доступ в хранилище будет только у одного определённого человека-токена. Разница заключается в том, что исчерпывающим образом описанное лицо — без применения клонирования — в природе встречается один только раз, а письменный знак, исчерпывающим образом определённый, может быть повторен сколько угодно — как, например, факсимиле. Таким образом, возможен вычислительный процесс, работающий со входными данными как с токенами, а следовательно символизм не является необходимым свойством вычислений.

---

<sup>9</sup>Оговорюсь: это метафора, не предполагающая телеологической интерпретации эволюции.

Более того, искусственные вычислительные устройства, способные идентифицировать тип в многообразных токенах (буквах или лицах), появились очень недавно и используют такие сложные статистические алгоритмы, что машине Тьюринга, чтобы сегодня закончить это вычисление, пришлось бы начать его в момент Большого Взрыва<sup>10</sup>. То же можно сказать о семантике. Владимир Шалак [159] предложил очень хороший модельный пример алгоритмического процесса — Золушка, перебирающая зёрнышки. Во-первых, конечно, зёрнышки в этом процессе выступают не как символы-типы, а как носители качеств «хорошее/плохое», т. е., как квалисайны<sup>11</sup> в терминологии Пирса. Но, если мы захотим, чтобы Золушка не только выполняла хозяйственную задачу в интересах мачехи, но и развлекала себя интересной сказкой с помощью того же занятия, мы можем, помимо означенных качеств, ввести дополнительные пары свойств «большое/маленькое» и, например, «ячменное/овсяное». Нам следует также спроектировать устройство, которое будет ассоциировать один набор указанных качеств с образом принца, другой — с образом тыквы, третий — с образом превращения и т. д. Тогда, изменив определённым образом алгоритм сортировки зёрен, мы превратим скучное занятие в занятное повествование. Но самое главное — мы покажем, что и приписывание значений физическим объектам тоже есть процесс алгоритмический.

Итак, аргумент от символизма вычислений уже не стоит на пути отождествления их как класса с классом алгоритмических процессов. Рассмотрим аргумент от телеологии. Подобно различию типа/токена, целеполагание может быть внутренне присуще алгоритмической системе, но может быть и частью внешней интерпретации. Внешнюю телеологическую интерпретацию можно приписать и вполне физическому не-вычислительному процессу: например, река меняет русло, *чтобы соответствовать изменившемуся рельефу местности*. В научной картине мира мы стараемся избегать телеологии, поэтому, если она внутренне присуща вычислительным процессам, то они не должны быть обнаруживаемы в природе. Мы можем приписывать телеологию любым природным процессам в модальности «как если бы». Река меняет русло, как если бы стремилась соответствовать рельефу, пчела облетает цветы, как если бы

<sup>10</sup>Т. е., чтобы не просто описать это вычисление (как тип), а реально заменить собою действующую нейросеть. Это, кстати, к вопросу о критериях отождествления вычислительных процессов как одного и того же, который я получил от Андрея Родина: отождествляются ли они как токены или всё же как типы. Крайне интересный вопрос, требующий отдельного исследования.

<sup>11</sup>Следовательно, никак не символы.

хотела помочь их оплодотворению и т. п. Искусственная нейросеть пересчитывает веса межнейронных связей, *как если бы* старалась научиться правильно различать письменные символы...

Уместно ли употребление сослагательного наклонения в последнем случае? А если это естественная нейросеть? Мальчик дёргает девочку за косички, *как если бы* хотел ей понравиться... В какой момент употребление «как если бы» становится излишним? Существует ли качественная граница (где-либо вообще и в частности) между бессознательным автоматизмом и сознательным целеполаганием, или здесь (как и везде) можно говорить только о постепенных количественных переходах?

Правильный ответ представляется следующим: сознательное целеполагание реализуется вспомогательными вычислительными процессами, сопряжёнными с основными. Эти вспомогательные процессы различаются по степени «сознательности» и выполняют функцию дополнительного внешнего контроля. Внутренняя (сопряжённая) телеологичность не делает контролируемый процесс более вычислительным, чем тот был до её подключения и, следовательно, не может быть видовым признаком, выделяющим подмножество вычислительных процессов на множестве процессов алгоритмических.

Если этот вывод верен, как и представляется мне, то смысл эпохальной статьи Тьюринга состоит в следующем: реальная или воображаемая машина, способная воспроизвести некий процесс, есть индикатор того, является ли данный процесс вычислением. «Машина» — это как раз то, что противостоит природной энтропии. Для того, чтобы воспроизвести падение камня, машина не нужна. Золушка, перебирающая зёрнышки — ничуть не худшая машина Тьюринга, чем а-машина, описанная в его статье<sup>12</sup>. Может быть, стоит признать, что «алгоритм», «вычисление», «машина» и «автомат» в каком-то отношении представляют синонимический ряд. И природа тоже умеет создавать машины.

Напротив, если вычисления необходимо предполагают участие сознания — в виде целеполагания, или как-то ещё, — то на вычислительной когнитивной науке можно ставить крест, ибо тогда *explanandum* содержится в *explanans* (как, собственно, и было на заре её формирования). Но после появления коннекционизма и, далее, искусственных нейросетей, как и всего множества

<sup>12</sup>Может быть, даже чуть лучшая благодаря дополнительному контролю со стороны «целеполагания».



био-вдохновлённых<sup>13</sup> вычислительных технологий, понятия субсимвольных, а потом и натуральных вычислений, кажется, смогли внести некоторую концептуальную ясность.

Причина трудностей с акцепцией вычислительного подхода к естественным процессам в российской академической среде кроется, в том числе, в специфике русского языка: у нас нет адекватного перевода латинского *computatio*<sup>14</sup>. У нас есть только "вы-числе-ние понимаемое, таким образом, как то, чем занимаются математики и специально созданные ими машины<sup>15</sup>. Вычислительный же подход предполагает продемонстрировать, что, по крайней мере иногда, этим занимается сама природа, и мы — как её дети и её часть.

Продолжая этимологическую аргументацию, надо отметить, что и в другом латинском варианте «вычислений» — *calculatio* — также нет «чисел» и «функций». Там есть только пифагорейская «галька» (*calculus*<sup>16</sup>), которую кротонские математики использовали как способ репрезентации чисел. Но нет никакой необходимости и это слово ограничивать арифметическим значением. В американском жаргоне 1950-х '*I calculate*' означало «я полагаю». Поэтому приверженность многих математиков и философов сугубо численному пониманию термина «вычисления» можно рассматривать как психологическую инерцию и дань традиции.

В номологическом видении мы приписываем природе «законы», а в алгоритмическом — «правила». Философская интуиция подсказывает, что, как и априорные формы Канта, всё это есть только на нашей стороне. Нам удобно (или нет) описывать мир, как если бы законы и правила в нём существовали. И тогда нет метафизической разницы между номологическим и алгоритмическим подходом. Это инструменты — как отвёртка и плоскогубцы — между которыми мы выбираем, сравнивая успешность их применения к определённым набору объясняемых фактов.

<sup>13</sup>Имеется в виду общепринятый англоязычный термин 'bio-inspired'.

<sup>14</sup>Этимология — не лучший способ аргументации, но всё же: 'putare' (лат.) — «думать», 'com-' — «со-», приставка, выражающая совместность. Я полагаю, что здесь не обязательно имеется в виду «совместное думание», а скорее комплексный вывод с учётом всех данных и соподчинённых процессов.

<sup>15</sup>Однако, по мнению, в частности, В.А. Воеводского, современная математика значительно более удалена от «числа», чем элементарная, поскольку её универсальным языком является теория множеств [170].

<sup>16</sup>Тот же корень — в слове «кальций».

В принципе и падение камня можно описать с помощью машины Тьюринга, но это будет откручиванием самореза плоскогубцами. А когда мы говорим «алгоритмы реализуются в физических системах», мы всего лишь имеем в виду, что абстрактно конструированные нами правилосообразные последовательности действий нам надо как-то интерпретировать на материале наших чувственных впечатлений, чтобы строить собственное поведение. И когда мы говорим «алгоритмическое действие запускает действие физического закона», мы всего лишь имеем в виду, что «физическую» часть нашей машины нет смысла описывать алгоритмически, поскольку до сих пор на этом уровне неплохо работало номологическое объяснение. Наука в большей степени зависит от культуры, привычек и случайно сформировавшихся традиций, чем мы иногда думаем. Законы природы формулируются или как качественные онтологические принципы — например, закон сохранения, — или в виде аналитических уравнений. В частных случаях они предписывают, например, некоторым типовым объектам скорость и траекторию движения, направление изменений, строго определённые следствия при наличии строго определённого набора причин. В общем виде они накладывают количественные ограничения на множество возможных миров: из позиции субъекта, «знающего» данный закон, достижимы только некоторые определённые возможные миры, а их многочисленные альтернативы — не достижимы.

Напротив, вычислительный подход предполагает взгляд на предмет исследования как на систему, осуществляющую вычислительные процессы. Возвращаясь к теме научных онтологий: преимущество этого подхода состоит также и в том, что вычислительное объяснение может быть основано на более простой и формальной онтологии, не предполагающей глубокого проникновения в «природу вещей». Вычислительная онтология в идеале состоит из «вычислительных примитивов» [171] — абстрактных узлов машины, выполняющих элементарные вычислительные операции. Очевидно, что эти подходы не взаимозаменяемы: с одной стороны, далеко не всякую систему или процесс есть смысл рассматривать как вычислительные хотя бы в каком-то смысле, а с другой — законы природы, конечно, налагают ограничения на множество возможных реализаций вычислительных алгоритмов. Но, в то же время, нельзя не отметить, что вычислительная интерпретация даёт определённую надежду на строгий формальный подход к исследованию вещей и процессов, — таких как

жизнь, сознание, мораль и социальность — перед которыми номологический взгляд останавливается в недоумении.

### 2.1.6 Вычислительный этап развития когнитивной психологии

Когнитивные исследования могут внести определённый вклад в исследование представления знаний в рамках (формальной) эпистемологии, поскольку, во-первых, их можно рассматривать как описательную теорию того же предмета, для которого эпистемология является нормативной теорией. И, во-вторых, понятие репрезентации играет определяющую роль в обеих дисциплинах, хотя и различается в оттенках своего значения. В свою очередь, философия сознания занимает место между обеими, исследуя их общую онтологию. Одну из релевантных философских проблем здесь составляет противостояние функционализма и натурализма, в котором сознание, понятое как разновидность вычисления, противопоставлено сознанию, понятому как причиносообразное природное явление.

О репрезентациях, на мой взгляд, имеет смысл говорить только в контексте вычислений и, следовательно, в рамках функционализма.

Вычислительные системы могут быть поняты как естественные или искусственные устройства, которые используют некоторые физические процессы на своих более низких уровнях в качестве атомарных операций для алгоритмических процессов на своих более высоких уровнях. Человеческая когнитивная система представляет собой многоуровневый механизм, в котором языковой, визуальный и другие процессоры надстраиваются над многочисленными уровнями более элементарных операций, которые в конечном итоге сводятся к элементарным нейронным спайкам. Гипотеза, отстаиваемая мною, заключается в том, что *знания представляют собой функцию не только отдельного вычислительного устройства, такого как мозг, но также и системы социальной коммуникации, которая, в свою очередь, может быть представлена как своего рода суперкомпьютер параллельно-сетевой архитектуры.* Следовательно, правдоподобный отчёт о производстве и обмене знаниями должен опираться на некоторую математическую теорию социальных вычислений, а также на теорию естественных или нейронных вычислений.

Естественный процесс может рассматриваться как вычислительный, если он придерживается некоторого алгоритма и не зависит от среды исполнения. По словам Брюса МакЛеннана, задача вычислительного подхода в науке состоит в том, чтобы изгнать «призрака в машине» из научных объяснений [151, р. 117]. В известной схеме Дэвида Марра вычисление может быть описано на уровнях (1) его цели, (2) его алгоритмов и представлений и (3) его физической реализации [172, р. 24-25]. Томмазо Поджио надстраивает марровскую схему такими уровнями, как (-1) развитие и (0) обучение, для лучшего соответствия текущим разработкам в области искусственного интеллекта [173, р. 1021]. С этим обновлением схема выглядит более актуальной для изучения естественных вычислительных процессов, как, например, происходящие в церебральных нейронных сетях.

В литературе по философии когнитивных наук в последнее время активно разрабатывается тема механистического объяснения [174—176], в отличие от динамического объяснения. Первое подразумевает описание своего предмета как системы функционально организованных компонентов. Если, как это иногда утверждается, вычислительный взгляд на природные явления требует не-тьюринговой (обобщённой — generic) теории вычислений, то в общем смысле вычисления могут быть истолкованы как последовательность состояний механизма, первое из которых считается вводом, в то время как вывод содержит все последующие. Алгоритмические правила, которые определяют переходы состояний, являются функциями переменных состояния, и они чувствительны только к различиям и вариациям внутри механизма, но не к его физической специфике. Физическая среда, способная реализовать данное естественное вычисление, должна обладать достаточным количеством степеней свободы.

\* \* \*

Таким образом, если под философией психологии мы понимаем анализ онтологической составляющей психологических теорий, то её история может быть схематически описана как переход от субстанциализма — представления об особых психических сущностях как предмете психологии — через его прямолинейное отрицание в бихевиоризме к функционалистскому пониманию психических явлений как вычислительных операций. Скептицизм Канта в отношении возможности психологии как науки, соответствующей всем критериям

научности, может быть интерпретирован как предупреждение о бесперспективности субстанционального подхода к психике.

Наука о сознании и когнитивных способностях, равно как и теория искусственного интеллекта, нуждаются сегодня в теории вычислений, которая представляла бы собой расширение машины Тьюринга, достаточное, чтобы охватить нейрофизиологические и социальные процессы. набросок такой теории, представленный здесь, включает концепцию уровней распределённой вычислительной системы, которые физически реализуются как модули сети — плотные ансамбли нейронов, между которыми — ансамблями — существуют слабые связи. Тогда сеть, образуемая связями между модулями, надстраивает уровень, более «высокий» по отношению к исходной сети, возможно, обладающий эмерджентными<sup>17</sup> свойствами. Процессы, протекающие на более высоком уровне, могут воздействовать на исходный уровень в порядке «нисходящей причинности», а вычисления на этом уровне осуществляются над репрезентациями результатов вычислений на низших уровнях. Так, целостные образы в зрительных отделах мозга формируются путём интеграции данных модуля обработки цветности и модуля обработки пространственных форм.

Усложнение структуры естественных вычислительных систем, надстраивание новых уровней и включение существующих сетей в более крупные метасети направляется естественными законами, не до конца изученными современной наукой. Предположительно, в основе этой закономерности лежит естественное стремление природных вычислительных систем к повышению эффективности вычислений путём снижения термодинамических затрат на их осуществление [177]. Но, как бы то ни было, исследования в этом направлении способны принести неожиданные и революционные результаты в широком спектре биологических, когнитивных и социальных дисциплин, включая теорию искусственного интеллекта.

### 2.1.7 Вывод из раздела 2.1

Сомнения Канта в возможности научной психологии имели основанием особенности качественно-субстанционального этапа её развития. Переход на

---

<sup>17</sup>См. определение на с. 306.

количественно-функциональные основания позволил в онтологическом плане **расширить сферу психического до пространства межсубъектной коммуникации**, а в методологическом — **продуктивно применить «вычислительную метафору» к этой науке.**

## 2.2 Основные когнитивные модели

Опорными концептами когнитивной науки в её классическую эпоху были «репрезентация» и «вычисление (компьютация, процессинг)» — т.е. сознанию приписывалась архитектура, аналогичная фон-неймановской машине: процессор, память, ввод/вывод и т.п. В сознании, согласно этому взгляду, есть структуры, представляющие (репрезентирующие) данные о внешних объектах, и есть процессоры, осуществляющие вычислительные процедуры над этими данными. Когнитивисты различных школ дискутируют относительно того, что представляют собой репрезентации и вычисления, однако данный подход можно считать достаточно общим, чтобы он мог охватить текущий мейнстрим когнитивной науки, включая некоторые разновидности коннекционизма — направления в когнитивных исследованиях, моделирующего ментальные процессы с использованием нейросетевых моделей. Вместе с тем, в текущих публикациях представлены и антирепрезентационалистские течения — такие как радикальный энактивизм [178; 179] или «динамические системы» [180; 181] — иногда определяемые как пост-когнитивистские.

Развитие когнитивных наук началось в середине XX века в качестве ответа на засилье бихевиоризма в психологии. Бихевиоризм формировался в рамках общей позитивистской тенденции, когда не было научных средств и возможностей изучать внутреннее содержание мозга и сознания. В этих условиях точная наука могла ориентироваться только на изучение наблюдаемого поведения. В 50—60-е годы XX в. появляются первые компьютерные технологии. Однако задолго до этого, ещё в XVII веке Томас Гоббс определил рассуждение как расчёт [182, с. 29–30]. В 1830-е гг. Чарльз Бэббидж высказал идею универсальной автоматической вычислительной машины. В 1936 г. Алан Тьюринг сформулировал концепцию универсальной машины, которая может вычислить любую вычислимую функцию [183].

### 2.2.1 Символизм

Когнитивная наука, появившаяся в середине 1960-х гг., возникает на основе «компьютерной метафоры». И Ноам Хомски со своей школой, и позже Джерри Фодор с последователями, считали, что над нейросетью мозга надстраивается блоковая схема, которая распадается на блоки, модули и другие составляющие архитектуры линейного компьютера. В рамках этой теории деятельность человеческого сознания рассматривается по аналогии с вычислительным устройством, совместимым с «машиной Тьюринга».

В 1976 г. Ньюэлл и Саймон выдвинули гипотезу «физической символической системы» как машины, которая развивает во времени некий набор символических структур. Такая машина (будь то человек или цифровой компьютер) «имеет необходимые и достаточные условия для общей разумной деятельности» [184, р. 116]. Один из классиков символической парадигмы в когнитивной науке Зенон Пилишин писал в 1984 г.: «Люди способны... действовать на основе репрезентаций, потому что они физически формируют такие репрезентации в виде когнитивных кодов, и потому что их поведение является каузальным следствием операций, произведённых на основе этих кодов. Поскольку это именно то, что делают компьютеры, отсюда следует, что познание является разновидностью вычисления» [185, р. xiii].

Классицистские подходы постулируют наличие принципиально идентифицируемых физических состояний когнитивной системы, взаимно однозначно соответствующих идентифицируемому ментальному состоянию. Эти репрезентации и представляют собой внутренние «символы», которыми оперирует сознание.

Итак, символизм как исторически первая форма когнитивизма исходил из достаточно прямолинейного понимания вычислений — из такого, каким оно сформировалось в компьютерной науке того времени с момента публикации знаменитой статьи Тьюринга 1936 г.: вычисление это преобразование последовательности символов в соответствии с некоторым алгоритмом. Для того, чтобы перенести это видение на человеческий когнитивный аппарат, нужно было предположить, что внутри этого аппарата есть нечто, функционально играющее роль символов, и нечто, функционально играющее роль алгоритмов. Как писал один из основоположников этого классического подхода,

Тем не менее, унитарная теория увидела в современном компьютере общего назначения важную метафору и, что, возможно, более важно, в символических языках программирования, которые показали, как единый набор принципов может охватывать широкий круг вычислительных задач. Также стало ясно, что множество вычислительных функций не ограничено, что означает, что общие принципы обработки необходимы для охвата широкого круга задач. Нет смысла создавать специальную систему для каждой мыслимой функции. [186, р. 2].

В литературе часто термины «классицизм», «символизм» и «компьютеризация» используют как синонимы, молчаливо предполагая или допуская, что тьюринговский взгляд на вычисления как на правилосообразные манипуляции с символами является единственно законным. Однако, как показывает Нир Фреско [187], помимо символьных вычислений, предпочитаемых в рамках классицизма, можно выделить субсимвольные вычисления, лежащие в основу коннекционистской парадигмы, которые, в свою очередь, могут быть понимаемы как цифровые или аналоговые в зависимости от архитектуры конструируемых нейросетей. Можно также говорить об особом роде вычислений, на концепции которых основывается вычислительная нейронаука, и которые, по мнению Фреско, не являются ни цифровыми, ни аналоговыми, а представляют собой вычисления *sui generis*.

Варианты классицизма/символизма основываются на одной из двух возможных моделей цифровых вычислений: модели формально-символьных манипуляций (ФСМ), восходящей к Тьюрингу, и модели физических символьных систем (ФСС), впервые сформулированной Ньюелом и Саймоном. Эти альтернативы не разнятся принципиально в понимании сути цифровых вычислений — несущественная разница между ними состоит в локализации уровня организации, на котором можно говорить о вычислениях, и, соответственно, в требованиях к вычислительной системе.



### 2.2.2 Коннекционизм

Однако в истории параллельно развивался другой подход к структурированию материальных объектов. Этот подход я бы назвал сетевой парадигмой. В 1736 г. Леонард Эйлер решил задачу о семи кенигсбергских мостах. В результате родилась математическая теория графов, которая до сих пор является основным математическим инструментом для анализа сетевых структур.

В середине 1980-х гг. появляется коннекционизм как альтернативное направление в когнитивной науке. Он возник как междисциплинарный подход к изучению познания, который интегрирует элементы из области искусственного интеллекта, нейрофизиологии, когнитивной психологии и философии сознания. В рамках этого подхода предполагается, что когнитивные явления можно объяснить с помощью набора общих принципов обработки информации, известной как параллельные распределенные вычисления (PDP).

С методологической точки зрения, коннекционизм — это рабочая рамка для изучения когнитивных явлений с использованием архитектуры простых процессоров, соединенных между собой с помощью взвешенных соединений. Согласно этой модели, каждый нейрон получает множество входящих сигналов от других нейронов. Нейрон интегрирует сигналы путем вычисления взвешенной суммы активации. На основании величины полной входной активации функция активации (например, пороговая функция) определяет уровень исходящей активации нейрона. Исходящая активация распространяется на последующие нейроны.

Для того чтобы сеть была обучаемой, был внедрён алгоритм обратного распространения ошибки. Алгоритм обратного распространения ошибки может быть использован для контролируемого обучения многослойной сети. Для этого типа обучения алгоритм представляет сеть, которая попарно связывает входящие паттерны возбуждения и желаемые исходящие паттерны. Алгоритм вычисляет ошибку на выходе, то есть разность между фактическим выводом сети и желаемых паттернов. Далее алгоритм обратно распространяет соответствующие сигналы ошибки через каждый слой сети. Эти сигналы ошибки используются для определения изменений веса, необходимых для достижения минимизации ошибки вывода.

Как писал в своё время Д. А. Поспелов,

“Правосторонняя машина на компьютер совсем не похожа. Она работает параллельно, используя ассоциативный принцип. В ее операциях нет четко выраженной цели, планирования на основе этих целей, программирования последовательности операций. В этой машине текут непрерывные процессы, аналогичные волновым, и конечный результат ее деятельности никогда не фиксируется в виде единственно возможного. На сегодняшний день у нас нет технических аналогов правосторонней машины. Мы не знаем, как ее моделировать, ибо пока еще слишком немного знаем об особенностях ее функционирования” [188, с. 7].

На тот момент, когда Дмитрий Александрович опубликовал этот текст, уже вышла основополагающая работа [55; 56], которая ввела в научный оборот термины “PDP” и «коннекционизм» и предопределила многие дискуссии в сфере когнитивной науки и философии сознания на два десятилетия вперед.

Группой исследователей, заложивших основы PDP, был поставлен эксперимент по обучению коннекционистской сети использованию неправильных английских глаголов в прошедшем времени. Лингвистам известен эффект так называемого *U-turn* в освоении людьми прошедшего времени неправильных глаголов. В процессе обучения коннекционистская сеть воспроизвела тот же самый эффект: то есть, в начале эксперимента сеть обнаружила хорошее знание тех форм, которые принимают неправильные английский глаголы в прошедшем времени. Зачем качество этого понимания ухудшилось. И только на последней стадии обучения оно вернулось на прежний уровень. По мнению сторонников коннекционизма, этот эксперимент показывает, что их подход в когнитивной науке указывает правильный путь, достаточно реалистично воспроизводя психологию реальных людей.

В рамках коннекционизма ментальная деятельность моделируется через распространение сигналов активации между простыми вычислительными единицами, что делает её возможной в условиях нечётких или недостаточных данных, контекстозависимых понятий и динамических репрезентаций. Под простыми вычислительными единицами имеются в виду «нейроны», которые умеют только входить в количественно измеряемые состояния активации и измерять «вес» связей друг с другом, создавая сложные сетевые конфигурации, описываемые столь же сложным математическим аппаратом. Каждая

такая конфигурация, описываемая математическим вектором, может служить репрезентацией ментального состояния. Но такая нейронная сеть, в отличие от компьютеров линейной архитектуры, практически не нуждается в предварительном программировании, а наоборот, способна к самообучению, в результате которого она становится способна на операции обобщения, классификации и прогнозирования.

Коннекционистские модели доказали свою эффективность в распознавании речи и образов, а также в исследовании памяти и процессов обучения.

Способность многочисленных узлов ранжировать входящие и исходящие связи по весу превращает стандартные программные коды, управляющие узлами, в бесконечно гибкую и самообучающуюся программу, управляющую сетью в целом. В отличие от принципов программирования, принятых для компьютеров линейной архитектуры, где все алгоритмы выполняются последовательно, и программист знает конечную цель программы (условие, при наступлении которого она завершает работу), сетевая программа не содержится ни в чьей голове в виде блоков, алгоритмов и целей. Она ориентирована на максимально эффективную адаптацию сети к изменяющимся условиям. При этом, даже если отдельные цепочки вычислений здесь осуществляются медленнее, чем в машине линейной архитектуры, сеть в целом выигрывает за счёт способности осуществлять их не только одновременно и параллельно, но и взаимообусловленно – когда в последовательности учитывается не только результат предыдущего шага, но и результаты параллельных процессов. Повторю, что такая программа вовсе не нуждается в рефлексии программиста от начального до завершающего шага. Самообучаясь, на определённом этапе она способна начать удивлять своего автора, если он существует.

Но для того, чтобы перекинуть концептуальный мостик между когнитивной наукой и философией, необходимо общее философское видение, включающее хорошо разработанный концептуальный аппарат. В то же время, по мнению Теренса Хоргана, “действительно, не существует «коннекционистской концепции сознания» в виде определённого множества фундаментальных предположений, отличающихся в конкретных, четко сформулированных отношениях от классицизма. Философы до сих пор обсуждали ряд несовместимых концепций сознания, которые могли произрастать из коннекционизма или сочетаться с ним” [189, р. 13].

Коннекционизм предложил модель, в которой нейроны и их связи играют роль репрезентирующих структур данных, а возбуждения и распространение активации нейронных ансамблей играют роль обрабатывающих алгоритмов. Коннекционистский вариант когнитивной науки, таким образом, основан на своего рода трёхместной аналогии между сознанием, мозгом и компьютером, где каждый из элементов модели может пролить новый свет на представления о других. Таким образом, в когнитивной науке нет единой вычислительной модели когнитивных способностей, поскольку различные школы в компьютерной науке и различные подходы к программированию предлагают различные модели ментальных процессов. Большинство из нас работает на компьютерах, использующих серийные процессоры, которые в каждый момент времени обрабатывают одну инструкцию (хотя и очень быстро), тогда как мозг и некоторые недавно появившиеся компьютеры основаны на идее параллельного процессинга<sup>18</sup>, который позволяет делать множество операций одновременно.

В последнее время происходит интеграция комплекса когнитивных наук с нейрофизиологией, которая порождает то, что можно было бы назвать комплексом нейрокогнитивных наук. Эта интеграция происходит отчасти в экспериментальной области, что вызвано появлением новых инструментов изучения мозга, таких как функциональная магнитно-резонансная томография, транскраниальная магнитная стимуляция и оптогенетика. Отчасти эта интеграция происходит на уровне теоретизирования, из-за прогресса в понимании того, как большие популяции нейронов могут выполнять задачи, которые до сих пор объяснялись когнитивными теориями, истолковывающими ментальную жизнь в терминах правил и понятий.

Коннекционистские сети, состоящие из простых узлов и соединений между ними, очень полезны для понимания таких психологических процессов, как видение аспектов, принятие решений, выбор объяснения, и порождение значений в процессе освоения языка. Моделирование различных психологических экспериментов показало психологическую актуальность коннекционистских моделей, которые, однако, лишь очень приблизительно соответствуют реальным нейrocеребральным сетям. Однако искусственные сети иногда используют и для тестирования высокоуровневых теорий. Так, представляется небезынтeресной попытка моделирования кантовской теории познания, в частности,

---

<sup>18</sup>См. определение на с. 306.

механизма получения синтетических суждений априори, с помощью искусственной нейронной сети, представленная в [190].

Теоретическая нейронаука представляет собой попытку разработать математические и вычислительные теории, а также модели структур и процессов в головном мозге человека и других животных. Она отличается от коннекционизма большей биологической точностью и реалистичностью в моделировании поведения большого числа нейронов, организованных в функционально значимые церебральные области. С развитием науки вычислительные модели мозга приближаются как к более реалистичным имитациям нейронов, отражающим электрический и химический аспекты их деятельности, так и к моделированию взаимодействия между различными областями мозга, таких как гиппокамп и кора головного мозга. Эти модели не являются строгой альтернативой компьютерно-математических моделей, пытающихся имитировать логический вывод, следование правилу, концептуальные структуры, аналогии и образы, но должны в конце концов соответствовать им и продемонстрировать нейрофункциональный базис когнитивных процессов.

### **2.2.3 Теория когнитивного К. В. Анохина как разновидность сетевого подхода**

Почему коннекционистский взгляд на процессы, связываемые нами с сознанием, оказывается привлекательным? По словам К. В. Анохина, «рассмотрение любого объекта как сети автоматически переводит проблему в плоскость физической и математической теории. Исследуя мозг как сеть, мы, с одной стороны, применяем к нему все современные методы экспериментальной нейронауки, а с другой — движемся в сторону теории мозга и сознания» [20, с. 85].

Теоретическая модель, предлагаемая Анохиным, интересна тем, что, надстраивая второй «этаж» над коннектомом, она на порядок увеличивает степень собственной сложности, а следовательно и объяснительную силу — постольку известно, что, согласно теореме Гёделя, менее сложная система не может объяснять более сложную. Модель когнитивного, как её описывает Анохин, выглядит следующим образом:

- (1) отдельные узлы коннектома, в том числе и довольно удалённые друг от друга, складываются в устойчивые субсетевые комплексы (коги), каждый из которых обеспечивает выполнение определённой когнитивной функции;
- (2) между когами устанавливаются сетевые связи, которые превращают их в узлы суперсети – когнитома, — которая надстраивается над собственно физической сетью нейронов;
- (3) концепция когнитома делает возможным эмпирическое изучение когнитивных функций мозга, прокладывая путь к научному решению психофизической проблемы.

Насколько я могу судить, научный потенциал теории когнитома достаточно велик. Более того, эта теория может внести вклад и в решение традиционных философских проблем, таких как проблема единства самосознания (проблема «я») или проблема причинной взаимосвязи ментального и физического. Однако существуют философские проблемы, которые, как представляется, остаются в данном случае за скобками.

Это, во-первых, комплекс проблем, связанных с так называемой «субъективной реальностью»: *qualia*, чувственный опыт, «личные смыслы» и т. п. Усложнение нейросетевой модели ничего принципиального не добавляет к извечной философской трудности: даже если будет с абсолютной точностью найдена нейронная структура, соответствующая видению, например, зелёного цвета, в ней не будет содержаться ничего, напоминающего качество этого ощущения. И если личные переживания смыслов, равно как и этические или эстетические переживания, имеют ту же «субъективную» природу (что, впрочем, не очевидно, как уже говорилось), то никакое максимально глубокое проникновение в тайны мозга не приблизит нас к решению философской проблемы невыразимости субъективного опыта.

Некоторая сложность возникает также и в связи с проблемой семантики языка: а именно, какие механизмы лежат в основе связи знака и значения. Если предположить, что значение знака сохраняется как некая постоянная конфигурация возбуждённых нейронов, постоянно сопровождающая физическое появление данного знака, то для того, чтобы этот знак мог использоваться в коммуникации, в мозгах всех его участников он должен вызывать абсолютно одинаковые сочетания активированных — неважно, нейронов или когов. Но если это эмпирически так, то возникает обоснованное предположение о врождённо-

сти языка, что во многом опровергается фактом многообразия национальных языков. Если же дело не во врождённости, то непонятно, чем обеспечивается и гарантируется абсолютное тождество нейродинамических образов конкретных значений в разных мозгах.

Ожидаемое возражение: преимущество теории когнитивного как раз и состоит в том, что у нас теперь нет необходимости говорить об обязательном тождестве физических реализаций семантических образов в нейродинамических связях. Коги могут задействовать какие угодно нейроны. Главная их задача состоит в обеспечении функционального тождества сочетаний когнитивных элементов, которые стоят за эффектом общего понимания значений знаков в коммуникации.

Но тогда возникает вопрос: а где содержится критерий этого функционального тождества? Проще говоря, если при произнесении слова «стол» в разных головах возбуждаются разные нейроны, то на основании чего мы можем утверждать, что это работают одни и те же коги? Если только на основании тождества самого знака, то мы просто выдаём *explanandum* за *explanans*.

Как мне кажется, теория когнитивного в принципе способна справиться с этой трудностью. Но для этого необходимо некоторое её расширение. Идея этого расширения пришла ко мне вместе с метафорой, которую употребил Д. Деннет в своей лекции в МГУ в 2012 г. Тогда он сказал, что язык устанавливает в наш мозг некоторые небольшие программы, подобные тем *java*-апплетам, которые наш браузер скачивает из интернета для выполнения некоторых *ad hoc* задач, — в этом и состоит то, что мы называем сознанием. Я бы немного видоизменил эту компьютерную метафору. Любая большая и сложная программа, предоставляющая высоко востребованные сервисы, как, например, поисковая система Google, рано или поздно сталкивается с желанием многих сторонних программистов использовать эти сервисы в своих продуктах. И тогда авторы этой программы создают специальный программный интерфейс — набор функций и библиотек, которые могут быть использованы во внешних программах для доступа к желанным сервисам. Подход, который я хотел бы предложить, состоит в том, что элементы языка — это не временные апплеты для выполнения *ad hoc* задач, как можно подумать при поверхностной интерпретации слов Деннета, но имена более или менее постоянных функций интерфейса, необходимо существующего между нейросетью мозга и человеческой сетью общества.

Для пояснения этой метафоры, возможно, потребуется в каком-то смысле скрестить позднего Витгенштейна с современной нейронаукой. Витгенштейн, как известно, считал, что значение слова есть его употребление, осмысленное употребление обеспечивается социально санкционированными правилами, а следование правилу есть социальный институт. Если общество, в свою очередь, понимать как сеть — что предполагается и поддерживается многими направлениями современной социологии, — то институты, связанные с хранением и обработкой языковых значений, феноменов культуры и т. п., можно представить как когнитивную надстройку над социальной сетью, своего рода социальный когнитом. И тогда слова и выражения языка выглядят своего рода двунаправленным интерфейсом, обеспечивающим доступ нейросети к жизненно важным для организма функциям социальной сети, и, наоборот, доступ социальных институтов к социально важным кодам головного мозга. Для этого язык — а точнее говоря, вся социально организованная практика его создания, освоения и употребления — устанавливает в мозг особые «апплеты», действие которых можно сравнить с символическими ссылками («ярлыками», если пользоваться жаргоном Windows): они функционально связывают перцепции слов как физических объектов с кодами, кодирующими социальные взаимодействия.

#### 2.2.4 Воплощённый разум и «динамические системы»

«Воплощённый» (situated) подход отказывается от классицистского понимания интеллекта как абстрактного, индивидуального, рационального и оторванного от восприятия и действия, противопоставляя этому понимание его как отелесненного (embodied), встроенного (embedded) и распределённого (distributed) [30; 60]. Иными словами, когнитивные процессы протекают не в мозге, а между мозгом, остальным телом и средой. Классицистские подходы в ИИ пытаются создать искусственного эксперта или искусственного физика, тогда как начинать следовало бы с уровня «интеллигентности» насекомых, с тем чтобы на основе этих работающих моделей перейти к воссозданию языка и абстрактного мышления. Научиться завязывать шнурки машине труднее, чем научиться решать математические задачи или играть в шахматы, а ведь эволюции понадобились миллиарды лет для появления чувствительности и мо-



бильности и только миллионы лет для выработки собственно человеческих способностей.

Динамический [180] подход и подход в рамках воплощённого интеллекта я также объединил бы в одно семейство. Первый предполагает описание когнитивной деятельности в терминах «пространства» состояний системы, определяемых переменными, зависимыми от времени. Следствием такого подхода становится замена теоретического языка, связанного с вычислениями и репрезентациями, языком, связанным с геометрией и динамическими состояниями. Обобщая, можно было бы сказать, что в динамических объяснениях возрастает роль математики за счёт роли логики.

Я вижу основания объединить последние два подхода в одну группу. Их общие особенности:

- принципиально антиклассицистская позиция — отказ от метафоры вычислений, возвращение к доброй старой каузальности;
- поворот от логики к математике;
- признание существенной роли биологической определённости человека в формировании его интеллектуальных качеств.

### 2.2.5 Предсказывающий разум

Согласно некоторым научным концепциям, образы, возникающие в чувственном восприятии, ни в коей мере не являются слепками действительности. На самом деле они являются постоянно уточняемыми и моделями, создаваемыми нашим мозгом. Эти модели снова и снова уточняются в соответствии со вновь поступающими чувственными данными. То есть, чувственный образ, который мы имеем и в котором мы отдаем себе отчёт, является результатом многократных итераций по перестраиванию мозгом имеющихся у него априорных моделей.

«Предиктивный процессинг» (ПП)<sup>19</sup>, по мнению многих, является одним из наиболее влиятельных и обладающих наибольшей объяснительной силой когнитивных подходов в настоящее время [61; 191; 192].

<sup>19</sup>Словосочетание ‘Predictive Processing’ стало наиболее употребимым в актуальной литературе обозначением этого направления.

С когнитивным классицизмом ПП объединяет то, то она также существенно опирается на понятия вычислений и репрезентаций. Однако, подобно коннекционизму и вообще нейросетевому видению, репрезентации рассматриваются не как символные, а как субсимвольные, выраженные некоторыми распределениями вероятностей, и, соответственно, вычисления понимаются как вероятностный (байесовский) вывод. Но то, чем ПП отличается от обеих конкурирующих компьютерно-рационалистских парадигм, состоит в понимании репрезентации как предсказания: когнитивная система, управляемая многоуровневыми аттракторами, создаёт гипотезы относительно каузальной структуры окружающей среды, позволяющие предсказывать входящие перцептивные данные. Количественное расхождение предсказанных и реально получаемых внешних данных есть та «свободная энергия», к минимизации которой стремится любая неравновесная система. Такая минимизация идёт двумя альтернативными путями: обновление генеративных моделей (когнитивные акты) или активный вывод с целью изменения внешних данных (поведение).

Новая парадигма, сформировавшаяся в середине 2010-х годов, претендует на преодоление ограничений предыдущих подходов. Один из его основателей, британский нейробиолог Карл Фристон, предлагает биологически правдоподобное объяснение, основанное на концепции обновления внутренних репрезентаций под воздействием сенсорных образцов. Теория постулирует существование генеративных моделей, которые порождают нисходящие предсказания. Последние встречаются и сравниваются с восходящими репрезентациями на более низком уровне, с тем чтобы вычислить ошибку предсказания [193, р. 392]. Естественное стремление свести к минимуму разницу между прогнозируемой репрезентацией и поступающими данными составляет суть так называемого «принципа свободной энергии» (ПСЭ), подробнее о котором ниже. Сторонники ПП выводят основные принципы своего подхода из известных в прошлом философских и психологических доктрин, таких как доктрины Альхазена, Канта и Гельмгольца [194, р. 210]. Что касается последнего, то ПП восходит к его идее «бессознательных выводов». Они формируются в раннем возрасте и составляют, по мнению естествоиспытателя, основу многих феноменов восприятия. Согласно Гельмгольцу, мы настраиваем наши чувства, чтобы с максимальной точностью различать вещи, которые на них воздействуют. Восприятие, таким образом, является результатом встречи внешнего воздействия с тем, что человеку уже известно [195]. Физик объясняет также само понятие «свободная

энергия» [196]. В конце XX века его именем была названа машина Гельмгольца — иерархический алгоритм неконтролируемого обучения, способный определять структуры, лежащие в основе различных шаблонов данных [197]. ПП представляет собой сложную систему взаимосвязанных понятий и концепций, освоить которую непросто. Насколько я могу судить, он основан на очень простой идее «предиктивного кодирования», к которой он добавляет концепции, заимствованные из термодинамики, статистики и теории информации. «Предиктивное кодирование» сводится к постулированию сгенерированных априорных моделей, которые последовательно обновляются путем взвешивания ошибок предсказания. Ошибка предсказания представляет собой разницу между предсказанием и вновь полученными данными. Обычно, когда указывают различные предшествующие теории в качестве предшественников ПП, например, «функциональные системы» П. К. Анохина [198], они имеют в виду эту общую идею, а не математические подробности, содержащиеся в полноценном ПП, которые, собственно, и определяют объяснительную и предсказательную силу теории. Так, например, Е. Е. Витяев, предлагая собственную формализацию концепции когнитива К. В. Анохина (см, параграф 2.2.3), основанную, в частности, на теории функциональных систем, указывает, что его «формализация принципиально отличается от [формального аппарата ПП], поскольку байесовские сети не поддерживают циклов» [199, с. 27]. ПП дополняет идею предиктивного кодирования, во-первых, концепцией оптимизации точности, которая модулирует вычисление ошибок предсказания на разных уровнях системы. Точность выборок оптимизируется за счет ранее полученного опыта, поэтому для расчетов требуется статистическая основа — «эмпирический байесовский подход». Это означает, что обработка является многоуровневой и зависит от контекста, и, таким образом, процессинговая<sup>20</sup> система, будь то клетка, человек или робот, способна аппроксимировать иерархический эмпирический байесовский вывод, благодаря чему она лучше адаптируется к неопределённой и постоянно меняющейся среде, чем система, выполняющая только точный байесовский вывод [194, р. 213]. С другой стороны, ПП во многом основан на концепции активного вывода. Чтобы свести к минимуму ошибки прогнозирования, система может, например, обновлять прогнозные модели, чтобы они соответствовали сенсорным входным данным. Но, в качестве альтернативы, она может активно менять пробы среды в поисках данных, ко-

<sup>20</sup>См. определение на с. 306.

торые лучше соответствуют прогнозу. Это означает, что она просто активно действует, откуда, собственно, сам термин. Согласно Фристон, живой организм может быть проще объяснен как действующая система, если мы предположим, что триггеры для активного вывода являются проприоцептивными данными, потому что они могут быть напрямую функционально связаны с рефлекторными дугами [200—202]. Таким образом, ПП оказывается единой теоретической основой, способной объяснить как восприятие, так и действие. Вышеупомянутые концепции являются частными теориями, апеллирующими к реальным или смоделированным механизмам, и поэтому могут быть напрямую фальсифицированы. Но в основе ПП лежит общий принцип, который действует так же, как известные принципы естествознания: он придаёт форму объяснениям частных теорий, но сам не поддается прямому опровержению. Это так называемый принцип свободной энергии (ПСЭ). Он постулирует «минимизацию свободной энергии» как главное стремление любой самоподдерживающейся или аутопоэтической системы, например, живого организма. Концепция свободной энергии, заимствованная из термодинамики, принимается и объясняется Карлом Фристоном как «удивление», которое должно быть минимизировано [203]. Большое количество удивления, т. е. взаимное несоответствие предиктивной модели и сенсорных данных, слишком дорого обходится когнитивной или, в широком смысле, живой — системе и должно быть сведено к минимуму. Эта потребность запускает как перцептивный, так и активный вывод. Оснащенный полным набором инструментов, ПП приводит не только к интересным эмпирическим объяснениям, таким как изменение настроения или шизофрени, но и к некоторым философским выводам. Таким образом, идея восприятия как постоянного вывода из сенсорных данных их вероятных причин представляет собой интересную переформулировку таких излюбленных философами предметов, как образ тела, чувство собственного тела и телесное самосознание. Указанные явления естественным образом объясняются предполагаемыми причинами interoцептивных и проприоцептивных ощущений. Согласно Якобу Хови, само «Я» в этом контексте может быть представлено «как подмножество предполагаемых причин сенсорного ввода, относящегося к собственным действиям, и, следовательно, возможность обсуждения того, заслуживает ли такой набор причин обозначения “я”» [194, р. 217]. Столь расширенная теория ПП, очевидно, включает знакомые сюжеты феноменологии в стиле Мерло-Понти, которая, как известно, была методологической основой для энактивизма и

«воплощённого познания» [60]. Интересно, что такой видный сторонник внедрения феноменологии в когнитивную науку, как Томас Метцингер, оказался в числе энтузиастов ПП и теперь является одним из редакторов всеобъемлющего тематического веб-ресурса [204], посвященного ПП и содержащего своего рода онлайн-энциклопедию по этой теме. Другой концептуальный столп ПП — это порождение мозгом набора генеративных моделей (ГМ), которые отвечают за создание возможных репрезентаций окружающей среды. Понятие таких моделей предполагает многоуровневую организацию, такую что иерархическая структура позволяет моделям на одном уровне подпитываться данными на более высоком уровне. Предполагается, что сенсорные данные находятся только на самом нижнем уровне иерархии, а самый высокий уровень генерирует только спонтанные случайные колебания [196, р. 75]. Те же авторы, в значительной степени опираясь на сложные математические вычисления, дают неформальное определение ГМ как «описание причинных зависимостей в окружающей среде и их отношения к сенсорным сигналам» [196, р. 61]. Более подробно ГМ объясняются в терминах плотности распознавания (R-плотность) и генеративной плотности (G-плотность), под которыми в обоих случаях понимаются плотности вероятности (распределения). Предполагается, что организм моделирует вероятность переменных окружающей среды, которая выражается как R-плотность. Но для этого он должен быть способен оценивать общие зависимости входящих сенсорных данных от состояний окружающей среды. Статистическая модель этих зависимостей выражается G-плотностью. Тогда взаимозависимость обоих видов плотностей создает генеративную модель. Что касается механистической реализации этих вычислительных моделей, они «создаются и параметризуются физическими переменными в мозге организма, такими как нейронная активность и синаптическая сила соответственно» [196, р. 57]. Таким образом, можно сделать вывод, что ПП фактически развивает, исправляет и обогащает первоначальную доктрину коннекционизма, используя, в частности, другие формальные инструменты, опираясь на ту же онтологию. Фристон демонстрирует интересное применение этой теоретической основы для моделирования языковой коммуникации, которая ранее была сферой классической символической когнитивной науки. По его словам, критерии оценки и тонкой настройки интерпретации чужого поведения те же, что лежат в основе действий и восприятий в целом, а именно минимизация ошибок прогнозирования. Концепция коммуникации в ПП основана на генеративной модели или

нарративе, которую разделяют агенты, обменивающиеся сенсорными сигналами. По словам Фристонa, модели, основанные на иерархических аттракторах, которые генерируют различные категории последовательностей, позволяют замкнуть герменевтический круг, просто обновляя генеративные модели и их предвидения, чтобы минимизировать предиктивные ошибки. Важно отметить, что эти ошибки можно вычислить, даже не зная истинного состояния другого собеседника, что, тем самым, решает проблему герменевтики [205, p. 129–130]. Фристон и его коллеги создали компьютерную эмуляцию двух певчих птиц с помощью программных агентов, чьи трели генерировались некоторыми моделями на основе аттракторов и рекурсивно уточнялись в процессе взаимного прослушивания. Модель показала, что птицы следуют нарративу, создаваемому динамическими аттракторами в их генеративных моделях, которые синхронизировались посредством сенсорного обмена. Это означает, что обе птицы могут петь «с одного нотного листа», сохраняя при этом последовательную и иерархическую структуру своего повествования. Именно это явление Фристон связывает с общением [193, p. 400]. Генеративные модели, используемые для определения собственного поведения, могут использоваться для определения убеждений и намерений другого при условии, что обе стороны имеют довольно похожие генеративные модели. Эта перспектива создает репрезентации набора преднамеренных действий и нарративов, предлагая коллективное повествование, разделяемое коммуницирующими агентами [193, p. 401]. Можно предположить, что объяснительные возможности ПП охватывают не только вопросы психологии и традиционной философии сознания, но и более новые темы «социального интеллекта» и социального познания. Безусловно, такая универсальность может вызвать опасения относительно фальсифицируемости теории. Краткая дискуссия по этому поводу содержится в [194, p. 221]. Но, как правило, такое беспокойство может быть связано с вопросом, может ли каждое поведение быть представлено как направляемое статистическими прогнозами субъекта. На самом деле кажется, что в этом случае трудно представить себе прямое опровержение, но всегда есть место для лучшей теории, которая продемонстрирует больший объяснительный потенциал. Более того, теория ПП ещё слишком молода: хотя в реальных экспериментах она демонстрирует больший или меньший успех, она неизбежно столкнется с необходимостью дать подробные механистические объяснения всех реализованных статистических моделей. Эти объяснения, безусловно, будут фальсифицируемы.

Как пишет Крис Фрит,

Скрывая от нас все бессознательные заключения, к которым он приходит, наш мозг создает у нас иллюзию непосредственного контакта с материальным миром. В то же самое время он создает у нас иллюзию, что наш внутренний мир обособлен и принадлежит только нам. Эти две иллюзии дают нам ощущение, что в мире, в котором мы живем, мы действуем как независимые деятели. Вместе с тем мы можем делиться опытом восприятия окружающего мира с другими людьми. За многие тысячелетия эта способность делиться опытом создала человеческую культуру, которая, в свою очередь, может влиять на работу нашего мозга» [206, с. 37-38].

### **2.2.6 Теория функциональных систем как предтеча «предсказывающего разума»**

От отечественных нейрофизиологов не может скрыться онтологическая и функциональная схожесть ПП и теории функциональных систем (далее — ФС) П. К. Анохина. Функциональные системы — это динамически складывающиеся «архитектуры», управляющие определёнными аспектами поведения живых организмов. В результате афферентного синтеза формируется программа действия и акцептор результата, работающий через некую обратную связь, называемую обратной афференцией. Поведение системы направляется стремлением к сокращению степеней свободы. Казалось бы, обе теории описывают одно и то же разными словами: системы с обратной связью, обменивающиеся информацией с внешней средой и стремящиеся к минимизации некоторых переменных. Однако далее я попытаюсь показать, что эти теории нельзя рассматривать как эквивалентные, и в пользу этого вывода свидетельствуют два принципиальные соображения.

Первое состоит в том, что онтология теории и собственно теория — не одно и то же. Предметная (или доменная) онтология, по сути, представляет собой таксономию типичных объектов, принимаемых теорией как существующие, их свойств и отношений. Её высказывания не истинны и не ложны, но конвенциональны, принимаемы по соглашению. Напротив, собственно теория

состоит в интерпретации неких формальных (например, математических) или содержательных, но проверяемых высказываний на объектах онтологии. Так, например, предложение «физические тела обладают массой» относится к онтологии и характеризует наш способ мироописания, а не мир как таковой, тогда как высказывание «действие равно противодействию» уже часть теории и может оцениваться на истинность/ложность<sup>21</sup>.

С этой точки зрения теории ПП и ФС обнаруживают некоторые очевидные схожести на уровне онтологий и некоторых содержательных теоретических принципов — например, тенденции к минимизации определённых показателей. Однако они используют разные математические аппараты: ПП тяготеет к байесовской статистике, тогда как классическая ФС была сформулирована на качественном уровне, а попытки её математизации начали предприниматься только в наше время. Это означает, что их собственно теоретические высказывания могут оказаться несоизмеримы.

Второе соображение связано с принципиальной многоуровневостью описания вычислительных процессов. В 1976 г. Дэвид Марр, классик в области вычислительной нейрофизиологии и теории зрительного восприятия, в своей впоследствии переиздававшейся статье, написанной в соавторстве с Томазо Поджио предложил понятие уровней вычисления<sup>22</sup>. В первом варианте концепции было названо четыре уровня, впоследствии их число сократилось до трёх, и именно в этом виде данная концепция стала одной из самых цитируемых (и критикуемых) в публикациях, посвящённых когнитивным вычислениям. Согласно Марру, любая вычислительная система может быть описана на трёх уровнях: (1) теория вычисления, описывающая его цель, приемлемость и стратегию, (2) репрезентация и алгоритм, где определяются формы представления входных и выходных данных, а также алгоритм преобразования одного в другое, и (3) физическая реализация алгоритма и репрезентаций.

Если располагать сравниваемые теории на марровских уровнях описания вычислений, то ПП будет располагаться на первом, телеологическом, и частично на втором, алгоритмическом, уровне. Теория ФС, напротив, очевидно алгоритмична, но некоторые части её онтологии тяготеют к уровню физических реализаций — третьему в системе Марра. Отсюда следует, что и кажущаяся онтологическая схожесть двух теорий имеет место только при условии некоторых

<sup>21</sup>Подробнее об этом см. 1.1.3

<sup>22</sup>Данная концепция рассматривается в деталях в разделе 2.3.2



дополнительных онтологических допущений, устанавливающих соответствия между двумя онтологиями. Если это так, то подозрения в несоизмеримости обсуждаемых теорий ещё более укрепляются.

Чтобы подтвердить их или опровергнуть, обратимся к некоторым деталям.

По мысли П. К. Анохина, функциональные системы (ФС) — это то, что динамически складывается в организме в ответ на некоторую задачу. Главный критерий и фактор её реализации — это ожидаемый результат, определённое представление которого уже содержится в ФС. На вход системы подаётся главный сигнал — поисковый стимул — и «обстановочная афференция» или восприятие поискового контекста. Эти данные передаются в блок афферентного синтеза, где важную роль играет содержание памяти относительно подобных ситуаций в прошлом. Далее следует принятие решения, и затем сигнал поступает в два параллельных блока: программа действия и акцептор результата действия. Главный критерий успеха действия для организма, согласно Анохину, это поддержание определённого уровня метаболизма — подобно тому, как, в соответствии с ПП, организм стремится к минимизации свободной энергии. Акцептор действия — соответственно, заранее заданный критерий оценки успешности действия. Программа действия даёт непосредственную обратную связь с акцептором действия. Далее реализуется само действие, получается результат, и его параметры также поступают в акцептор. Если результат соответствует ожиданиям организма, определяемым потребными характеристиками метаболизма, то программа выполнена и организм переходит к чему-то иному. В противном случае происходят новые итерации с неизбежной подстройкой исходной модели.

Однако метаболизм — это физиологическая детерминанта. Но в теории ФС есть и математический предел, минимизация которого рассматривается как принципиальная телеология изучаемых процессов. Это степени свободы системы в целом, её подсистем, их элементарных единиц. В этом пункте аналогия со «свободной энергией» ещё более очевидна. Анохин пишет: «мы должны вскрыть те детерминирующие факторы, которые освобождают компоненты системы от избыточных степеней свободы» [207, с. 68]. И далее: «важным последствием включения результата в систему как решающего операционального фактора системы является то, что сразу же делаются понятными механизмы освобождения компонентов системы от избыточных степеней свободы» [207, с. 71]. Заранее данная репрезентация результата есть то, что способствует этой минимизации

на основе обратной связи: «Главное качество биологической самоорганизующей системы и состоит в том, что она непрерывно и активно производит перебор степеней свободы множества компонентов, часто даже в микроинтервалах времени, чтобы включить те из них, которые приближают организм к получению полезного результата» [207, с. 72].

Аналогично обратной связи в механических системах, в органических Анохин выделяет обратную афферентацию как способ получения организмом информации о достигнутом результате действия. Этот контур может в том числе «санкционировать последнее распределение в системе эфферентных возбуждений, обеспечивших достижение полезного результата» [207, с. 70]. Последовательное применение этой схемы ведёт к сокращению числа степеней свободы системы в пользу тех, которые способствуют достижению заранее представленного результата. Таким образом создаётся определённая структура, приемлемая для целей организма в данной конкретной нише с её вероятностными инвариантами. Такой подход, по мнению создателя теории ФС, выгодно отличается от общей теории систем Людвиг фон Берталанфи, поскольку последняя не объясняет, каким образом внутри системы создаётся структурная определённость, равнозначная отказу от многих и многих степеней свободы [207, с. 72].

Афферентный синтез, согласно Анохину, осуществляется с участием четырёх компонентов: мотивация, обстановочная афферентация, пусковая афферентация и память. Результатом этого синтеза является постановка цели, которой в дальнейшем будет следовать живая система. В результате отбора приемлемых степеней свободы эфферентные возбуждения передаются на определённые группы мышц, и, таким образом, совершается нужное действие [207, с. 88–92].

Если ПП опирается на концептуальный каркас физики, то теория ФС в данном случае предлагает скорее логическое видение, аналогичное тому, как в некоторых эпистемических логиках приобретение знания рассматривается как эквивалент исключения некоторых возможных миров. Правильным путём, как представляется, был бы синтез обоих подходов, возможно, через установление концептуальной связи между больцмановским и шенноновским пониманием энтропии («удивления», в терминологии Фристана).

Возможно, физическая интерпретация через минимизацию свободной энергии может помочь в поисках такого механизма. Минимизация степеней

свободы может работать скорее как абстрактная концептуальная модель. Её необходимо дополнить описанием каузальных связей внутри схемы. Согласно одной из интерпретаций, шенноновская энтропия есть мера количества «да/нет»-вопросов, необходимых для определения содержания сообщения [208, р. 3398]. Это примерно соответствует характеристике системы в терминах степеней свободы в понимании Анохина. Энтропия в понимании Больцмана — это число микросостояний, соответствующих некоторому макросостоянию. В некотором концептуальном приближении можно сказать, что первое понимание этого термина имеет отношение к формально-алгоритмическому описанию того, физическое описание чего предполагает вторая интерпретация. Однако полное описание, например, живых систем на материальном уровне потребовало бы использование средств органической химии, клеточной биологии, теории эволюции и т. д., при том что эволюция сама по себе является механизмом сокращения числа возможных состояний. Применение вычислительного подхода — т. е. описание всей системы в её динамике как системы алгоритмов (эволюционных, генетических и т. п.) — сделало бы теорию значительно более компактной и функциональной.

В то же время теория ФС пока (на момент написания рассматриваемого текста) не в состоянии ответить на вопрос о механизме минимизации степеней свободы. «Как происходит это освобождение от избыточных степеней свободы? Почему момент принятия решения часто имеет характер внезапной интуиции? В настоящее время мы еще не можем ответить на эти вопросы, но выработанный подход к ним дает полную гарантию того, что принятие решения в биологических системах с большой и малой иерархией является вполне анализируемым и доступным для объективной науки феноменом» [207, с. 93].

Общий алгоритм работы ФС демонстрирует очень близкую аналогию с таковым ПП. Согласно П. К. Анохину, процесс оценки полученного реального результата осуществляется путём сличения прогнозированных параметров и параметров реально полученного результата. Специальный динамически формирующийся аппарат производит оценку полученного результата. Эта оценка определяет дальнейшее поведение организма. Если результат соответствует прогнозу, то организм переходит к следующему этапу «поведенческого континуума». «Если же результат не соответствует прогнозу, то в аппарате сличения возникает рассогласование, активирующее ориентировочно-исследовательскую реакцию, которая, поднимая ассоциативные возможности мозга на высокий уро-

вень, тем самым помогает активному подбору дополнительной информации» [207, с. 97].

Е. Е. Витяев относительно недавно предпринял попытку математизировать теорию ФС [209]. Опуская символические детали, я попробую описать общий подход. Прежде всего, в данной интерпретации вводится понятие ФС одного действия, которая складывается в организме под выполнение одного элементарного действия. Она определяется целевым состоянием системы (ЦС), набором правил (П) и оценкой вероятности достижения этой цели при применении этих правил (В). ЦС определяется набором предикатов, каждый из которых принимает одно из двух возможных значений: истина или ложь. В может быть вычислена на основе статистики достижения подобных целей в прошлом при применении П, хранимой в памяти. На этапе афферентного синтеза происходит идентификация наличного состояния системы с хранимыми в памяти аналогами. На этапе принятия решения из той же памяти извлекаются правила, применение которых ранее чаще приводило к успеху. Акцептор действия, в свою очередь, содержит некоторую антиципацию желаемого результата. Он принимает данные о реальном результате совершённого действия и сравнивает их с данными антиципации. При их несовпадении выбранные правила понижаются в весе, т. е. в вероятности выбора их в дальнейшем. При совпадении ожидания и реальности правила, напротив, закрепляются.

Далее Витяев вводит понятие комплексной функциональной системы. Она определяется представлением об общей цели конкретной деятельности, комплексом функциональных систем одного действия и значением вероятности достижения цели этими средствами. Таким образом комплексная функциональная система являет собой иерархию вложенных ФС одного действия, подобно тому, как у Фристана организм представляет собой ограждение Маркова, состоящее из вложенных ограждений Маркова. Например, согласно некоторым представлениям, органеллы эукариотической клетки эволюционно возникли из прокариотических клеток, объединившихся в единый организм. Витяев специально указывает, что, согласно концепции Н. А. Бернштейна, ведущим уровнем организации движения является верхний уровень, который соответствует доминирующей мотивации. ФС нижних уровней активизируются ФС верхних уровней [209, р. 627]. Таким образом, оба подхода — ПП и ФС — в существенной степени основаны на концепции нисходящей причинности.

Сравнение двух концепций предсказывающего познания позволяет сделать некоторые выводы. Согласно некоторым подходам, чья популярность возрастает в последнее время, научные теории можно подразделить на номологические и алгоритмические (вычислительные) [159; 210]. С этой точки зрения, теория ПП является скорее вычислительной и, по этому признаку, далее отстоит от традиционного, основанного на законах, естествознания, чем теория ФС. Несмотря на встречающиеся в текстах Фристана и последователей отсылки к конкретным нейробиологическим воплощениям, теория в целом формулируется и применяется как описание некоторого типичного вычисления (в данном случае, свободной энергии), которое на верхнем марровском уровне достаточно индифферентно по отношению к своим алгоритмическим и физическим реализациям (хотя некоторые версии ПП достигают второго марровского уровня, описывая статистические алгоритмы). Напротив, теория ФС в значительной степени привязана к онтологии живого организма, наделённого нервной системой и памятью и способного к целенаправленным действиям. Можно сказать, что вычислительный подход теории ПП избавляет её от необходимости объяснять нейробиологические детали применять её общие формулы далеко за пределами собственно когнитивной сферы, для объяснения которой она изначально строилась. Теория ФС, в свою очередь, существенно опирается на представления об устройстве живого действующего организма, но при этом признаёт собственную неспособность объяснить некоторые из нейробиологических механизмов, обеспечивающих достижение искомых параметров. Таким образом, её можно в основном локализовать на втором, алгоритмическом уровне трёхуровневой схемы Марра. Вместе с тем, тот факт, что теория ФС в большей степени ссылается на устройство реальных живых организмов, демонстрирующих когнитивные способности, позволяет использовать математизированную версию этой теории в проектировании мультиагентных систем. Но этот же факт обуславливает некоторые её части фактами материальной реализации, относящимися к третьему уровню схемы Марра.

Теория функциональных систем в большей степени ссылается на устройство реальных, в основном естественных, когнитивных архитектур. Витяев утверждает, что использовал её в проектировании реальных мультиагентных систем. Таким образом, она ближе к традиционному естествознанию, и, на первый взгляд, у неё лучше обстоят дела с фальсифицируемостью.

В ней присутствуют *ad hoc* решения, как, например, «предшествующий опыт», хранимый памятью. Т. е., мы должны предполагать наличие предшествующего опыта, наличие памяти как его хранилища.

Теория ПП более проста, более обща, потенциально обладает большей объяснительной силой, хотя с уверенностью утверждать это пока ещё рано. Математические средства обеих теорий формализованы, но недостаточно формальны. Под формализованной теорией здесь понимается теория, в языке которой некоторые термины естественного языка заменяются формальными символами, но логическая структура при этом принципиально не меняется. Формальная же теория не только обильно использует символизмы, но и правила перехода от одних высказываний к другим в такой теории являются чисто синтаксическими и не зависят от свойств моделей, на которых интерпретируется данная формальная теория. Поскольку ни теория предиктивного процессинга, ни теория функциональных систем не являются формальными в строгом смысле слова, и различаются по степени формализованности, их, на самом деле, пока трудно сравнивать на предмет фальсифицируемости.

Таким образом, у нас нет концептуальных и методологических оснований утверждать, что теории функциональных систем и предиктивного процессинга эквивалентны. Они располагаются на разных уровнях схемы Марра, используют разную математику и, возможно, в разной степени являются фальсифицируемыми. Исходя из этих соображений, можно предположить, что объяснения и предсказания этих двух теорий будут различаться.

\* \* \*

Как показано выше, смена подходов (парадигм) в истории когнитивной науки была вызвана не в последнюю очередь осознанием недостаточности классической — символично-серийной — концепции вычислений для порождения удовлетворительных объяснений её фактов. Это осознание и разочарование, с одной стороны, породило многочисленные антикомпьютералистские концепции, вроде популярного ныне «радикального энактивизма». С другой стороны, пока единственной альтернативой компьютерализму остаются традиционные естественные науки — физика, химия, биология, нейрофизиология — каждая из которых достаточно сложна, и отдельную проблему составляет их междисциплинарная интеграция. Напротив, в случае появления приемлемой вычислительной теории сознания мы сможем ограничиться достаточно простой

онтологией, описывающей только элементы материальной реализации вычислительных алгоритмов, а в качестве формального языка может быть принят тот или иной вариант вычислительной математики, которая в последнее время претерпевала интересные развития.

### 2.2.7 Вывод из раздела 2.2

Разработка **концепции вычислений, адекватной предмету когнитивных наук, и перевод этих наук на компьютерно-информационные основы** поможет преодолеть ограничения классических естественно-научных методологий, не справляющихся со сложностью предмета.

## 2.3 Проблема вычислений в когнитивной науке

Как было показано выше, «когнитивная революция» в науках о познании состоялась как реакция на необихевиоризм в тот момент, когда реальная компьютерная революция дала психологам и лингвистам концептуальные средства для научного исследования сознания. До тех пор признаком научности считалось стремление избегать психологической терминологии. Первоначальный вариант этой неоменталистской картины был, наверное, слишком прямолинейно списан с компьютерной науки и включал два основных элемента: *вычисления*, которые производятся над символическими *репрезентациями*. То есть, вычисления и репрезентации стали двумя концептуальными столпами, на которых зиждется когнитивистская парадигма. Однако, чтобы правильно определить место и роль каждого из этих понятий, нужно разобраться в концептуальных тонкостях, которые связаны не в последнюю очередь с более широкими сдвигами в современных естественных науках. В большей степени это касается понятия вычислений, которое играет всё большую роль в физике, астрофизике, нейрофизиологии, биологии и даже социальных науках. При этом имеется в виду не инструментальная или методологическая их роль, а самая что ни

на есть онтологическая — когда вычислениями занят не исследователь, а сам предмет исследования.

### 2.3.1 Вычисления и проблема их определения

Первая абстрактная модель того, что представляет собой вычисление, появилась как попытка решить проблему вычислимости функций, поставленную Дэвидом Гильбертом. Алан Тьюринг [183] попытался представить себе деятельность человека-«вычислителя» ('computer', в его терминологии) в самом абстрактном виде: как последовательную запись и преобразование символов некоторого конечного алфавита (например, цифр) в клетках, нанесённых на бумажный лист. Поскольку, согласно нашим интуитивным представлениям, чтобы быть вычислением, этот процесс должен подчиняться определённым правилам, — т. е., быть жёстко детерминированным, — то, по мысли Тьюринга, человека можно заменить некоей весьма абстрактно описанной печатающей головкой, а бумагу — бесконечной лентой, разделённой на клетки, в каждой из которых может быть записан только один символ. И вот перед нами машина, чьё устройство определяется несколькими конечными множествами: возможные действия (напечатать, стереть, сдвинуться влево или вправо), возможные состояния, алфавит и правила. Если каждый шаг вычисления однозначно определяется уже имеющимся символом в считываемой клетке ленты и текущим состоянием машины, то такая машина считается детерминированной. Если машина способна считать с ленты не только символы (данные), но и правила их обработки (программу), то такая машина считается универсальной. Тьюринг показал, что его машина способна вычислить любую вычислимую функцию. На основе концепции универсальной машины Тьюринга Джон фон Нойманн описал принципиальную архитектуру современного компьютера. Концепция Тьюринга, касающаяся вычислений и вычислимости, появилась в контексте математической дискуссии. Отсюда происходит уверенность (с моей точки зрения, чисто психологическая) в том, что вычисления выполняются непременно сознательными субъектами (скажем, математиками) с определённой целью. Хотя на самом деле, как мне представляется, общая идея Тьюринга (возможно, даже не полностью осознаваемая им самим) заключалась в следующем: если человек



(«вычислитель», в его терминах) выполняет некую последовательность символических операций, которая также может быть выполнена машиной, то это и есть вычисление. По умолчанию, вычисление предполагает алгоритм. Но если у вас есть алгоритм, вы в любом случае можете построить машину для его выполнения.

Характерными особенностями тьюринговой модели вычислений являются её символизм, серийность и алгоритмичность: вычислитель, кем или чем бы он ни был, совершает операции над символами, обрабатывая один из них на каждом шаге вычисления в соответствии с некоторым алгоритмом. Однако возможно пойти дальше по пути абстрагирования: вместо «Некое устройство или человек обрабатывает серии символов в соответствии с алгоритмом», этот процесс можно описать как « $X$  что-то делает с  $Y$  в соответствии с  $Z$ ». На 'generic'-уровне вычисление предстаёт как некое динамическое процессуальное трёхместное отношение. Это отношение можно описать двумя парами соотносимых категорий: серийное (последовательное) — параллельное (распределённое) и дискретное — континуальное (последнюю пару иногда не совсем корректно отождествляют с парой цифровое — аналоговое).

Тогда общие виды вычислений можно представить в виде таблицы:

Таблица 1 — Виды вычислений

<b>Вычисления</b>	<b>серийные</b>	<b>параллельные</b>
<b>дискретные</b>	тьюринговы	нейросетевые
<b>континуальные</b>	технологические	нейробиологические

В таблице представлены типы, которые можно было бы назвать идеальными. В реальности сложные вычислительные устройства часто имеют смешанный характер. Так, некоторые искусственные нейросети допускают действительные числа в качестве значений своих переменных, т. е., по сути являются (отчасти) континуальными по своей вычислительной архитектуре. Нейроны в биологических системах — например, в мозге — обмениваются сериями спайков, каждый из которых описывается континуальной амплитудой, но сами их последовательности дискретны [211, р. 467]. А участие химических нейромедиаторов в усилении или подавлении электрической активности нейронов существенно добавляет «аналоговости» всему процессу. Благодаря пониманию ограниченной применимости тьюринговой концепции для описания природных

вычислительных процессов, в последнее время стали появляться теоретические альтернативы. Так, в ряде публикаций формулируется концепция вычислений на основе абстракции «механизма» [174; 212]. Механизм как вычислительная система понимается как пространственно-временное единство функционально определённых составных частей или элементов с достаточно большим числом возможных состояний. Начальное состояние механизма рассматривается как «вход» вычислительного процесса, конечное состояние — как его «выход».

Благодаря своей абстрактности, эта модель, действительно, на первый взгляд выглядит как концепция вычислений, способная продемонстрировать компьютерный характер достаточно широкого спектра природных процессов. Однако она упускает из виду особенность вычислительных процессов, которую можно назвать многоуровневостью. Это отличительное свойство вычислительных — в отличие от других систематических — процессов впервые было замечено Дэвидом Марром, который писал о несводимых друг к другу уровнях всякого вычислительного процесса [172, р. 24–25]. Примерно то же имел в виду Б. Дж. Коулэнд, говоря об аналоговой системе физического устройства как «честной» (не-нестандартной) модели архитектуры и алгоритма некоторого вычисления [169]. Обе концепции предполагают, что собственно вычислительный уровень процесса, предстающий перед разумным наблюдателем как символический и целесообразный, надстраивается над алгоритмическим уровнем, состоящим из сколь угодно сложной комбинации примитивных шагов, и уровнем физической реализации этого вычисления, где и цель, и алгоритм реализуются в виде электрических, химических, квантовых и других естественных взаимодействий. При этом определённое вычисление с его символикой и целью может быть реализовано более чем одним алгоритмом, а определённый алгоритм, в свою очередь, может быть реализован более чем одной физической системой. Концепция механизма, на мой взгляд, не обладает концептуальными средствами, достаточными для различения этих уровней.

Попутно замечу, что вопрос, является ли «вычисление» некоей метафорой для описания многоуровневых процессов в самоорганизующихся системах, или такие системы суть в буквальном смысле вычислительные устройства, имеет отношение к дискуссии между научным реализмом и антиреализмом, которая не специфична для вычислительного подхода, но касается также, а может быть, прежде всего, номотетических теорий.

Для природных вычислений важен не столько их дискретный/континуальный характер, сколько их параллельная архитектура. Последняя является наглядным, хотя и косвенным, свидетельством эволюционного происхождения естественных вычислительных систем. В отсутствие инженера-конструктора, действующего сознательно и целенаправленно, вычислительная мощность может наращиваться только путём сложения мощностей уже существующих процессоров. Случайные объединения клеток или особей, если они приводят к наращиванию вычислительной мощности и/или экономии энергетических затрат на вычисления, отбираются и закрепляются. Поэтому виды с более развитым мозгом вытесняют тех, кто отстаёт по этому показателю, из определённых экологических ниш. Таким же образом получают эволюционные преимущества виды с более эффективной социальной организацией. В последнем случае объём мозга уже может не иметь решающего значения, поскольку вычислительные задачи распределяются по хорошо организованной сети, состоящей из когнитивно нагруженных особей (ср. историю неандертальцев и кроманьонцев).

Моделирование вычислительных систем человеком изначально пошло в направлении, обратном естественной эволюции. Нам оказалось легче научить компьютер символическим, логическим и математическим операциям, чем распознаванию образов, звукам, хождению и управлению мелкой моторикой. Аналогично, если искусственные компьютеры нужно специально готовить к работе в сети, естественные вычислительные устройства (клетки или организмы) на определённом этапе своей эволюции с готовностью объединяются в «строчки», комплексы или сети. Естественная причинность, как она понимается в традиционном естествознании, не способна объяснить это развитие. Но, если вычислительные системы онтологически реализуются как надстраивание уровней управления и цепочки нисходящей причинности, то поиск большей эффективности вычислений ведёт к их параллелизации. Если это так, то нужна теория, объединяющая физику и теорию вычислений (возможно, связывающая нарастание «вычислительности» природных процессов с неравновесной термодинамикой).

Учитывая соображения, изложенные выше, можно было бы остановиться на следующем определении вычисления, которое учитывало бы вызовы, стоящие перед науками о жизни, обществе и сознании. Оно исходит из предположения, что любая достаточно сложная структура обладает некоторы-

ми формальными свойствами, которые имеют значение при взаимодействии структуры с её окружением. Если при таком взаимодействии зависимость формальных свойств одной структуры от формальных свойств другой описывается некоей вычислимой функцией, то имеет место передача (и преобразование) информации, сопровождающаяся вычислительными действиями. Тогда, введя с помощью подчинённых дефиниций понятия механизма и репрезентации, можно определить, что:

*(Df3)* Вычисление — это процесс, осуществляемый вычислительной системой.

*(Df3.1)* Вычислительная система есть один из множества возможных механизмов некоторой репрезентации.

*(Df3.1.1)* Механизм есть взаимно однозначное соответствие между множеством возможных действий и множеством возможных состояний.

*(Df3.1.2)* Репрезентация есть отображение (mapping) формальных свойств одного процесса в формальные свойства другого.

*(Df3.1.1.1)* Процесс есть любая последовательность действий или состояний чего бы то ни было.

*(Df3.1.1.2)* Формальные свойства суть свойства, которые могут быть описаны на языке логики или математики.

Это определение сохраняет классицистскую когнитивную позицию в том отношении, что вычисление продолжает пониматься как правилосообразные операции над репрезентациями. Но сами репрезентации «отвязываются» от какого бы то ни было психологизма, и получают чисто формальное определение. Такой подход соответствует нашим интуициям, в соответствии с которыми всякое вычисление всегда есть вычисление чего-то, причём это что-то понимается скорее как цель, а не как референт. И, наконец, представленное определение ясно показывает отношение включения между множествами алгоритмических и вычислительных процессов, поскольку процессы, которые могут быть представлены как подчиняющиеся правилам, но не имеющие целью формальные свойства других процессов, могут считаться алгоритмическими, но не вычислительными. Так, всякий алгоритм можно представить как множество импликаций вида «если  $A$ , то  $B$ ». Тогда, если  $A$  и  $B$  суть формальные свойства различных, но взаимосвязанных процессов, то их алгоритмическая взаимозави-

симость образует естественное или искусственное вычислительное устройство . Если же, например,  $A$  есть свойство вещи, а  $B$  — действие с нею, то процесс, выполняющий такие правила, можно считать алгоритмическим, но не вычислительным.

### 2.3.2 Уровни вычислений

Понятие уровней вычисления было впервые предложено в 1976 г. Дэвидом Марром, классиком в области вычислительной нейрофизиологии и теории зрительного восприятия, в его впоследствии переиздававшейся статье, написанной в соавторстве с Томазо Поджио [213]. В первом варианте концепции было названо четыре уровня, впоследствии их число сократилось до трёх [172, р. 24-27], и именно в этом виде данная концепция стала одной из самых цитируемых (и критикуемых) в публикациях, посвящённых когнитивным вычислениям. Итак, согласно Марру, любая вычислительная система может быть описана на трёх уровнях:

- (1) теория вычисления, описывающая его цель, приемлемость и стратегию,
- (2) репрезентация и алгоритм, где определяются формы представления входных и выходных данных, а также алгоритм преобразования одного в другое, и
- (3) физическая реализация алгоритма и репрезентаций.

Эвристическая ценность трёхуровневой концепции, по мнению её автора, состоит в том, что «поскольку эти три уровня лишь нестрого соответствуют друг другу, некоторые феномены могут быть объяснены только на одном или двух из них» [172, р. 25]. Например, как считает Марр, феномен послесвечения (после взгляда на горящую лампу) может быть объяснён только на физическом (физиологическом) уровне (3), тогда как иллюзия, известная как куб Неккера, предполагает неравновесную нейронную сеть с двумя возможными устойчивыми состояниями (уровень 2) и принципиальную возможность двоякой трёхмерной интерпретации двумерных изображений (уровень 1).

Модель Марра показывает, что, во-первых, ФСМ и ФСС не являются теоретическими альтернативами, а локализуют вычисления на разных уровнях, при том, что ФСС делает особый акцент на физической реализуемости вычислительной системы. Но ещё более интересно, что Марр использует свою схему против популярного, в том числе среди коннекционистов, тезиса о принципиальном различии (фон-неймановского) компьютера и мозга — тезиса, в котором подразумевается, что первый осуществляет серийные вычисления, а второй — параллельные. Для Марра «различие между серийным и параллельным есть различие на уровне алгоритма; оно совершенно не фундаментально — всё, что запрограммировано параллельно, может быть переписано серийно (но не обязательно наоборот)» [172, р. 27]. Ещё более интересно, что одно из наиболее существенных возражений (которое, впрочем, можно рассматривать как дополнение) против классической трёхуровневой схемы Марра было сделано его другом и соавтором Томазо Поджио спустя почти тридцать лет после выхода первого издания книги «Зрение». В послесловии к переизданию книги в 2010 г. и позже, в отдельной статье [173], Поджио утверждает, что, если в 1970-х им с Марром казалось, что компьютерная наука может многому научить нейрофизиологию, то теперь «стол поворачивается», и многие открытия вычислительной нейрофизиологии, прогресс которой во многом был обеспечен трудами Марра, могут внести существенный вклад в общую теорию вычислений.

Так, важнейшим, с его точки зрения, упущением теории Марра было отсутствие ответа на вопрос, каким образом живой организм может научиться необходимым вычислительным алгоритмам. Некоторые из современных статистических теорий машинного обучения предлагают правдоподобные ответы на этот вопрос. Поджио полагает, что эмпирически подтверждённая иерархическая организация отделов мозга, ответственных за обработку зрительной информации, представляет собой эволюционно развитую структуру, чья цель — снижения уровня сложности зрительных образцов (sample complexity) для выработки репрезентации, подходящей для эффективных вычислений.

В наше время по причине упомянутого «поворачивания стола» появляются теории вычислений, способные пролить свет на алгоритмы лежащие в основе обучения и даже эволюции. Поэтому Поджио предлагает следующий апдейт марровской схемы:

- (1) эволюция,
- (2) обучение и развитие,
- (3) вычисления,
- (4) алгоритмы,
- (5) биологический субстрат (wetware), физический субстрат (hardware), схемы и компоненты.

Главная мысль статьи, похоже, состоит в следующем: мы не сможем создать мыслящие (intelligent) машины, пока мы вынуждены программировать их интеллектуальные действия. Мы должны не закладывать в них готовые алгоритмы, а сделать их способными вырабатывать эти алгоритмы во взаимодействии с внешней средой, т. е. обучаться. Но, поскольку сама способность к обучению не дана от века, мы также должны воспроизвести в машинах эволюционные механизмы, обеспечивающие прогресс обучаемости. И сегодня это становится возможным.

В стройной концепции, изложенной в статье, есть, на мой взгляд, не очень глубоко скрытое противоречие. По крайней мере, в двух местах в тексте статьи Поджио настаивает, что знание алгоритмов обучения делает необязательным знание алгоритмов более низкого уровня. Одно из высказываний звучит так: «Теоретические и практические успехи <...> машинного обучения предполагают, что возможно разрешить трудные проблемы зрительного восприятия (и интеллекта) без понимания [их] вычислительных и алгоритмических аспектов (without a computational and algorithmic understanding)» [173, p. 1019]. Вместе с тем, он обозначает собственные предпочтения в области различных подходов к машинному обучению. Так, он отказывается обсуждать в рамках статьи какие-либо статистические теории машинного обучения, помимо теории «в стиле Вэпника и Смейла», поскольку байесовские теории хороши только для общего феноменологического описания интеллектуального поведения, но неспособны напрямую решить проблему сложности (зрительных) образцов и обобщения.

Что же касается — и это и есть вторая часть противоречия — сетей глубокого обучения<sup>23</sup>, Поджио заявляет, что они не более чем «отвлечение»

<sup>23</sup>Имеются в виду Deep Learning Networks, прямые наследники коннекционистских сетей 1980–90-х гг.

(distraction), несмотря на их сегодняшнюю популярность. Аргумент для такой оценки он заимствует в тексте Марра, и сводится этот аргумент к тому, что в теории нейронных сетей, «прежде всего, нерешённым остаётся вопрос, какие функции должны быть внедрены и зачем» (цит. по [173, р. 1019]). Т. е. нейронным сетям ставится в вину отсутствие обязательного понимания содержания вычислительного и алгоритмического уровня марровской схемы при достаточной разработанности уровня обучения. Но именно эта необязательность вычислительно-алгоритмического понимания на той же самой странице статьи была объявлена родовым признаком любой хорошей теории машинного обучения! Кроме того, насколько я могу судить по публикациям и фактам реальной жизни (точной статистикой я не владею), упомянутые Поджио «практические успехи» машинного обучения представлены сегодня в основном именно нейронными сетями глубокого обучения.

На мой взгляд, неоднозначная по качеству аргументация сторонника одной из разновидностей статистической теории машинного обучения (Machine Learning) против глубоко-нейросетевого подхода к обучению (Deep Learning) объясняется, если так можно сказать, «парадигмальным» предубеждением сторонников классицизма, к которым, безусловно, относятся Марр и Поджио, против коннекционизма и его современных нейросетевых воплощений — предубеждением, которое впервые проявилось в уже упоминавшейся полемической статье Фодора и Пилишина. И это при том, что уровневый подход Марра–Поджио как раз и помогает расставить эти и некоторые другие когнитивные парадигмы по их законным местам. Так, очевидно, что коннекционистские модели эмулируют уровень физических реализаций и в какой-то степени затрагивает уровень первичных алгоритмов, формально управляющих нейросетевым аппаратом: таких как пороговая функция, расчёт весов, обратное распространение ошибки и т. д. Если при этом возникает вычислительная система, успешно демонстрирующая некоторые когнитивные способности и способная к (само)обучению — с тренером или без, — но при этом мы не всегда можем реконструировать схемы и алгоритмы, соответствующие первому и второму уровням схемы Марра, то, согласно изначальной концепции уровней, мы, тем не менее, получили хорошее объяснение некоторому множеству феноменов.

Напротив, статистические конкуренты глубокого обучения, насколько я могу судить, концентрируются на втором уровне схемы Поджио — уровне вычислений, связанных собственно с обучением. К сожалению, автор схемы не



специфицировал, если можно так сказать, «природу» второго уровня: должен ли он представлять собой только отображение (mapping), в терминах Марра [172, р. 24], одного вида информации в другой, или он должен содержать конкретные и точные формальные алгоритмы. В первом случае статистические теории должны были бы перенесены на четвёртый, алгоритмический, уровень его схемы, что, может быть, было бы правильно, но, похоже, Поджио не это имеет в виду. Из контекста понятно, что предпочитаемая им статистическая теория должна располагаться на втором уровне его схемы, в значительной степени предопределяя содержание третьего, вычислительного, уровня (или первого в схеме Марра), поскольку, по его словам, наше понимание (а следовательно, и программирование) вычислительного и алгоритмического уровней более не является необходимым.

Важно отметить принципиальное, или, если угодно, концептуальное, отличие противопоставляемых подходов: Machine Learning (ML), основанного на той или иной математико-статистической теории, и Deep Learning (DL), выросшего из теории и практики нейросетей. Между ними есть важные технические и математические различия, но для философского анализа важно другое: если DL представляет собой, пусть абстрактную, но имитацию биологической реальности; модель, в которой программист задаёт только самые общие принципы и константы параллельных вычислений, не предписывая никаких собственно когнитивных действий и способностей — т. е. создаёт аппаратную платформу (hardware), которая не нуждается в программном обеспечении в привычном смысле слова, то создатель ML-систем продолжает действовать, по сути, как старый добрый программист, только вынужденный работать с существенно более сложными программными средствами. Для того, чтобы избавиться себя от заботы о собственно когнитивных алгоритмах, он или она вынуждены писать алгоритмы обучения машины когнитивным алгоритмам, а перемещаясь на уровень вверх, соответственно, эта необходимость заменяется необходимостью писать алгоритмы эволюции алгоритмов обучения алгоритмам когнитивных действий. Необходимость подобной пирамиды программирования обусловлена самой сутью классицистского подхода «сверху-вниз»: важно понять телеологию процесса, спроецировать её в алгоритм, выполняемый на «железе», и, если получилось, объявить, что найдено объяснение сложным природным феноменам. Это как если бы компьютерные игры, изображающие автогонки, использовались для объяснения устройства и движения реальных автомобилей.

### 2.3.3 Типы вычислений

Чтобы выйти за пределы ФМС и ФСС, составляющих основу классицизма, обратимся к популярной ныне концепции механистической системы как онтологической основы вычислений в более широком их понимании. Крейвер и Бектель [174, р. 469–470] выделяют четыре признака механизма и, соответственно, четыре требования к механистическому объяснению: (1) феноменальный аспект — задача, которую решает система, (2) структурный аспект — всякий механизм состоит, как минимум, из двух элементов, (3) каузальный аспект — элементы механизма находятся в причинном взаимодействии — и (4) пространственно-временной аспект — для механизма в целом важна локализация его элементов в пространстве и во времени. Вычисления в таком случае понимаются как последовательность состояний механистической системы, в которой ввод (input) — это исходное состояние системы, а выводом (output) в разные моменты времени являются её последующие состояния. Соответственно, как считает Фреско, более слабые, чем символизм, концепции цифровых вычислений опираются уже не на тьюринговскую концепцию, а на алгоритмические описания деятельности механистических систем. К альтернативным концепциям цифровых вычислений относятся, в частности, конечные автоматы, машина Ганди, дискретные нейронные сети и суперкомпьютеры. Возможность этих и подобных реализаций позволяет утверждать, что когниция есть алгоритмическое вычисление, которое не обязательно должен быть символическим [187, р. 361].

Итак, постепенное обобщение разновидностей компьютеризации выглядит следующим образом: (а) символизм, (б) слабый цифровой компьютеризация и (в) общий компьютеризация, включающий цифровой, аналоговый, квантовый и нейрофизиологический.

Некоторую проблему представляет точная локализация коннекционистских вычислений на этой шкале общности. Нейросети имеют разную архитектуру и обрабатывают разные типы данных. Какие-то их разновидности могут рассматриваться как подмножество цифровых (несимвольных) вычислений, какие-то — как подмножество аналоговых. Но в любом случае коннекционизм, если, согласно важной оговорке Уильяма Рэмзи, речь идёт о тренируемых, а не программируемых сетях, не имеет точек пересечения с символьным подходом: «Когда коннекционисты изучают роль весов [межнейронных связей] в сетевых

вычислениях, их первый вопрос — не “Какое правило кодирует данный вес?”, а скорее “Какой каузальный вклад вносит данная связь в обработку данных?”» [153, р. 49]. И это, конечно, показывает, что, даже в дискретных коннекционистских сетях вычисления могут быть рассматриваемы как цифровые только в слабом смысле — не на основе абстракции машины Тьюринга, а на основе, например, абстракции механистической системы или каких-либо других альтернативных моделей.

Наиболее интересную проблему представляет собой точная идентификация и локализация нейрофизиологических вычислений в рамках предложенной таксономии. На первый взгляд, эта разновидность должна быть родственна коннекционистским (нейросетевым) вычислениям и, соответственно, подчиняться логике и математике, применяемым в коннекционистских сетях. Однако это не так, и тому есть несколько причин. Во-первых, искусственные нейроны представляют собой очень простые процессоры, по большей части виртуальные, тогда как нейроны мозга существенно более сложны физиологически, химически и генетически. Во-вторых, важной частью обучения искусственной нейросети является алгоритм обратного распространения ошибки, что предполагает возможность передачи данных между слоями в обратном направлении. Напротив, устройства ввода и вывода биологического нейрона — соответственно, дендриты и аксоны — предполагают только однонаправленную передачу сигнала. В-третьих, в коннекционистских сетях передача данных идёт последовательно от слоя к слою, тогда как архитектура мозга значительно сложнее и во многом не до конца ясна самим нейрофизиологам.

Но наиболее интересное отличие состоит в том, что, если коннекционистские сети в принципе подчиняются главному принципу Марра, утверждающему относительную независимость уровней рассмотрения вычислительной системы, то в нейрофизиологических вычислениях материальная архитектура биологической нейросети играет решающую роль в определении возможных алгоритмов вычислений. Таким образом, если марровский классицистский подход характеризуется в ныне популярных терминах как подход «сверху вниз» (top-down), то нейрофизиологические вычисления демонстрируют противоположный — bottom-up — подход «снизу вверх», где понимание сути вычислений предопределяется знанием фактов на уровне физической реализации [187, р. 367].

При том, что вычислительная нейрофизиология остаётся пока в высокой степени дискуссионной областью, практикуемый здесь подход «снизу вверх»

позволяет предположить, что нейрофизиологические вычисления представляют собой особый тип, не локализуемый не только в дихотомии «сильный (символьный) — слабый компьютеризм», но и в дихотомии «цифровые — аналоговые вычисления». Серии пиков нейронных возбуждений (spike trains) демонстрируют как аналоговые, так и цифровые свойства: кривая электрических потенциалов на внешних «портах» клеток континуальна, но сама последовательность пиков образует дискретную структуру. Можно предположить, что вычислительная нейрофизиология как научная дисциплина ещё слишком молода, и ей ещё предстоит определиться, с какого рода вычислениями ей приходится иметь дело.

### 2.3.4 Антикомпьютеризм

Учитывая все эти сложности, неудивительно, что раздаются голоса, отрицающую какую-либо эвристическую ценность за вычислительным подходом как таковым. На первый взгляд, действительно, если, обнаружив ограниченную применимость классического тьюринговского понимания вычислений в различных предметных областях, мы делаем уступки физике в виде замены машины Тьюринга на абстракцию механистической системы, в виде отказа от строгости трёхуровневой схемы Марра в пользу определяющей роли физического уровня, а также будучи неспособны отнести нейрофизиологические вычисления к какому-либо определённом типу, то стоит ли вообще настойчиво придерживаться этой абстракции как объяснительного принципа с тем только, чтобы отдать дань духу времени? Не лучше ли вернуться к старой доброй физике с её прямолинейной каузальностью?

Как было сказано выше, это направление было принято радикальными коннекционистами, радикальными энактивистами и сторонниками динамических систем (ДС). Первая из этих точек зрения утверждает, что коннекционистские сети никоим образом не являются вычислительными устройствами, не является таковым и человеческий мозг. Напротив, как сети, так и мозг являются «сенсорно-моторным процессором, который научили вести себя определенным образом, только благодаря взаимодействию с непредвиденными обстоятельствами этого изменяющегося мира» [214, р. 313]. Ту же или аналогичную позицию

разделяют радикальные энактивисты: “Под влиянием феноменологии, теории динамических систем и прогресса робототехники сторонники энактивистского и «воплощённого» образа мысли отвергают привычные объяснительные рамки ортодоксальной когнитивной науки в пользу альтернативных платформ” [179, р. 1]. Можно предположить, что все три «посткогнитивистские» тенденции гораздо более совместимы друг с другом, чем их оппоненты. Так, ДС могут рассматриваться как формальный инструмент для теоретизирования на основе радикально-коннекционистских или энактивистских онтологий. Фреско считает, что ДС сами по себе вообще не имеют каких-либо средств механистического объяснения, хотя в принципе совместимы с механистическими моделями [187, р. 371]. Это означает, что данная теория в её чистом виде не имеет встроенных средств демонстрации компонентной структуры своей модели, включая причинное взаимодействие компонентов и поэтому вынуждена заимствовать такие модели из других радикальных парадигм. Что касается последних, чей решительный успех ещё впереди, есть естественное подозрение, что полный отказ от вычислительных описаний означает шаг назад в научном объяснении сознания, так как традиционные каузальные и натуралистические теории пока не привели к распутыванию связанных с ним философских узлов. Я не хочу сказать, что та или иная версия компьютеризации есть гарантированный путь к решению проблемы, но, по крайней мере, стоит попытаться настроить это базовое понятие таким образом, чтобы оно соответствовало известным эмпирическим фактам и работающим моделям.

### 2.3.5 Термодинамическая стоимость вычислений

Одним из возможных способов такой настройки является увязка общей концепции вычисления с его термодинамической стоимостью. Согласно [177], последняя связана с понятием плотности энергии физических и биологических систем, причём живые существа демонстрируют большую энергетическую эффективность по сравнению с астрономическими объектами, а человеческие общества оказываются самыми расточительными из всех вычислительных систем [177, р. 3]. Меньшие затраты энергии на единицу массы в неравновесных системах, процессы в которых на основании некоторых критериев можно

рассматривать как вычисления, означают большую термодинамическую эффективность последних. Но вычислительная задача, скажем, для биологических систем, по-прежнему представляется довольно сложной, поскольку они фактически выполняют многоуровневые вычисления, где более высокие уровни «поглощают» более низкие [177, р. 6]. Авторы моделируют свою теорию на термодинамической стоимости трансляции РНК как самого простого и очевидного вида биологических вычислений, но будущие исследования, по их мнению, должны охватить некоторые из более сложных явлений жизни, познания и социальных взаимодействий высших млекопитающих — всё это подлежит объяснению средствами вычислительной теории, основанной на неравновесной статистической физике.

Одно из возможных эвристических предположений может заключаться здесь в том, что вышеупомянутая многоуровневая архитектура биологических вычислений способна пролить свет на то, что я хотел бы обозначить как онтологию вычислений как таковых. Если природа использует в разных системах своего рода механизм надстраивания уровней, то каждый более высокий уровень вызывает некоторые следствия в более низких — это функциональное отношение часто называют *нисходящей причинностью*. Но должен быть также учтен его восходящий аналог, так как любой более высокий уровень иерархии фактически является некоторой функцией более низких. Понятие термодинамической стоимости вычислений помогает понять этот механизм надстраивания уровней как эволюционную тенденцию к сокращению энергетических затрат вычислений. Переход на уровень вверх позволяет влиять на положение вещей на более низком уровне с меньшими энергетическими затратами за счет нисходящей причинности — аналогично тому, как усилитель руля влияет на рутинные задачи вождения автомобиля. И, поскольку закономерности более высокого уровня в некотором смысле представляют факты более низких, в этом отношении его функционирование может быть принято как обработка информации или вычисление. Такой взгляд имеет смысл, особенно в отсутствие более традиционной научной теории, которая могла бы связать эти разные уровни.

\* \* \*

В контексте того, что было сказано об онтологии надстраивания уровней в естественных вычислительных системах, было бы разумно предположить, что

структурные элементы более высокого уровня функционируют как репрезентации некоторых состояний нижнего уровня, но это репрезентативное отношение не является семантическим — оно является особым типом причинной связи между уровнями, который позволяет снизить термодинамическую стоимость системных взаимодействий. В этом смысле температура представляет собой репрезентацию броуновского движения молекул, подходящую для чувствительных биологических организмов.

Принятие нами «слабой» концепции репрезентации влечет за собой другое рациональное предположение: не существует прямых инструментов или прямых способов репрезентации субъекту «мира» или содержащихся в нём «вещей». На пути к тому, что мы называем «знанием», информация проходит через несколько передающих уровней и репрезентирующих сред и, таким образом, через несколько уровней репрезентации. Таким образом, то, что в конечном итоге появляется как единое, структурированное и классифицированное «видение мира», является результатом чрезвычайно сложного взаимодействия репрезентирующих, интегрирующих и контролируемых подсистем нашей когнитивной фабрики, не говоря уже о социальных вычислениях, которые здесь почти не упоминаются — таких как язык и различные формы передачи коллективного опыта.

\* \* \*

Итак, представленные здесь соображения позволяют прийти к следующим заключениям:

1. Классические объяснения вычислений выглядят сравнительно простой когнитивной теорией, которая готова к непосредственной реализации в компьютерных моделях. Но на концептуальном уровне они чреватны парадоксами и не имеют биологически реалистических интерпретаций, если говорить о познавательной деятельности человека или животных.
2. Более слабые версии компьютеризации частично приносят в жертву вычислительный дискурс в пользу физики и биологии, что приближает их к естественнонаучным взглядам на эволюцию и прижизненное обучение живых существ. Но в результате основные понятия когнитивной науки теперь менее чем достаточны для «хорошей» теории, оставаясь, самое большее, необходимыми в ней. И потому,

3. Необходима разработка «родовой» (generic) теории вычислений, которая представила бы вычисление как потенциально многоуровневый процесс предполагающий представленность одних процессов в других с помощью некоторых репрезентационных форм и механизмов.

### 2.3.6 Вывод из раздела 2.3

Когнитивная наука нуждается в опоре на «вычислительную метафору», но последняя должна выйти за тесные тьюринговы рамки, избавиться от других частностей, связанных с особенностями отдельных дисциплин, и стать **общей (родовой) теорией, объединяющей предметные дисциплины, изучающие процессы в сложных системах.**

## 2.4 Репрезентации в когнитивных науках

### 2.4.1 Понятие репрезентации в исторической перспективе

Компьютерная революция XX в. открыла новые перспективы не только в математике и прикладных науках, но и в психологии, и даже в лингвистике. Поскольку компьютер воспринимался как интеллектуальная машина, он естественным образом стал моделью человеческого интеллекта, породив «компьютерную метафору» в психологии. И, поскольку тьюрингова модель вычислений представляла их как правилосообразные операции с символами, науки о языке также оказались включёнными в новую парадигму. Необихевиоризм как непосредственный предшественник когнитивной психологии видел идеал научности в отказе от менталистских терминов. Новой психологической школе удалось восстановить их в правах без малейшего намёка на метафизику. Это случилось благодаря возрождению на новой основе старой идеи Томаса Гоббса: мышление есть вычисление. Вычисления в их классическом тьюринговом понимании суть операции над символами, которые представляют некое внеш-



нее содержание и, следовательно, являются репрезентациями. Таким образом, вычисления и репрезентации — два опорных концепта в рамках когнитивных подходов в различных науках. Вокруг них разворачиваются основные философские дискуссии об основаниях когнитивной парадигмы.

Тема вычислений как специального предмета теории — прежде всего и в основном математической — привлекла внимание в 1930-х гг., когда Алан Тьюринг, думая над проблемами гильбертовского обоснования математики, предложил теоретическую абстракцию, известную как машина Тьюринга [183], позволившую лучше обозначить класс вычислимых функций. Поскольку идеи Тьюринга легли в основу теории и практики современных компьютеров, его подход долгое время считался — а многими и по сей день считается — теорией вычислений в собственном смысле слова. Однако с появлением новых подходов — квантовые, аналоговые, генетические, параллельные вычисления — встал вопрос о границах тьюринговой парадигмы, породив широкий спектр исследований в области нетьюринговых моделей [215]. Таким образом, проблема вычислений — их определения, онтологии, спектра применимости понятия — постепенно стала общенаучной и в какой-то степени философской. В общем виде вычислительные процессы, в отличие от простых динамических, можно охарактеризовать как алгоритмические и существенно независимые от материальных реализаций. Иными словами, любые структурные изменения, которые можно описать как последовательность состояний, смена которых подчиняется определённым правилам («если — то»), и которые можно реализовать более чем в одном материальном субстрате, суть вычисления. Естественным образом появилась концепция «натуральных вычислений» [151] — вычислительных процессов в природе, таких как трансляция РНК и последующий синтез протеинов. Это дало начало таким научным дисциплинам, как вычислительная термодинамика, вычислительная биология, вычислительная астрономия, экономика и т.д. Так как в самом понятии вычислений присутствует изрядная доля антропоморфизма, дело часто представляется так, что вычисления непременно связаны с обработкой информации. Но, поскольку эта тема до сих пор является в высшей степени дискуссионной, ограничимся лишь следующим соображением: об информационном аспекте в вычислениях можно говорить, если имеет место инвариантная функциональная взаимозависимость изменений в различных структурах — как в уже упомянутых РНК и протеинах. Можно с некоторыми оговорками сказать, что информация как бы добавляет телеологическое изме-

рение в мир: изменение  $a$  в структуре  $A$  происходит для того, чтобы вызвать изменение  $b$  в структуре  $B$ . Вопрос же о том, является ли информация необходимым элементом вычислений, безусловно, заслуживает отдельного обсуждения, которое, впрочем, выходит за пределы, положенные целями настоящего исследования. Однако, как станет ясно далее, сама его постановка проливает свет на концептуальные проблемы когнитивных наук. Так, если некая структура или последовательность элементов обрабатывается в вычислительном процессе именно как информация о состоянии другой структуры, то первая как бы представляет — репрезентирует — вторую, т. е. является её репрезентацией в данном вычислительном процессе.

Поскольку первые инкарнации когнитивной психологии и лингвистики в 1950-х гг. существенно направлялись восходящей звездой компьютерной теории и промышленности, вычисления и репрезентации стали двумя концептуальными столпами, на которых зиждется когнитивистская парадигма. Они оставались таковыми на классическом этапе развития когнитивных наук, который в литературе принято обозначать терминами «символизм», «классицизм» и «компьютериализм». В основу современных компьютеров архитектуры фон Ноймана была положена машина Тьюринга, идея которой основана на алгоритмической обработке последовательностей символов. Несмотря на то, что, по некоторым свидетельствам, сам Алан Тьюринг незадолго до смерти начал разрабатывать альтернативные модели вычислений — в том числе распределённые и генетические — его соответствующие рукописи не были вовремя опубликованы, и слава за эти открытия досталась другим людям [216]. Ирония истории состоит в том, что Тьюринг *de facto* заложил основы нетьюринговых вычислений, но об этом мало кто знает. Серийные же символьные вычисления, теория которых связана с его именем, оказались «на коне» по исключительно практическим соображениям — в те времена (а во многом и в наши тоже) любая альтернатива была более ресурсоёмкой [217]. Исследования Питтса и Маккаллока в области нейронов и «перцептрон» Розенблатта (соответственно, в 1940-е и в 1950-е гг.) не принесли результатов, пригодных к использованию в промышленности или в обороне, и попытки реализации распределённых вычислений в нейронных сетях были приостановлены до начала 1980-х гг. Поэтому компьютер фон-ноймановской архитектуры стал на десятилетия главной метафорой для когнитивных способностей живых субъектов.

В 1986 г. выходит фундаментальный труд группы авторов под названием «Параллельный распределённый процессинг» [55; 56]. В двухтомнике были продемонстрированы теоретические разработки и результаты экспериментов с когнитивными моделями, представляющими собой эмулированные нейронные сети. В таких моделях данные, подаваемые на входной слой сети, случайным образом прокладывают себе путь через внутренний слой к выходному слою, меняя по пути количественные характеристики связей между нейронами, называемые «весами». Вес каждой связи означает сравнительную вероятность прохождения сигнала по этой связи в будущем. Данные, получаемые на выходе, интерпретируются как ответ сети на вопрос, заданный ей на входе. Если ответ неправильный с точки зрения экспериментатора-тренера, то запускается алгоритм «обратного распространения ошибки», и уже вычисленные веса связей вновь пересчитываются. Так происходит обучение сети на большом количестве данных, пока сеть не научается давать правильные ответы. В качестве данных может использоваться что угодно: изображения животных, форма прошедшего времени неправильных глаголов и т. п.

Новый подход в когнитивных исследованиях долгое время после его появления был известен как коннекционизм. Его успехи в биологически реалистичных объяснениях когнитивных способностей поставили новые вопросы перед философами-методологами: вместо правилосообразных операций с символами мы имеем теперь многочисленные итерации пересчёта весов межнейронных связей, вместо простого семантического отношения символа к значению — векторные величины активации нейронных ансамблей. Стала очевидной потребность в переосмыслении опорных концептов — вычислений и репрезентаций, референты которых в новой когнитивной модели локализовались не на уровне процессинга<sup>24</sup>, а на уровне архитектуры. Сигналом к такому переосмыслению была известная в истории науки атака Джерри Фодора и Зенона Пилишина на семантические возможности коннекционистских сетей [218]. В этой статье авторы — основоположники классицистского взгляда на когнитивную архитектуру — выдвинули аргументы, основанные на комбинаторном характере ментальных репрезентаций, следствием которого являются их продуктивность (способность порождать неопределённо большое количество мыслей из конечного числа элементов), систематичность (способность мыслить некоторые содержания, отличные от данного, но семантически связанные с

---

<sup>24</sup>См. определение на с. 306.

ним) и композициональность (зависимость значения целостной репрезентации от значений её составных частей). По их мнению, этот аргумент показывает, что архитектура сознания не является коннекционистской на когнитивном уровне, поскольку коннекционистские репрезентации не демонстрируют этих свойств. Впрочем, они допустили, что коннекционизм может давать представление о нейронных (или «абстрактно неврологических») структурах, на которых реализована классическая когнитивная архитектура.

Затем последовала дискуссия с участием Пола Смоленски [219] который утверждал, что критики не учли распределённый характер коннекционистских репрезентаций и что коннекционистские исследования должны предложить новые формализации фундаментальных вычислительных понятий, которые до сих пор были формализованы одним определённым образом в традиционной символической парадигме — очень важное замечание в контексте нашего обсуждения основополагающих понятий когнитивной науки. Часто цитируемая статья Уильяма Рэмзи [153] многими рассматривается как важный вклад в интересующую нас дискуссию. Рэмзи показал, что однонаправленные сети с обратным распространением ошибки не нуждаются в понятии репрезентации как таковом. Это породило целый букет антирепрезентационалистских концепций, в том числе отрицающих какую-либо роль этого понятия в когнитивных науках вообще. Следующим вызовом классическому когнитивизму оказались нетождественные, но соотносимые концепции, известные как энактивизм и *embodied mind*. Сторонники этих подходов, в многом черпающие вдохновение в феноменологической философии, противопоставляют пониманию когнитивного аппарата как встроенного компьютера, манипулирующего символами, новый подход, располагающий когниции<sup>25</sup> в пространстве между мозгом, телом и окружающей средой. Сторонники этого направления сосредотачиваются скорее на моторике и бессознательных реакциях, а не на символьных вычислениях. Как и в коннекционизме, антирепрезентационализм здесь не необходим, но возможен. Так, «радикальный энактивизм», манифест которого сформулирован в работах Дэниела Хутто [178; 179; 220; 221], провозглашает отказ не только от репрезентационализма, но и от компьютерационализма как такового, обозначая новую парадигму как пост-когнитивизм.

Сюда же примыкает относительно новое направление, часто обозначаемое как «динамические системы», имея в виду используемый там математиче-

---

<sup>25</sup>См. определение на с. 306.

ский аппарат [180]. Когнитивные процессы помещаются здесь в многомерное «пространство состояний», измерения которого представляют параметры системы, выраженные переменными с изменяющимися во времени значениями. Когнитивные процессы описываются как сложные траектории в этом  $n$ -мерном динамическом пространстве. Когнитивная наука должна обнаруживать инварианты в этих движениях и устанавливать функциональные связи между ними — во многом подобно тому, как действуют естественные науки довычислительной эпохи. Если попытки проникнуть в конкретные механизмы когнитивных процессов предполагают обращение к понятиям вычислений и репрезентаций, то в динамическом подходе они оказываются в общем и целом неуместны. На другом конце спектра позиций относительно роли репрезентаций располагается, пожалуй, новейшее из конкурирующих направлений — Predictive Processing [61—63] — возвращающее нас, по мнению некоторых авторов в лоно строгого и крайнего репрезентационализма [222]. Предиктивный процессинг<sup>26</sup> основывается на модели когнитивного аппарата как иерархии статистических вычислительных машин, задача которых — предсказание входных данных. В качестве математического формализма используется байесовская статистика. Эти машины детерминируют другие машины, расположенные рядом и ниже в иерархии, создавая антиципаторную модель мира. Входящие данные оцениваются системой по энергетической интенсивности их воздействий и передаются дальше поверхностных датчиков только в случае их несовпадения с построенной моделью. Причём статистические предсказания начинаются уже на уровне сенсорных модулей, преодолеть уровень которых способны только не ожидаемые ими данные. По мнению создателей и адептов этого направления, заложенная здесь методология исключительно перспективна не только в изучении традиционных тем когнитивной психологии, но и в поиске решений «трудных проблем», таких как квалиа, эмоции и даже психические расстройства. Однако в задачу настоящего исследования не входит какая-либо оценка конкурирующих концепций: их широкий спектр интересует нас в отношении к основополагающим когнитивным понятиям. И, как мы видим, по отношению к проблеме репрезентаций, существующие парадигмы могут быть выстроены по трехступенчатой шкале: (1) сильный репрезентационализм — (2) слабый репрезентационализм — (3) антирепрезентационализм. Точное позиционирование интересующих нас подходов в

<sup>26</sup>Более подробно об этом см. параграф 2.2.5.

этой системе координат возможно, однако, только при условии учёта глубоких концептуальных изменений, происходящих сегодня в естественных науках.

#### 2.4.2 Трудности и парадоксы в современном понимании репрезентаций

В классической когнитивной науке, как указывает Нир Фреско [187, р. 356], считается, что репрезентации должны обладать двумя важными свойствами — быть физически реализуемыми и быть интенциональными. Интенциональность понимается также классически: как наличие значения или содержания, т. е. того, репрезентацией чего она является. Физическая реализуемость предполагает наличие физически допустимого транспорта (vehicle) репрезентаций, в качестве которых могут выступать вычислительные структуры или состояния мозга. В рамках такого понимания репрезентации, действительно, являются физически воплощёнными сущностями, обладающими семантикой, т. е. символами. Над ними возможны Тьюринг-вычислимые операции, и вся модель когнитивных актов оказывается полностью аналогичной работе компьютера фон-неймановской архитектуры. К несомненным достоинствам классической модели можно отнести её непосредственную компьютерную реализуемость: один из классиков символического подхода Дж. Р. Андерсон создал компьютерную платформу АСТ-R для моделирования когнитивных функций с целью последующей экспериментальной проверки моделей [186]. Объяснительный принцип построен на научной абдукции: если модель показывает те же результаты, что и живой испытуемый, то с высокой степенью вероятности, когнитивный аппарат испытуемого имеет ту же структуру, что и компьютерная модель.

И вот здесь начинаются проблемы. Исторически когнитивистская парадигма торжествует после победоносной полемики Н. Хомского и Б. Ф. Скиннера в конце 1950-х гг. [223]. Теория врождённой генеративной грамматики как бы смогла объяснить продуктивность человеческого языка — способность составить и понимать ранее не слышанные высказывания. Лингвистика победила необихевиористскую психологию в союзе с восходящей компьютерной наукой. Неудивительно, что новый когнитивный подход имел ярко выраженные линг-

вистические родовые черты: представление о психической жизни как о потоке вычислительных операций над символами, обладающими семантикой.

Закономерным следствием этого видения явилась концепция «мысленного языка» (*Mentalese*), выдвинутая Джерри Фодором, согласно которой нашим операциям с внешними символами соответствуют внутренние манипуляции с символами-репрезентациями, такие, что символы имеют ярко выраженную знаковую природу, а сами операции аналогичны высказываниям в естественном языке. Именно язык мысли, согласно Фодору, является основой нашего понимания языка общения [54; 224]. Такая позиция не должна непременно уводить в дурную бесконечность, тем более, что Фред Эттнив ещё в 1959 году представил механистическую модель когнитивного аппарата, позволяющую избежать *парадокса гомункула* за счёт перераспределения функций между уровнями организации системы [225]. Однако, несмотря на преодоление парадокса на уровне дизайна, на концептуальном уровне его опасность сохраняется, если его сформулировать в следующем виде:

(ПГ1:) Чтобы распознать внешнее содержание в символе, необходимы когнитивные способности. Но именно их мы пытаемся объяснить с помощью этой схемы.

Предположим, можно найти техническое объяснение, каким образом когнитивная система распознаёт синтаксические свойства внутренних символов. Но откуда берётся содержание, делающее их интенциональными? И кто это содержание считывает? Одним словом, возникает серьёзное подозрение, что объясняемое содержится в объяснении. Как замечает Фреско, «внешние репрезентации зависят от внешнего субъекта (*knower*): субъект приписывает объектную семантику структурам данных, строчками и символам» [187, р. 358]. Не случайно поэтому исследования в рамках символической (классицистской) парадигмы более всего преуспели в объяснении языковых способностей и языковой деятельности.

Кроме того, остаётся непреодоленной двусмысленность самого понятия репрезентации: имеется ли в виду состояние когнитивного аппарата или психическое (феноменальное) состояние, или, иными словами, — говорим ли мы об объективных или субъективных репрезентациях? Последние выглядят более законной областью понятия репрезентации, поскольку выполняют роль представителя объективных положений дел перед мысленным взором субъекта (хотя

и это не всегда справедливо). Что же касается объективно регистрируемых состояний когнитивного аппарата, то, на мой взгляд, такое понимание репрезентации играет в классицизме скорее нормативную роль: все исходят из того, что таковые репрезентации должны быть, поскольку они предусмотрены в используемых компьютерных моделях.

В рамках различных проектов ‘brain reading’ — «чтения мозга» — имеются результаты эмпирических исследований, демонстрирующие возможность построения функциональных отношений между паттернами активации определённых отделов мозга и внешними стимулами. Так, в [226] была сделана попытка с помощью математического моделирования продемонстрировать функциональную связь между произнесённым словом и паттерном активации верхней височной извилины, ответственной за высокоуровневую обработку семантически нагруженной акустической информации. Пациентам, оперируемым в связи с эпилепсией или опухолями мозга, были имплантированы датчики в эту область, с помощью которых стало возможно реконструировать структуру нейронных возбуждений, возникающих при восприятии пациентом реальных или специально придуманных слов. Далее были построены несколько математических моделей, описывающих функциональные связи между паттернами активации и волновой формой звучащего слова. Затем наиболее релевантная модель использовалась для обратной реконструкции акустического образа из нейронных импульсов. Был получен неоднозначный результат: звуковой образ «восстановленных» слов не распознавался слушателями, но визуально, тем не менее, волновая форма полученного сигнала соответствовала таковой реально произносимых человеком слов. Исследователи предположили, что, по мере совершенствования технических и математических средств, станет возможной разработка средств коммуникации с пациентами, лишёнными речи, например, вследствие паралича.

Очевидно, что для конкретной научной области этот результат является промежуточным. Но в рамках концептуального исследования можно предположить, что эмпирический поиск увенчался успехом, и найден способ трансляции как звукового образа слова в нейронные ансамбли, так и обратно. Тогда мы должны принять, что структура ансамбля возбуждённых нейронов есть в точном смысле слова объективно регистрируемая репрезентация звука произносимого слова. И это, скорее всего, будет справедливо. Другое дело, что так понятая «репрезентация» не является достаточно операционализированным по-



нятием в контексте когнитивной проблематики и не предоставляет достаточных концептуальных средств для решения философских и когнитивно-психологических проблем, связанных с сознанием и его многочисленными загадками.

Тому есть несколько причин. Во-первых, в таком случае паттерн нейронного возбуждения в такой же мере является репрезентацией звука произносимого слова, в какой, наоборот, звуковые колебания суть репрезентация структуры возбуждения нейронного ансамбля. И уже при этом соображении в самом понятии репрезентации нет ничего специфически когнитивного. Во-вторых, подобное расширенное, или «слабое», понимание репрезентации ведёт к позиции, которую можно было бы назвать панрепрезентационализмом, по аналогии с панкомпьютеризмом. Мозговые структуры на тех же основаниях могут считаться репрезентациями внешних событий, на каких синтезированный протеин может считаться репрезентацией цепочки нуклеотидов РНК — или наоборот, что не принципиально. Таким образом, в основу объяснения мира когнитивных явлений кладётся понятие, охватывающее широкий спектр некогнитивных явлений. С точки зрения логики объяснений и определений, такое понятие может в лучшем случае служить родовым, ничего не говоря о видовых отличиях объясняемого феномена. Иными словами, для теории сознания — философской или психологической — это понятие в такой степени разработанности не может быть достаточным.

### 2.4.3 Позиция «слабого» репрезентационализма

В свете этих соображений более рельефно обозначаются методологические недостатки классического репрезентационализма. На мой взгляд, классическое понимание репрезентации исходит из примитивной схемы когнитивного субъекта, окруженного объектами, которые отражаются в представлениях. На самом деле все гораздо сложнее. Вся концепция семантически нагруженных внешних репрезентаций объектов недостаточно обоснована и убедительна; скорее можно было бы говорить о субъективных внешних репрезентациях: например, о цвете как репрезентации некоторого спектра электромагнитного излучения и т. п. Подобно субсимвольным вычислениям, на которые опирается коннекционизм, было бы правильно говорить о субсимвольных репрезентациях — как

внешних, так и внутренних: например, векторная сумма весов нейронной сети может рассматриваться как представление категориальной структуры данных, на которых сеть была обучена.

Тогда репрезентации имеют смысл только в контексте вычислительных процессов в системах, где такие процессы релевантны:

*(Df4)* Структура А есть репрезентация структуры В в рамках некоторого вычисления, тогда и только тогда, когда в этих же самых рамках А и В связаны устойчивой инвариантной функцией.

Принимая такое «слабое» определение репрезентации, мы оказываемся дальше от конечной цели когнитивной теории, поскольку так понятые репрезентации не обязательно позволяют нам объяснить процесс получения знания в его окончательном виде. Но это единственный способ избавиться от «гомункула» и увидеть познание как процесс в сложной многоуровневой вычислительной системе. Другая трудность может быть связана с излишне антропоморфным пониманием «семантического» отношения репрезентации к тому, что репрезентируется. Подобно тому, как в человеческом мире знаки и их значения связаны между собой конвенционально, здесь также люди должны знать, а, следовательно, быть обучены этим связям. Если мы позаимствуем эту семантическую теорию для классической версии когнитивной науки, мы рискуем столкнуться с другой версией парадокса гомункула<sup>27</sup>:

*(ПГ2:)* Для того, чтобы символическое вычисление было семантически эффективным, когнитивная система должна «знать» семантические отношения между символами и их референтами. Но любое знание представляет собой (или имеет в своей основе) репрезентацию. Тогда для каждой репрезентации требуется другая, поддерживающая её репрезентацию, и так далее до бесконечности.

В случае чисто синтаксического вычисления мы избегаем этого парадокса, но оставляем механизм, благодаря которому психические состояния вообще имеют содержание (т. е., интенциональность) необъяснённым. Но если, потерпев неудачу с компьютерно-информационными теориями, мы отступим к чисто и прямолинейно причинно-следственным объяснениям репрезентации, то мы в конечном

<sup>27</sup>Первую версию этого парадокса см. на с. 183

итоге упустим смысл всего когнитивного предприятия. Таким образом, это еще один аргумент в пользу более слабого понимания репрезентации, с тем чтобы она, не будучи необходимой в рамках философии сознания, тем не менее, оставалась в когнитивной науке как полезный объяснительный принцип. В контексте современных компьютерных теорий было бы разумно предположить, что репрезентации являются не только объектом, но прежде — результатом вычислений. Вычислительные процессы, осуществляемые на более высоком системном уровне, принимают выходные данные процессов более низкого уровня в форме репрезентаций, но это репрезентативное отношение не является семантическим — оно представляет собой особый тип причинной связи между уровнями, что позволяет повысить эффективность системных взаимодействий, например, понизив их термодинамическую стоимость [177]. В этом смысле, скажем, ощущение теплоты представляет собой репрезентацию броуновского движения молекул, а зрительные образы — репрезентацию результатов нейробиологических вычислений, осуществляемых в сетчатке и зрительных нервах, при этом обе репрезентации подаются в качестве данных на вход сенсомоторных процессоров.

«Слабый» репрезентационализм в моей версии предполагает одно важное философское следствие: понятие репрезентации в том виде, в котором оно сохраняется в когнитивном тезаурусе, недостаточно для выражения таких глобальных тем, как познавательное отношение субъекта к миру. Это понятие уместно при интерпретации когнитивной «фабрики» как иерархии вычислительных процессов для обозначения формы детерминации одних процессов другими. Последовательные репрезентации образуют многоуровневый путь прохождения данных. Общий «выход», данный нам в виде единой упорядоченной картины мира, является результатом сложного взаимодействия взаимно детерминированных когнитивных процессов, которые обмениваются данными и результатами вычислений в приемлемых для них репрезентативных формах. За пределами настоящей главы остаются социальные вычисления<sup>28</sup>, которые — в виде языка, производного от него рационального дискурсивного мышления, норм и культурных сценариев — достраивают уникальную когнитивную сферу *homo sapiens*.

\* \* \*

---

<sup>28</sup>Подробнее о них см. параграф ??.

Представленные в этом разделе соображения, на мой взгляд, позволяют прийти к следующим заключениям:

1. Репрезентации, как они понимаются в классической когнитивной науке, более всего приспособлены к непосредственным компьютерным реализациям на современном уровне развития технологий. Но на концептуальном уровне они чреватны парадоксами и слишком зависимы от лингво-семантической объяснительной схемы.
2. Радикальные антикомпьютационалистские программы в когнитивной науке частично или полностью приносят в жертву вычислительный дискурс в пользу невычислительной нейрофизиологии и биологии, что приближает их к традиционным естественнонаучным взглядам на эволюцию и прижизненное обучение живых существ. Но этот же шаг лишает их главного преимущества когнитивного подхода — расширения эмпирической сферы за счёт компьютерного моделирования психических процессов. И потому,
3. С этой точки зрения наиболее перспективной выглядит слабая версия репрезентационализма, в рамках которой понятие репрезентации схватывает информационные обмены в многослойных вычислительных процессах сложных когнитивных систем, но однако не является недостаточным для окончательного объяснения психической жизни.

#### 2.4.4 Вывод из раздела 2.4

Репрезентации — необходимый элемент вычислительных процессов и, в качестве такового, имеют место в когнитивных актах. Однако именно поэтому **понятие репрезентации не выражает видовой специфики когнитивного.**

## 2.5 Дискуссии о биологической или вычислительной природе мозга

### 2.5.1 Школа Коштыянца и «биологический мозг»

Со стороны как некоторых философов, так и психологов, нейрофизиологов и вообще специалистов в области конкретных наук можно услышать следующее возражение против нейросетевой концепции когнитивных способностей: мозг — значительно более сложное устройство, чем любая самая изощрённая нейросеть. Во-первых, его архитектура не предполагает слоёв, по которым последовательно проходит информация в искусственных нейронных сетях (ИНС); во-вторых, биологический нейрон может передавать информацию только в одном направлении, тогда как для большинства ИНС критически важно т. н. «обратное распространение ошибки». Но, что даже более важно, в-третьих, мозг — это вообще не электрическое, а биохимическое устройство, что означает два принципиальных обстоятельства: (1) главными движущими силами мозговой деятельности являются не электрические импульсы (спайки), а нейротрансмиттеры, и (2) по крайней мере значительная часть когнитивной деятельности мозга совершается не в нейронных сетях или ансамблях, вообще — не в пространстве между нейронами, а внутри отдельных нейронов, где взаимодействуют нейрхимическая и генетическая «машины». В общем и целом, мозг устроен гораздо сложнее искусственных нейронных сетей, электрические процессы там управляются химическими (выработкой нейромедиаторов) и генетическими (экспрессией генов), в то время как коннекционистские модели эмулируют только электрические взаимодействия. «Антиэлектрическая» позиция некоторых психологов и нейрофизиологов часто обозначается кратким названием «биологический мозг».

В этих аргументах нетрудно увидеть, если так можно сказать, нейрофизиологическую детализацию старой философской полемики Дж. Сёрла против Д. Деннета (см. её обзор в [70, с. 31–32]: тезис о биологической каузальности сознания против функционалистского представления о «множественных реализациях»).

Аргументация Сёрла и других сторонников биологического натурализма в философии, равно как и сторонников «биологического мозга» в нейрофизиологии, имеет следующую логическую структуру:

1. Функции, зависящие от свойств материального субстрата их носителей, не доступны для моделирования (т.е. воспроизведения в других материальных средах).
2.  $X$  зависит от свойств субстрата.
3. Следовательно,  $X$  не доступен для моделирования.

Если под  $X$  понимаются когнитивные функции, в общем случае — сознание, а под субстратом — биохимически определённые сущности (нейротрансмиттеры, биология и генетика нейрона и т.п.), то отсюда заключают к невозможности «сильного искусственного интеллекта» (Сёрл) или к нерелевантности искусственных нейросетей в качестве моделей психической деятельности и когнитивных способностей (адепты «биологического мозга»).

Но, как и всякий классический силлогизм, данное рассуждение уязвимо для ошибки, известной как «удвоение термина». Ошибка имеет место, когда средний термин — в данном случае это «быть зависимым от свойств субстрата» — понимается по-разному в каждой из посылок. Далее я попытаюсь показать, что эта ошибка весьма вероятна и в нашем случае. Я предлагаю различать по крайней мере два типа субстратной зависимости: сильную, т.е. зависимость от материала, и слабую, т.е. зависимость от регулятора. Например, пищеварение является собою очевидный пример сильной зависимости: только определённые белки, жиры и углеводы плюс определённые ферменты могут привести к желаемым химическим реакциям. А когда мы строим деревянный дом, то можно дискутировать об уникальности дерева как строительного материала, но, например, утеплители уже ближе к слабой зависимости по своей функциональной роли, поскольку со сравнимыми результатами можно использовать минеральную вату, стекловату, пеноплекс и др. Безусловно, физико-химические свойства этих материалов важны для достижения цели, но эта зависимость носит более абстрактный характер, что обеспечивает широкий набор альтернативных решений. То же можно сказать о радиолампах, полупроводниковых диодах и транзисторах как о средствах выпрямления тока в радиотехнике — используются совершенно разные вещества с разными химическими и физическими свойствами, но их сочетания дают один и тот же эффект.

Поэтому для окончательного обоснования притязаний «биологического мозга» его сторонникам следовало бы показать что в этом отношении химия мозга сродни пищеварению, и мы имеем дело с очевидной сильной зависимостью. При этом важно было бы исключить сомнения в том, что если мозг для достижения определённого эффекта всегда использует один и тот же нейротрансмиттер, это происходит именно как следствие сильной зависимости, а не потому, что организм просто не умеет производить аналоги. Хотя уже сейчас из популярных публикаций можно заключить, что морфий является функциональным аналогом эндорфина, а алкоголь — функциональным аналогом ацетилхолина. Но, как бы то ни было, пока нам ничто не мешает встать на альтернативную точку зрения, согласно которой управление возбуждением и подавлением нейронов, равно как и утепление домов, — это функции, с которыми могут справиться химически различные вещества в физически различных конфигурациях. И, конечно же, чем абстрактнее свойства веществ и вещей, важные для их функциональных взаимодействий, тем легче моделировать эффект этих свойств другими материальными средствами.

Однако попытаемся проследить линию аргументации в пользу «биологического мозга». Д. А. Сахаров пишет: «Формирование новой концептуальной схемы возможно <...> на основе идеи, что механизмы нервной регуляции унаследованы от донервных регуляторных систем и имеют химическую природу» [227, с. 334]. Эта идея восходит к “энзимохимической гипотезе” советского физиолога Х.С. Коштоянца, которая «априорно наделяла сигнальные молекулы нервной системы биохимической индивидуальностью и предлагала выводить своеобразие каждого нейромедиатора из его донервной регуляторной функции» [227, с. 334]. Иными словами, для управленческой деятельности мозга принципиально важны не электрические, а химические сигналы, которыми обмениваются его клетки, и эта химическая машинерия связывает появление мозга с более ранними этапами эволюции. В свою очередь, А.С. Базян считает, что кодирование информации в мозге происходит не на уровне нейронных сетей, а на уровне внутриклеточных реакций. Нейросети не смогут имитировать сложные интеллектуальные функции. Приведу объёмный отрывок, детально выражающий содержание его подхода:

Суть молекулярно-химического кодирования сводится к тому, что модуляторные рецепторы запускают внутриклеточные процессы

фосфорилирования и, управляя синаптическими процессами и порогом возбудимости нейронов, формируют пространственно-временную топологию нейронной сети и, тем самым, индуцируют специфическое для эмоционального или мотивационного состояния поведение. Вторым этапом процессы фосфорилирования включают трансдукционный сигнал, который модифицирует экспрессию генов и сохраняет в долговременной памяти процессы первого этапа, т. е. фактически пространственно-временную топологию нейронной сети и специфическое поведение. Эмоциональное или мотивационное состояние представляет собой сложную высокоорганизованную систему и включает много структур с их нейронами. Для каждого типа нейронов высокоорганизованной системы вырабатывается свой молекулярно-химический код, который является клеточным аналогом эмоционального состояния. Это происходит вследствие того, что каждый нейрон представляет собой высокоорганизованную молекулярно-химическую систему. Эмоциональное или мотивационное состояние консолидируется каждым нейроном через модификацию экспрессии его генов. Сложная высокоорганизованная система консолидируется в долговременной памяти каждой точкой [228, с. 168–169].

Из приведённого отрывка следует, что существуют механизмы долговременной памяти, которые работают на уровне отдельных нейронов и действуют через химическое управление экспрессией генов. Далее, нейроны подразделяются на типы, каждому из которых соответствует определённый молекулярно-химический код. И, наконец, сложные функциональные химические структуры, включающие нейромедиаторы, рецепторы и фосфорилирование (т.е. активацию/деактивацию) белков, характерны тем, что «формируют пространственно-временную топологию нейронной сети». В контексте дискуссии о релевантности нейросетей мы должны признать, что устройство мозга неизмеримо сложнее. Однако, на этом уровне изложения неизбежно возникают следующие вопросы: Какая именно информация сохраняется в химико-генетической памяти отдельного нейрона? Что мы узнаем, если сможем её расшифровать?

Если вся сложнейшая химия предназначена для оперативного изменения конфигурации сети нейронов (подобная мысль содержится и у Д.А. Сахарова),



то для чего создаётся сама сеть? И для чего вообще нужны электрические импульсы, испускаемые нейронами, если всё делает химия?

С инженерной точки зрения способность сети к оперативной реконфигурации — существенное преимущество, пока, насколько мне известно, не реализованное в ИНС, которые — по крайней мере, в промышленном исполнении — создаются «под задачу» и не могут взять на себя универсальное управление. Но, с другой стороны, подумаем об биологических эволюционных механизмах вообще. Мы, люди, объединены в единую информационную сеть, называемую обществом. И эволюция распорядилась так, что мы вынуждены обмениваться сигналами, преобразуя сложную химико-генетико-электрическую машинерию мозга в движения гортани и языка, которые, в свою очередь, преобразуются в акустические колебания, которые далее улавливаются ушной мембраной и трансформируются в электрические (всё-таки!) нервные сигналы, которые обратно декодируются мозгом реципиента, чтобы быть «услышанными» как осмысленное сообщение. Мягко говоря, отнюдь не оптимальное техническое решение. Но природа смогла предложить нам только его, именно потому что действует вслепую, методом проб и ошибок, и, в отличие от обычного инженера, её никто не может уволить из-за отсутствия конкуренции на этом рынке труда.

Не было ли бы естественно предположить по аналогии, что и для относительно несложной задачи оперативной реконфигурации сети «на лету» и наделения памятью каждого отдельного нейрона природа подготовила слишком сложный и не очень эффективный химический «механизм», просто потому он использовался и на донейронных стадиях эволюции, как справедливо указывают сторонники «биологического мозга»? Подобно тому, как многие из нас для некоторых домашних технических работ используют не то, за чем нужно долго ехать в магазин, а то, что уже под рукой, и что, пусть хуже, но поможет быстро достичь желаемого. Если моё предположение имеет смысл, то возможно также, что концепция «биологического мозга» увязает в сложных, но малосущественных деталях, настаивая на том, что в них и таится суть когнитивных феноменов. Вообще, когда говорят, что ИНС не объясняют деятельности мозга, потому что мозг не электрическая, а химическая машина, фокус дискуссии смещается к субстанциальной определённости: сети — это сфера прохождения и маршрутизации электрических сигналов, а нейромедиаторы — это сфера химических реакций. С феноменальной точки зрения разница принципиальна.

Но она производит впечатление только до тех пор, пока мы отождествляем сложные когнитивно-нагруженные сети, каковой, безусловно, является и мозг, с конкретной их технической реализацией. Нет ничего невозможного в том, чтобы и электрические импульсы, и химические реакции выступали в качестве транспорта для информационно-вычислительных процессов. Но только нужно оговорить два обстоятельства.

1. Из сегодняшнего горизонта знаний наиболее перспективным путём поиска решений когнитивных проблем представляются теории, в существенной степени опирающиеся на ту или иную концепцию вычислений<sup>29</sup>. Вычисления — алгоритмические процессы, связанные с передачей и обработкой информации. Для них существенна *субстратонезависимость* — один и тот же вычислительный алгоритм может быть реализован в разных средах; и *алгоритмичность* — результат достигается путём последовательного или параллельного применения набора элементарных процедур (вычислительных примитивов), правила применения которых и составляют алгоритм.

Заметим, что вычисления определяются здесь через дискуссионное в определённой степени понятие информации. Цели настоящего исследования не предполагают участие в дискуссии по поводу удовлетворительного определения этого понятия. Для этих целей следующее уточнение было бы достаточным:

(Df5) Информация — определённая структура, связанная с некоторой другой структурой устойчивой инвариантной функцией.

Так, например, физический знак «слово» содержит информацию, поскольку если я изменю его структуру так, чтобы получить «волос», последуют предсказуемые изменения в структуре понимания и действий моего читателя. Причём в общем случае сам читатель вовсе не обязателен для того, чтобы некоторая структура оказалась информационно нагруженной. Напротив, антропоморфизация этого понятия играет с ним злую шутку и препятствует его операционализации в контексте наук, для которых оно важно. В конце концов, изменение последовательности нуклеотидов РНК также окажет предсказуемое влияние на синтез протеинов, хотя никакой семантики в этом процессе не усматривается, несмотря на метафорическое употребление слов «буква» и «строчка» в контексте генетики.

---

<sup>29</sup>Проблема вычислений и вычислительных подходов в когнитивной науке подробно рассмотрена в разделе 2.3.

Здесь возможно возражение: ведь пуля, пробивающая стекло, тоже делает это в силу своих структурных особенностей, поскольку её удельный вес и плотность, помогающие ей одержать победу в столкновении с менее прочным материалом, объясняются её молекулярной и атомной структурой. И след от пули на стекле также поддаётся структурному описанию на разных уровнях организации. Что мешает представить взаимодействие пули и стекла как информационный процесс?

Столкновение свинца и стекла возможно описать как линейное причинное взаимодействие, и для такого описания вполне достаточно средств классического естествознания: механики, термодинамики, химии и т.п. Конечно, все знают, что и пуля, и стекло представляют собой сложные структуры с огромным количеством элементов. Но и структурные свойства свинца, существенные для пробивания стекла, и такие же свойства стекла, обеспечивающие образование пулевого отверстия, интегрированы в простые величины: скорость, кинетическая энергия, плотность, температура и т.д. Для получения ожидаемого эффекта не важны такие специфические свойства их структур, как топология элементов и связей между ними, пути прохождения взаимодействий между отдельными элементами — мы отвлекаемся от всего этого, заменяя статистические или системные величины однозначными динамическими, взаимозависимость которых описывается простыми линейными уравнениями, и вся картина происходящего выглядит поэтому достаточно «крупнозернистой». Напротив, то, что мы называем информацией, имеет место, когда фиксируется инвариантная функциональная зависимость именно между пространственно-временными и/или иными структурными деталями взаимодействующих систем.

2. Есть основания утверждать, что когнитивные способности являются эмерджентными<sup>30</sup> эффектами сетевой структуры, обеспечивающей параллельные вычисления. Почему это важно?

Живые организмы, у которых только мы изначально и обнаруживаем когнитивные способности, являются продуктами эволюции — альтернативная точка зрения происходит скорее из сферы идеологии, а не науки. Это значит, что методом строительства всё более сложных адаптивных систем выступает не разумное проектирование, а слепые пробы и ошибки. Более того, само появление мозга как основного носителя когнитивных свойств означало революцию в эффективности биологических адаптивных систем, поскольку обучаемые

---

<sup>30</sup>См. определение на с. 306.

организмы меньше гибнут в результате ошибок, и, соответственно, биоматериал экономится, а скорость эволюции увеличивается экспоненциально. Всё это вместе говорит о возрастающей эффективности (в т.ч. энергетической) вычислительной системы.

Человек-инженер, работая над созданием вычислительной системы и увеличением её мощности, поступает как индивид, уже тренированный в рамках социальной коммуникации: он берёт машину Тьюринга и заставляет её совершать всё больше операций в единицу времени, увеличивая частоту и добавляя ядра процессора — всё ради того, чтобы быстрее и лучше обрабатывать последовательные цепочки символов, как, собственно, и поступаем все мы в рамках нашей социальной коммуникации — в точности как научила нас природа, о чём говорилось выше. Собственно и метафора «вычислений» заимствуется учёными из этой же сферы. И, подобно тому как принципиально не менялась схема автомобиля с 1910-х, несмотря на фантастически возросшие мощность и комфорт современных экземпляров, так и современные искусственные вычислительные системы — компьютеры — всё также получают данные от устройств ввода, обрабатывают их в соответствии с загруженными алгоритмами и сохраняют их в ячейках памяти. Все инженерные ухищрения до сих пор были направлены на ускорение процессов, происходящих в той же самой схеме: ввод — процессинг<sup>31</sup> — запись в память — вывод.

Если бы природа исполняла некий инженерный план, то, во-первых, принципиальная схема такого устройства уже должна была быть реализована в нашем когнитивном аппарате. А во-вторых, алгоритмы обработки тоже должны были быть заранее кем-то созданы и загружены в него же. Но ничего из этого природа позволить себе не может. Слепая эволюция может лишь, однажды случайно создав примитивный и маломощный вычислительный процессор, добиваться увеличения его вычислительной мощности, а значит и адаптивной эффективности, просто увеличивая их количество в некоторой связанной группе, подобно тому как в некоторых архаичных языках множественное число образуется через удвоение слова. И на этом пути природу ожидает одно замечательное открытие, до которого не додумался человек разумный: совокупная вычислительная мощность сложносоставного вычислительного устройства не равна сумме таковых мощностей примитивных процессоров — возникает эмер-

<sup>31</sup>См. определение на с. 306.

джентный<sup>32</sup> эффект за счёт параллельной обработки информации, и те данные, которые даже не могут быть представлены (репрезентированы) отдельному процессору, совокупная сеть отражает, обрабатывает и даже запоминает с готовностью. Не случайно современные неклассические подходы в теории и практике вычислений, основанные на параллельном процессинге — искусственные нейронные сети, мультиагентные системы — называют в литературе ‘bio-inspired’ — вдохновлённые биологией.

Есть достаточно серьёзные доводы в пользу предположения, что когнитивные способности во всём их многообразии суть эмерджентные эффекты именно сетевой архитектуры биологических вычислительных систем, осуществляющих параллельный процессинг. Именно распределенные процессоры, работающие как статистические машины, способны к самообучению на основе категоризации данных, порождения вероятностных гипотез и уточнения их в результате обработки ошибок. Поэтому дело не в качественной определённости процессов, обеспечивающих эти способности — электричество или химия, — дело в архитектуре вычислительной системы, способной самостоятельно, без внешнего программирования работать с вероятностными распределениями данных.

Здесь мы можем столкнуться ещё с одним возражением: аргумент против сетевой основы когниций<sup>33</sup> состоит даже не в субстрате мозговых процессов, а в уровне, на котором следует ожидать проявления когнитивных способностей, в их элементарном носителе — сторонники «биологического мозга» полагают, что таковым может выступить отдельный нейрон, который не обладает сетевой структурой, а электрический потенциал, транслируемый им вовне образуется вследствие некоторых химических и генетических событий. Адепты этой точки зрения ссылаются на т. н. «нейрон бабушки» — эффект, когда целостный когнитивный образ появляется как следствие возбуждения одного единственного нейрона.

Философ не имеет ни средств, ни морального права критиковать научные теории, основанные на эмпирических данных и экспериментах: его поле — концептуальные схемы. Но нельзя не отметить, что даже специалисты в области поведенческой специализации отдельных нейронов подчёркивают, что, при всей сложности последних как био-химико-генетических устройств и информационных процессоров, информация, которая доступна каждому из них в

<sup>32</sup>См. определение на с. 306.

<sup>33</sup>См. определение на с. 306.

отдельности, располагается на, если так можно сказать, «служебных» уровнях структурной иерархии мозга, тогда как собственно когнитивная обработка информации осуществляется на более высоких уровнях организации (см., напр., [229—231]).

С другой стороны, имеются результаты исследований, которые напрямую связывают степень развитости когнитивных функций со степенью связности нейронной сети, измеряемой средствами теории графов. Так, в недавнем исследовании нейронных основ когнитивных и интеллектуальных различий между людьми, опирающемся на солидный эмпирический базис и серьёзные математические средства анализа, было установлено, что общая интеллектуальность (intelligence) испытуемых зависит «от между- и внутримодульной связности в нейронных кластерах, расположенных в лобных, теменных и других кортикальных и подкорковых областях мозга, которые ранее рассматривались в качестве локализованных нейронных субстратов интеллекта» [111, р. 6].

С другой стороны, успехи моделирования многих когнитивных функций дают некоторые основания для правдоподобных предположений, что именно такой механизм лежит в основе естественных когнитивных процессов. При этом химические и генетические механизмы живого мозга играют существенную роль в его общем функционировании — но не благодаря их субстратной определённости, а благодаря их функциональной роли, которая, скорее всего, доступна моделированию в иных субстратах.. Это значит, что для искусственного воспроизводства всей сложности реального когнитивного аппарата не обязательно использовать те же химические вещества или цепочки ДНК. Достаточно математически описать их функциональные роли и воспроизвести эти роли в компьютерных моделях.

Но окончательные ответы, конечно же, могут дать только научные исследования, опирающиеся на солидный эмпирический базис и корректное моделирование<sup>34</sup>.

---

<sup>34</sup>Буквально на последнем этапе подготовки текста диссертации пришла новость о том, что учёные БФУ им. Канта создали первый в мире химический нейрокомпьютер [232]. Это устройство спроектировано как нейросеть, но носителем информации выступают не электрические импульсы, а химические реакции. Создатели признают, что как компьютер оно получилось медленным. Однако сам факт подтверждает тезис о многообразной реализуемости: вычислительные алгоритмы могут исполняться в различных материальных средах.

## 2.5.2 Когниции как эмерджентные эффекты сетевых вычислительных архитектур

Как показывает опыт изучения специальной литературы, эмпирические данные, на которые опирается концепция «биологического мозга», неплохо интегрируются в альтернативные концепции нейросетевых вычислений. Так, несколько лет назад группа авторов [171] поставила перед собой амбициозную задачу: проанализировать вычислительную структуру мозга с целью обнаружения элементарных вычислительных операций — «вычислительных примитивов» — и их нейродинамических коррелятов. Этот теоретический ход интересен тем, что, в то время как нейросетевые модели часто называют «биологически вдохновлёнными», поскольку их вычислительная архитектура была навеяна знаниями об устройстве живого мозга, предложенный здесь противоположный подход можно было бы охарактеризовать как «компьютеровдохновлённый», поскольку он стремится увидеть в биологическом материале элементы, характерные для искусственных процессоров.

Авторы исходят из того, что при ближайшем рассмотрении в мозге имеются сотни различных типов нейронов, а отдельные синапсы содержат сотни различных протеинов. Это наводит их на мысль, что кору мозга следовало бы понимать не как единое устройство, а как широкий набор типовых вычислительных примитивов — элементарных единиц процессинга<sup>35</sup>, подобных множествам базовых инструкций в микропроцессоре — возможно, соединённых параллельно, как и в том реконфигурируемом типе интегральной схемы, который известен как программируемая пользователем вентиляционная матрица. В качестве возможных кандидатов на роль вычислительных примитивов авторы предлагают схемы для смещения фокуса внимания, для кодирования и манипуляции последовательностями, для нормализации отношения активности отдельного нейрона к активности множества нейронов. Сюда же могли бы быть отнесены схемы-гейты, переключающие потоки информации между различными отделами коры, а также устройства хранения рабочей памяти и принятия решений. Они могут также включать в себя схемы для хранения рабочей памяти, принятия решений, хранения и преобразования информации посредством

---

<sup>35</sup>См. определение на с. 306.

распределённого нейронного кодирования (population coding) и кодирования переменных.

Но особого внимания заслуживает такое элементарное вычислительное действие как временное или постоянное связывание двух битов информации: переменной и её произвольного значения. Такое связывание играет центральную роль как в понимании (в т. ч. продуктивном) языка, так и в дедуктивном рассуждении. В качестве кандидатов на нейронные механизмы такого связывания рассматриваются: временная синхрония нейронных ансамблей, мультипликация векторов, кодируемых нейронными популяциями, точно контролируемые взаимодействия между префронтальной корой и базальными ганглиями. Возможные механизмы включают также взаимосвязанные системы анатомически определённых регистров (групп нейронов, определяющих временные хранилища памяти) с различными схемами кодирования. Эмпирическому исследованию в этой области помогают такие новые технологии, как оптогенетика. Адекватное описание механизма связывания переменных, по мнению авторов, может оказаться незаменимым для установления связей между нейронами и когнитивными процессами более высокого уровня.

Например, наблюдаются систематические различия в экспрессии генов между отделами коры, которые увеличиваются как функция расстояния между ними. Другие молекулярные механизмы, такие как альтернативное склеивание нейрексинов (протеинов, которые помогают организовать формирование нейрональных синаптических связей), обеспечивают потенциальные пути, с помощью которых кажущиеся на первый взгляд тонкими молекулярные различия могут приводить к важным качественным изменениям синаптической связи.

Нейронаука, уверены авторы, должна развивать такие экспериментальные инструменты как подробные карты мозга и вычислительные инфраструктуры, на поддержку которых нацелены сегодняшние инициативы в исследовании мозга, но также и новый набор научных методов для понимания того, как, хотя бы в общем виде, системы могут переходить от сетей нейронов к символическому познанию. С этой целью междисциплинарное исследование возможности построения таксономии и филогенеза представленных в коре вычислительных примитивов продвинуло бы наше понимание далее к конечной цели — к расшифровке того, как комплексы таких элементов определяют поведение.

Авторы более раннего исследования [233] сосредоточились на межклеточных сигнальных сетях, которые координируют деятельность отдельных



клеток мозга. Многочисленные сигнальные пути, согласно их данным, позволяют клетке принимать, обрабатывать и реагировать на информацию. Часто компоненты различных путей взаимодействуют, что приводит к образованию сигнальных сетей. Была построена компьютерная модель *биохимических сигнальных сетей* в соответствии с экспериментально полученными и опубликованными константами и проанализирована с помощью вычислительных методов, чтобы понять роль этих сетей в сложных биологических процессах. Как оказалось, сигнальные сети демонстрируют эмерджентные<sup>36</sup> свойства, такие как интеграция сигналов в разные временные масштабы, генерирование явно различимых выводов в зависимости от мощности и продолжительности ввода и самоподдерживающиеся контуры обратной связи. Обратная связь в этих сетях может привести к бистабильному поведению с дискретными устойчивыми действиями, четко определенными порогами ввода для перехода между состояниями и длительным выходом сигнала и модуляции сигнала в ответ на переходные (transient) раздражители. Эти свойства сетей сигнализации повышают вероятность того, что информация для «выученного (learnt) поведения» биологических систем может храниться во *внутриклеточных биохимических реакциях*, которые включают в себя сигнальные пути.

В этом изложении я выделил курсивом ссылки на те факты, которыми оперируют сторонники «биологического мозга», но которые, как мы видим, прекрасно ассимилируются современными нейровычислительными теориями.

\* \* \*

Возвращаясь к силлогизму, сформулированному в параграфе 2.5.1, отметим, что удвоение термина всё же имеет место: когнитивные способности демонстрируют слабую зависимость от нейротрансмиттеров, медиаторов и других химических субстанций, тогда как в первой посылке, конечно, предполагается сильная зависимость.

Многочисленные вычислительные модели, учитывающие химический аспект деятельности мозга, показывают, что когнитивные акты всё же существенно отличаются от пищеварения — именно тем, что представляют собой вычислительные процессы. Эмпирические исследования поставляют многочисленные данные, которые могут быть интерпретированы в пользу любой из конкурирующих концепций. Это означает лишь, что конкуренция, как и везде в

---

<sup>36</sup>См. определение на с. 306.

научной сфере, должна основываться на достижении преимуществ в двух основных параметрах научной теории: объяснительной силе и простоте онтологии. Что касается концепции «биологического мозга», то необходимо заметить, что задачей философа не может быть установление её истинности или ложности — это дело профессионалов. Я лишь ставил перед собой цель подвергнуть анализу притязания этой теории на роль единственно истинного объяснения когнитивных эффектов, порождаемых мозгом, и, соответственно, отрицание истинности и релевантности альтернативного — нейросетевого — подхода. Как мы увидели, свидетельства, приводимые биологическими натуралистами, недостаточны для окончательного выбора в пользу их интерпретации и в принципе неплохо уживаются и с нейросетевым видением. В то же время, реальные технологические успехи искусственного интеллекта в наши дни связаны именно с нейросетевыми моделями когнитивных способностей. Все эти обстоятельства позволяют заключить, что адептам «биологического мозга» пока не удаётся ни доказать истинность собственной теории, ни опровергнуть конкурирующие подходы.

### 2.5.3 Вывод из раздела 2.5

Концепция «биологического мозга» представляется слишком сильной абстракцией, для которой не просматривается достаточно оснований. Для настоящего исследования важно, что **аргумент «биологического мозга» не достаточен для дискредитации вычислительного подхода в когнитивной нейронауке.**

## 2.6 Вывод из главы 2

Психология, исторически преодолевшая качественно-субстанциалистские основания в пользу количественно-функционального подхода, естественным образом приняла «вычислительную метафору». Трудности с адаптацией этой метафоры свидетельствуют о необходимости нового понимания вычисления. Результатом этого этапа исследования является **онтологическая концепция**

вычисления как потенциально многоуровневого процесса, предполагающего представленность одних процессов в других с помощью некоторых репрезентационных механизмов. В свою очередь, репрезентации могут пониматься как необходимый элемент многоуровневых вычислительных процессов в сложных, в том числе когнитивных, системах.

### Глава 3. От когнитивного к социальному

Традиционно материальный субстрат когнитивных способностей отождествляется с мозгом. Как показано в предыдущей главе, энтузиасты когнитивной революции искали и находили в компьютере аналогии, способные объяснить его функционирование. Позже адепты «воплощённого познания» и энактивизма расширили сферу когнитивного до сенсомоторных систем — фактически, до анатомических границ организма, — правда, зачастую за счёт отказа от вычислительных метафор. Позднейшие «экологические» расширения когнитивного проекта, часто смыкаясь с методологическими принципами «предиктивного процессинга» (см. с. 145), ещё более раздвинули границы, включив в поле реализации когнитивных функций непосредственное окружение организма, наполненное аффордансами (см. с. 306) [57; 220; 234; 235]. Логично предположить, что и социальное окружение может быть исследовано как ещё одно расширение материальных основ когнитивных функций. Определение философских и методологических основ такого исследования составляет содержание настоящей главы.

#### 3.1 Социальная онтология

##### 3.1.1 Номиналистический взгляд на общество

Социальная онтология, как и всякая другая, должна ответить на вопрос: что является «первыми сущностями» — в терминологии Аристотеля, — которые не нуждаются в определении чем-либо иным или в отношении к чему-либо иному для своего существования; и которые отличны от «вторых» сущностей, которые существуют как эффекты комбинаций и взаимодействий первых сущностей. В области социального знания, где предмет исследования — общество — мыслится как состоящее из индивидов, мы находим два принципиально различных ответа на вопрос о существовании: (а) существуют индивиды, а социальные связи представляются или мыслятся ими, и (б) существуют социальные связи

как устойчивые формы взаимодействия индивидов, и они не сводимы к представлениям и мыслям индивидов о них.

Эту проблему можно переформулировать так: всё ли то существует, что является значением терминов, входящих в истинное высказывание? Возьмём для примера непосредственно истинное высказывание: «В этой комнате есть столы». Согласно подходу к его интерпретации, который можно было бы назвать реалистическим, понятию «стол» соответствует сущность «стол как таковой», отнесение к которой делает этот и тот объект столами. Данная сущность существует как комплекс общих признаков своих экземпляров и в своём абстрагированном – очищенном от индивидуальных различий – виде представляет собой один и тот же объект.

Реалистическому подходу противостоит подход номиналистический: существуют этот и тот объекты, которые могут обладать сходными признаками. Здесь экзистенциальное высказывание «есть такие  $x$ , что...» относится к неопределённому множеству объектов, идентифицировать которые можно только индексикалами («тот», «этот»,  $a_n$  и т.п.). Эти объекты могут случайным образом обладать набором признаков, некоторые из которых оказываются общими определённым подмножествам объектов, или воспринимаются как общие. Такие пересечения подмножеств объектов, обладающих общими признаками, соответствуют осмысленным истинным высказываниям: «Существуют объекты, которые, обладая всеми признаками стола (согласно общепринятому перечню), находятся в этой комнате».

Большинство социальных теорий привычно опираются на реалистические онтологии — реалистические в понимании, известном из истории философии: признающие существование универсалий и полагающие их реальными значениями теоретических терминов. Реализм в общественных науках имеет даже более прочные эпистемологические корни, чем в науках естественных, поскольку социальное знание, вырастая из повседневных квази-теорий, наследует их словарь, наполненный именами абстрактных сущностей: «солидарность», «справедливость», «правительство» и т. п. Альтернативный — номиналистический<sup>1</sup> — взгляд, признающий существующими только партикулярии, обозначаемые именами собственными или индексикалами, имеет меньше шан-

<sup>1</sup>Дискуссии вокруг современного или, как его иногда называют, неоминализма в метафизике и философии математики имеют уже продолжительную историю — по крайней мере, с начала прошлого века: см. [72; 236–244]

сов на широкое признание, поскольку — в сфере собственно общественных наук — такой подход вынужден будет отказаться от привычного социального словаря, а в масштабе научного знания вообще, как утверждает Гидеон Розен, «любой довод в пользу обязывающего номинализма неизбежно будет скептическим доводом: доводом в пользу пересмотра неизменной (и, по привычным стандартам, беспроблемной) позиции здравого смысла и устоявшейся науки» [72, р. 71]. Однако далее я попытаюсь показать, что возможно примирить номиналистскую установку с привычным словарём науки, если правильно провести различие между универсалиями и партикуляриями.

В первом приближении, и согласно расхожей точке зрения, социальный номинализм должен состоять в убеждении, что в этой сфере не существует ничего, кроме биологически определённых индивидов, наделённых психикой. То же, что мы называем обществом, государством, политическим режимом, экономической системой и т.п., существует только как идеи в их головах. Однако такая точка зрения уязвима в двух отношениях: онтологическом и эпистемологическом.

Первая уязвимость состоит в том, что не определён критерий тождества для социальных идей. Если один гражданин некоторого государства считает его в высшей степени справедливым и соответствующим народным чаяниям, а другой называет его «кровавым режимом», то очевидно, что идеи государства в их головах не являются тождественными, и тогда — что, они живут в разных государствах?

Вторая уязвимость заключается в том, что в этом случае социальные теории не могут быть признаны объективно истинными или ложными. Невозможна, например, ситуация, когда большинство граждан ошибочно приписывают какое-либо свойство своей политической системе (скажем, свойство быть демократией), поскольку социальная реальность по определению такова, каковой они её воспринимают.

Критическая интенция социального номинализма должна состоять в том, чтобы дезавуировать некоторые расхожие псевдосущности, не подлежащие эмпирической проверке. Методологическая интенция социального номинализма должна состоять в том, чтобы предложить эффективно операционализируемую онтологию. Онтология наивного номинализма, описанная чуть выше, не удовлетворяет ни одной из этих интенций, поскольку, с одной стороны, любая псевдосущность вполне может рассматриваться как комплекс идей в головах

большой массы индивидов, и тогда она не более, но и не менее реальна, чем то же государство. С другой стороны, индивиды, обладающие идеями, обладают также полным набором интенциональных свойств, что ставит их над естественной причиносообразностью и наделяет «свободой воли», «активностью сознания» и другими философскими конструктами. А это исключает использование каких-либо эффективных формализмов в социальной теории, поскольку любой сбой в объяснительной функции теории можно объяснить «свободой» и «активностью» её объектов.

Корректно построенная социальная онтология должна рассматривать социальных индивидов как простые объекты, обладающие только однозначно определяемыми и количественно измеряемыми свойствами, а также способные вступать в отношения друг с другом, которые — отношения — также должны быть количественно измеряемы. Классическая экономическая теория в этом отношении имеет тот недостаток, что помимо простейших однородных элементов социальной реальности предполагает дополнительные сущности — товары, деньги и т.п. — обрекая себя тем самым на роль *нишевой теории*. Общая социальная теория не должна иметь никакой предметной определённости, чтобы быть интерпретируемой на любой предметной области.

Правильным путём конструирования социальной теории был бы следующий: (a) создание минималистической онтологии, удовлетворяющей «бритве Оккама», (b) формулировка основных принципов (ограничений), законов (объяснительных принципов) и/или алгоритмов (правил преобразования), (c) обоснование наличной устойчивой структурности изучаемой реальности, (d) демонстрация выводимости наблюдаемых фактов из того, что содержится в (b).

В контексте декларируемого здесь онтологического номинализма неоднозначной выглядит проблема объективного существования социальных систем (c) вне человеческих представлений. Существует то, что может быть самостоятельным звеном в причинно-следственных цепочках. Очевидно, что это относится к социальным структурам (институтам). То обстоятельство, что они не могут существовать без поддержки когнитивных агентов, не отменяет их онтологического статуса как объективно существующих, эмпирически обнаруживаемых и участвующих в причинно-следственных связях функциональных комплексов — инвариантов поведения тех же агентов. Но если агенты мыслятся как онтологические партикулярии, то признание существования чего бы то ни было наряду с ними может быть истолковано как отход от номинализма

или, по крайней мере, как непоследовательность. Однако это подозрение легко преодолевается средствами уточняющего анализа.

Существуют факты — устойчивые положения (отношения) вещей ( $c$ ), — а не субстанции (подробнее см. параграф 1.2.3). Положения вещей можно представить как сколь угодно сложные комплексы отношений элементарных объектов. Несмотря на то, что описания некоторых таких комплексов могут быть полностью идентичны, тем не менее, соответствующие им факты не абстрактны, но конкретны — соответственно, являются не универсалиями, а партикуляриями — поскольку составляющие их объекты фундаментально различны, т. е. занимают различные места в логическом пространстве. Так, если определено некоторое отношение  $R$  на множестве  $A$ , то комплексы  $a_m R a_n$  и  $a_o R a_p$  — суть различные экземпляры типа, заданного этим отношением. При этом имя типа семантически определяется как «синкатегорематическое: значимое в контексте, но ничего не обозначающее» [236, р. 105]. Так, если вернуться к социальной реальности, термин «государство» не обозначает ничего реального, в отличие от терминов «Соединённые Штаты Америки» или «Республика Беларусь», которые обозначают конкретные экземплярификации типа «государство», т. е. социальные партикулярии. Такое решение хорошо тем, что, опираясь на максимально абстрактную онтологию, оно переносит задачи содержательного описания на собственно теорию. Это, в свою очередь, означает лучшую эмпирическую верифицируемость/фальсифицируемость теории, поскольку типы её объектов задаются не конвенционально, а как легко проверяемые образы фактов.

Может возникнуть вопрос: почему нельзя онтологически разъять сложные комплексы, чтобы теория имела дело с более простыми партикуляриями, и можно было бы, например, редуцировать социологию к психологии? Полноценная редукция возможна, только если теория низшего уровня может объяснить все феномены и факты высшего уровня. Если это не так, то признание зависимости высшего уровня от низшего, его «надстроенности» над феноменами низшего порядка, ничего не меняет в онтологической структуре мира (см. об этом [37; 245]).



### 3.1.2 Сеть как предельный случай социальной онтологии

Итак, хорошо работающая наука – это теория, основанная на простой эффективно формализуемой онтологии, на которой интерпретируется некий (логический или математический) формальный аппарат. Причём, если такая теория работает хорошо, это ещё не значит, что её онтология представляет собой «истинную» картину мира – это значит лишь, что мы – случайно или нет – подобрали удачную, работающую настройку нашей теоретической оптики. Но главное для философа состоит в том, что такая теория не нуждается в мистических или метафизических сущностях в качестве объяснительных инструментов.

Что касается социальных наук, их сциентизация сдерживается представлением об атомах социальной материи как об индивидах, наделённых свободой воли, то есть, способностью повести себя не так, как этого требует правило, каким бы оно ни было. Иными словами, социальная онтология в первом приближении не соответствует описанному выше идеалу: система отношений простых объектов, собственными, нереляционными свойствами которых можно пренебречь. Первой социальной наукой, которая продвинулась в этом направлении, оказалась классическая экономика (политэкономия), которая представила общественных индивидов в виде рациональных агентов, имеющих несколько простых базовых потребностей и способных к рациональной кооперации ради их удовлетворения. К этой онтологии оказались применимы математические методы, теория на её основе, как оказалось, обладала достаточной прогностической силой, хотя и не умела объяснить всех фактов.

С развитием методов компьютерного моделирования наступает новый этап в развитии социологии: социальные онтологии, приспособленные для решения проблем этой науки, теперь стало возможным моделировать программными средствами, что лучше и эффективнее эмпирических исследований в том случае, когда нужно ответить на общий вопрос – на правильном ли мы пути.

С реализацией описываемого нами идеала тем хуже обстоит дело, чем дальше мы продвигаемся в глубь территории гуманитарных наук.

Онтология, доступная для формализации, не торжествует пока в социально-гуманитарной сфере, поскольку эта сфера в основном находится в плену «народной психологии»: здесь очень часто, если не сказать всегда, придаётся онтологический статус тому, что обозначают психологические или интенци-

ональные предикаты: «знает», «сомневается», «боится», «любит» и т.п. И поскольку эти предикаты приписываются единому субъекту (в обоих значениях этого слова — логическом и эпистемологическом), возникает представление об их единой субстанциальной основе — «сознании». Поверхностный критик может предположить, что за этим аргументом должен последовать призыв к редукции: достаточно свести психологию к нейрофизиологии, а её — к электрохимии и т.д., и научное объяснение восторжествует. Однако, подчеркнём ещё раз, редукционизм сталкивается с эффектом нисходящей причинности: описав события, например, психологии в терминах нейрофизиологии, мы, возможно, не будем в состоянии воспроизвести причинную связь психических событий. Аргумент сторонников концепции нисходящей причинности сводится к тому, что более высокий уровень организации порождает собственные причинные связи, которые как бы «сверху» управляют событиями более низкого уровня, заставляя, например, нейрофизиологические и даже электрохимические процессы обслуживать психические функции [246—248].

Эта трудность означает лишь, что мы — если продолжить метафору — некорректно настроили нашу теоретическую оптику. Теория не должна предполагать, что её элементарные объекты «на самом деле» являются сложно-составными, даже если существует другая респектабельная теория, для которой они таковыми и являются. Корректно построенная теория должна описать свои объекты как простые сущности, наделённые немногими измеряемыми свойствами, а также способностью вступать в отношения, типы которых также немногочисленны и подробно описаны теорией. В предлагаемой мною социальной онтологии элементарные объекты не обладают никакими свойствами, кроме способности вступать в отношения, тип которых также единственен и максимально абстрактен. В рамках этой онтологии объекты становятся узлами, отношения — рёбрами, структура в целом — сетью.

\* \* \*

Корректно построенная социальная онтология должна рассматривать социальных индивидов как простые объекты, не обладающие нереляционными свойствами, если речь идёт об общей социальной теории, и обладающие только однозначно определяемыми и количественно измеряемыми свойствами, если речь идёт о локальных теориях, — а также способные вступать в отношения

друг с другом, которые — отношения — также должны быть количественно измеряемы. Правильным путём конструирования социальной теории был бы следующий: (1) создание минималистической онтологии, удовлетворяющей «бритве Оккама», (2) формулировка основных «законов», то есть устойчивых количественных функциональных зависимостей одних параметров от других, (3) вывод наличной устойчивой структурности изучаемой реальности из сформулированных законов, (4) демонстрация выводимости наблюдаемых фактов из законов в конъюнкции с другими, уже подтверждёнными фактами.

Существует то, что может быть самостоятельным звеном в причинно-следственных цепочках. Очевидно, что это относится к социальным структурам (институтам). То же, что они не могут существовать без поддержки когнитивных механизмов, не отменяет их онтологического статуса как объективно существующих, эмпирически обнаруживаемых и участвующих в причинно-следственных связях функций — инвариантов поведения социальных элементов. Необходимо также иметь в виду, что социальные и ментальные сущности существуют не как статические, а как динамические структуры, как инварианты событий и структурных взаимодействий. Их онтологический статус во многом связан с грамматикой языка объяснения. Давая имя некоему инварианту событий или социального поведения, мы как бы превращаем его в псевдоонтологическую сущность (имя существительное) — и вот у нас уже «социальные институты» рассматриваются как квазиобъекты — элементы социальной реальности. Правильно построенный язык описания социальной онтологии должен корректно различать объекты, свойства, отношения и функции.

Внедрение подобной эффективной онтологии в структуру социальных теорий сдерживается представлением об атомах социальной материи как об индивидах, наделённых свободой воли, то есть, способностью повести себя не так, как этого требует правило, каким бы оно ни было. Иными словами, социальная онтология в первом приближении не соответствует описанному выше идеалу: система отношений простых объектов, собственными, нереляционными свойствами которых можно пренебречь. Первой социальной наукой, которая продвинулась в этом направлении, оказалась классическая экономика (политэкономия), которая представила общественных индивидов в виде рациональных агентов, имеющих несколько простых базовых потребностей и способных к рациональной кооперации ради их удовлетворения. К этой онтологии оказались

применимы математические методы, теория на её основе, как оказалось, обладала достаточной прогностической силой, хотя и не умела объяснить всех фактов.

С развитием методов компьютерного моделирования наступает новый этап в развитии социологии: социальные онтологии, приспособленные для решения проблем этой науки, теперь стало возможным моделировать программными средствами, что лучше и эффективнее эмпирических исследований в том случае, когда нужно ответить на общий вопрос – на правильном ли мы пути.

Теория не должна предполагать, что её элементарные объекты «на самом деле» являются сложносоставными, даже если существует другая уважаемая теория, для которой они таковыми и являются. Корректно построенная теория должна описать свои объекты как простые сущности, наделённые немногими измеряемыми свойствами, а также способностью вступать в отношения, типы которых также немногочисленны и подробно описаны теорией.

В ходе исследования мы подошли к идее эволюционно развитой суперструктуры, которая сочетает в себе индивидуальные механизмы управления адаптивным поведением на основе обучения и индивидуальную подсистему хранения и обработки управляющих кодов, функционирующую на основе языка и письменности. Мозг индивида, включённого в эту суперструктуру, должен научиться обрабатывать не только данные, получаемые из природной среды, но и социальные сигналы. Генетический механизм сделал возможным производство многочисленных экземпляров одного вида. На первой стадии это имело смысл как производство избыточного биологического материала, с тем чтобы сделать возможным отбор случайно возникших полезных изменений. Появление мозга сделало возможной эффективную адаптацию на уровне и в течение жизни отдельного индивида. Появление общества изменило функциональную роль индивидов: теперь они уже не столько расходный материал, сколько субстанция единой сложной сети с эмерджентными свойствами.

### 3.1.3 Вывод из раздела 3.1

Внедрению номиналистической онтологии<sup>2</sup> в науки об обществе мешают традиционные философские презумпции относительно того, что общество

---

<sup>2</sup>Основы номиналистической онтологии изложены в разделе 1.3.2

состоит из крайне сложно организованных индивидов, наделённых волей и сознанием. Однако, именно **когнитивные теории, построенные на той же номиналистической онтологии, способны не только справиться с интенциональностью и относительным индетерминизмом социальных акторов, но и способствовать реальной интеграции когнитивных и социальных наук.**

## 3.2 Концепция метасети

### 3.2.1 Социология сетевых обществ

Подход, который условно можно назвать «сетевой парадигмой»<sup>3</sup>, имеет долгую историю в различных отраслях научного знания, не всегда связанных с вычислениями. В 1736 г. Леонард Эйлер решил задачу о семи кенигсбергских мостах. В результате родилась математическая теория графов, которая до сих пор является основным математическим инструментом для анализа сетевых структур.

Вместе с тем, методы исследования сетей находили себе дорогу и в социальной науке. В 1973 г. Марк Грановеттер опубликовал знаменитую статью «Сила слабых связей» [155]. В этой статье он пытался найти теоретический мостик между социологическими теориями низшего и высшего уровня. Таким связующим звеном, по мысли автора, должна стать теория, согласно которой социальные воздействия лучше всего распространяются и приводят к более существенным результатам, будучи транслируемы по так называемым слабым сетевым связям, в отличие от сильных — например, семейных — связей. Он писал: «В этой статье я буду доказывать что анализ процессов в межличностных сетях составляет наиболее плодотворный мостик между микро- и макроуровнем. Тем или иным способом, именно через эти сети малые взаимодействия транслируются в более масштабные паттерны, которые, в свою очередь, влияют на малые группы» [155, p. 1360]. Далее он пишет: «Что бы ни распространялось, это может достичь большего количества людей и преодолеть большие социаль-

---

<sup>3</sup>Сетевая (коннекционистская) парадигма в когнитивной науке была рассмотрена в разделе 2.2.2.

ные дистанции... если это распространяется через слабые связи, скорее чем через сильные» [155, р. 1366]. Интересно в этой связи то, что Грановеттер для построения теории использовал методологию и математический аппарат, применяемый для анализа сетевых отношений, заложив, по мнению многих, основы социологии социальных сетей.

В социальной науке присутствует понятие социальной сети, которое не связано с современной реальностью Интернета, но имеет долгую историю в социологии, социальной психологии и социальной антропологии благодаря, прежде всего, таким авторам, как Джеймс Барнз (автор термина), Я. Л. Морено и А. Рэдклиф-Браун. Социальная сеть противопоставляется жёсткой институционализированной социальной структуре как, прежде всего, система неформальных человеческих связей, которая может быть как горизонтальной, так и вертикально интегрированной.

Но импульс к изучению целостных социальных систем (вплоть до глобального мира) дал, безусловно, Мануэль Кастельс. Его теория сетевого общества стала заметным в социологических и социофилософских кругах явлением в середине 1990-х гг. Тогда же появились первые русские переводы его работ и теоретические обзоры, выполненные русскоязычными авторами ( см. [249], а также [250]) .

Нужно сказать, что точность и аналитичность как некие идеалы научного знания не всегда успевают за вольным полётом мысли Кастельса, из-за чего высказанные им в многочисленных работах взгляды и догадки, обильно сдобренные эмпирическим материалом, подчас трудно изложить в виде некоторой последовательной теоретической схемы, избегая противоречий. Ценность трудов этого американского каталонца состоит, скорее, в обилии интересных гипотез и инсайтов, которые могут оказаться бесполезными для будущих теоретиков. Так, Иан ван Дейк, комментатор и оппонент Кастельса, пишет, что если Маркс и Вебер строили свои теории в виде точных и выверенных концептуальных схем, то теоретические разработки Кастельса «гораздо более поверхностны [sketchy], а причинные связи, которые он раскрывает, не достигают того уровня абстракции и обобщения, который мы находим у Маркса и Вебера» [251, р. 128]. Ради справедливости заметим, что сам по себе высокий уровень абстракции и обобщения не всегда идёт на пользу, но определённая особенность стиля мышления Кастельса подмечена верно. Тем не менее, ему

удалось обогатить теоретическую социологию и социальную философию весьма ценными, пусть и недостаточно систематичными, идеями.

Переход от уже ставшей общим местом идеи информационного общества к теории сетевого общества намечается в самой знаменитой трилогии Кастельса «Информационный век: экономика, общество, культура» [252–254]. Здесь автор концентрируется на трёх независимых процессах, которые в конце 1960-х — начале 1970-х сыграли решающую роль в становлении нового общества: речь идёт об ИТ-революции, экономическом кризисе, затронувшем как капитализм, так и «этатизм» (коммунизм), а также о расцвете таких новых социальных движений, как феминизм и «зелёные». Как считает Кастельс, революция в области информационных технологий внесла свой вклад в крах этатизма, вылившийся в распад СССР и советского блока, а также в обновление и придание новой силы, гибкости и эффективности капитализму. Появление новых социальных движений стало результатом кризиса национального государства, демократии, традиционных институтов гражданского общества и патриархальных устоев. А сочетание этих тенденций привело к формированию новой социальной морфологии — сетевого общества, новой экономики — глобальной информационной экономики и новой культуры — культуры «реальной виртуальности».

Как впоследствии ещё неоднократно укажет Кастельс, сетевая форма социальной организации существовала и в иные времена и в иных местах, однако информационные технологии создали базис для её всеобъемлющей экспансии по всей социальной структуре. Сетевое общество построено по принципу сети сетей, поэтому сети внутри общества могут превалировать одна над другой. Вместе с тем, причинное воздействие сетевых связей становится более важным, чем те конкретные интересы, которые они представляют. Кастельс называет это «приматом социальной морфологии над социальным действием» [253, р. 469].

Новая информационно-технологическая парадигма характеризуется, согласно Кастельсу, несколькими важными признаками: (1) информация выступает своего рода «сырьём», которое подвергается обработке; (2) используется способность информации к повсеместному проникновению; (3) для систем, использующих информационные технологии, характерна «сетевая логика»; (4) характерна для них также гибкость; (5) происходит конвергенция технологий. Так возникает новая форма социальной организации, в которой главным ресурсом экономики и системы власти становится производство и обработка информации [253, р. 21].

Далее, Кастельс приходит к выводу, что власть более не концентрируется в институтах, таких как государство, в организациях, таких как компании, или в центрах символического контроля — корпоративных СМИ и церквях. Она растворяется в глобальных сетях благосостояния, энергетики, информации и визуализации. При этом она не исчезает [253, р. 359] Новая власть содержится в информационных кодах и образах восприятия, создаваемых человеческими сознаниями и выстроенных в личности. Эти сознания и личности вступают в противоречия с доминирующей «логикой» сетевого общества [253, р. 358]. И на основании этого видения Кастельс формулирует основные тезисы, касающиеся главных социальных противоречий сетевого общества.

В первом из них речь идёт об оппозиции «Сети» и «Личности [the Self]». Кастельс объясняет это противоречие как разрыв между абстрактным универсальным инструментализмом сетевой «логики» и конкретными исторически обусловленными идентичностями, между функцией и смыслом [252, р. 3]. Этот разрыв порождает движения сопротивления функциональной логике сетевого общества: религиозный фундаментализм, национализм, регионализм, экологические движения, феминизм и движения гендерной идентичности. Они появляются в результате кризиса, который испытывают в информационную эпоху такие институты гражданского общества, как рабочее движение и политические партии, ввергая заодно в кризис национальное государство и национальную демократию.

Второй тезис касается пространственно-временного разрыва в существовании разных классов сетевого общества: капитала и труда, глобальных институтов и локальных социальных движений. Если первые из них живут в текущем [instant] времени компьютерных сетей, то вторые — по циферблатам старых хронометров повседневной жизни. К тому же, всё более различаются и их пространства: первые живут в киберпространстве или в космополитической среде высококомобильных сливок общества и информационной элиты, вторые — в старых, привязанных к географии, городах и посёлках. Это даёт основание Кастельсу говорить о «вневременном времени» и о «пространстве потоков» . Третий тезис Кастельса говорит о том, что все эти противоречия кого-то включают в сетевое общество, а кого-то исключают, причём исключают целые сообщества, экономики и страны, проявляясь как возрастающее социальное и информационное неравенство, охватывающее весь мир. Коллапс Советского Союза случился, потому что его экономика не вписалась в инфор-



мационное и сетевое общество, тогда как ускоренное развитие Тихоокеанского региона говорит о том, что эти экономики туда вписались. В третьем томе трилогии Кастельс говорит о больших пространствах Азии, Африки и Южной Америки, которые, беднея и стагнируя, образуют низшие классы современного глобального общества. Не став частью официальной глобальной экономики, эти территории являются важными опорами международной криминальной экономики, среди «отраслей» которой трафик наркотиков, контрабанда, нелегальная торговля оружием, отмывание денег и проституция. По иронии, криминальная экономика активно использует новые коммуникационные технологии, образуя «порочные информационные связи».

Говоря о новой культуре, формируемой сетевым обществом, Кастельс называет её культурой «реальной виртуальности». Здесь «пространство потоков» (капиталов, информации, технологий, организационных взаимодействий, образов, звуков и символов) превосходит и заменяет собой пространство мест, «форма, функция и смысл которых содержатся в них самих в границах физического соприкосновения» [252, р. 423]. Пространство потоков продуцирует вневременное время с его разрывами последовательности событий в сети, ментальностью и одновременностью, ветвлением гиперссылок и т. п.

Пространство потоков и вневременное время образуют культуру «реальной виртуальности», представляющую собой «систему, в которой реальность полностью схватывается, оказывается полностью погружённой в виртуальное производство образов, в мир притворства, где кажимости — не просто проекции на экране, через которые сообщается опыт, но сами становятся опытом» [252, р. 373]. Сетевое общество, таким образом, «деперсонифицирует [disembodies] социальные отношения, вводя культуру реальной виртуальности» [254, р. 349].

«Логика» сетевого общества, таким образом, охватывает все сферы социальной и культурной жизни. Она стремится к саморасширению и всепоглощению, маргинализируя остатки старого общества. Попытки реконструировать их цели и ценности образуют «движения сопротивления», как ретроградные, так и прогрессивные, которые вступают в схватку с сетевым обществом за время, пространство и технологии [253, р. 358].

Кастельс считает, среди прочего, что информационное общество причинно воздействует на капитализм как способ производства, который оно омолаживает, и этатизм, который оно приводит к историческому концу. Еще одна интересная идея состоит в том, что «способы развития придают форму всей

сфере социального поведения, включая, конечно, и символическую коммуникацию» [252, р. 18].

Свою теорию сетевого общества Кастельс определяет как дальнейшее развитие идеи информационного общества, авторами которой явились Дэниел Белл и Аллен Турен, и которой он в своё время сам отдал должное (см. об этом [251]). Однако в середине 1990-х в своей главной трилогии Кастельс, говоря об «информационной эпохе», характеризует соответствующее ей общественное устройство уже как сетевое общество. Позже он объяснит, что отказался от популярного термина «не потому, что знания и информация не занимают центрального места в нашем обществе, а потому что они всегда его занимали, во всех исторически известных обществах» [255, р. 4]. Следовательно, специфика социальных структур современной быстро меняющейся эпохи не может быть установлена через терминологическую отсылку к информации или знаниям.

С другой стороны, и сетевые принципы организации также появились не вчера, а, напротив, сопровождали человечество на протяжении всей его истории, никогда до сих пор, согласно Кастельсу, не определяя лицо какой-либо эпохи. Ранние сетевые формы социальной организации, с одной стороны, благодаря своей гибкости и адаптивности, лучше «следовали тропой эволюции» [255, р. 4], но, в то же время, проигрывали альтернативным социальным моделям там, где требовалась мобилизация и концентрация значительных ресурсов. Перелом произошёл только в наше время, с появлением цифровых информационных технологий, которые придали невиданную ранее эффективность сетевым сообществам, лишив альтернативные формы социальной организации их конкурентных преимуществ.

В этом контексте Кастельс определяет сетевое общество как социальную структуру, складывающуюся на основе микроэлектронных технологий и цифровых компьютерных сетей [255, р. 7]. От века существующие в социуме сетевые «матрицы» получают от сетевых информационных технологий мощный импульс к развитию и вытесняют вертикальные модели.

«Сеть есть формальная структура... Она представляет собою систему взаимосвязанных узлов... Сети — открытые структуры, которые развиваются путём добавления или изъятия узлов в соответствии с меняющимися требованиями программ, приписывающих сетям цели их деятельности. Конечно, решения по поводу этих программ принимаются совместно [socially] за пределами сети. Но, как только программа заложена в логику сети, последняя будет эффектив-

но следовать её инструкциям, добавляя, удаляя и меняя конфигурацию, пока новая программа не заменит или не изменит коды, командующие операционной системой» [255, р. 7].

Здесь напрашиваются сразу два возражения. Первое: если управляющая программа привносится в сетевую структуру извне, то — откуда? Мы исследуем современное общество при помощи сетевой модели и, выбрав её в качестве стратегии, должны полагать, что она или исчерпывает исследуемую систему, или составляет её наиболее важную, определяющую часть. Иначе странно было бы именовать это общество сетевым. Но тогда получается, что в случае внешнего программирования, на котором настаивает Кастельс, управляющая программа привносится или из другого общества (что иногда случается), или из несетевой части того же самого.

Нужно сказать, что существование последней прямо предполагается Кастельсом: сетевое общество «распространяется по всему миру, но не включает всё человечество» [255, р. 5]. Однако из контекста ясно, что эта оставшаяся часть человечества принадлежит предшествующей эпохе и вряд ли способна породить программы, управляющие более прогрессивной структурой. К тому же, тезис о внешнем программировании противоречит другому определению сети от того же автора — как самоконфигурирующейся системы коммуникации [256, р. 2]. Действительно, исследователи сетевых моделей, например, искусственного интеллекта подчёркивают самообучаемость как их важнейшее свойство. В примыкающей области — в исследованиях нейрокомпьютеров, архитектура которых имитирует нейронную сеть головного мозга, — насколько мне известно, отсутствует программирование в привычном нам понимании, уступая место «обучению на примерах». Одним словом, предположением о внешнем программировании сети Кастельс перечёркивает преимущества сетевой организации как таковой, связанные с её большей гибкостью и самообучаемостью.

В этом же русле лежит и моё второе — более формальное — возражение: о «кодах, меняющих операционную систему». Трудно до конца понять эту социально-компьютерную аналогию (возможно, под ОС имеется в виду культура), но ОС в известных ныне нейрокомпьютерах достаточно проста (она управляет связями нейронов, прежде всего их «весом») и мультиплицируема, что позволяет запускать параллельно задачи, которые традиционные, линейные компьютеры могут решать только последовательно. Поэтому нет нужды

её менять. А замена текущих приложений может в основном осуществляться путём самообучения, что и делает сетевую организацию более эффективной.

Отдельного замечания заслуживает тезис Кастельса о том, что «сети составляют фундаментальную основу всех видов жизни» [256, р. 2]. Это утверждение возникает как продолжение мысли о том, что сеть как модель не специфична как для современного общества, так и для общества вообще — природа также поставляет примеры сетевых организаций. Контекст возражений не вызывает, но сам тезис нуждается очень серьёзном сопряжении с теоретической биологией. Иначе мы рискуем оказаться в плену метафизических догм. На самом деле, конечно же, мы имеем дело с различными объективными явлениями, к которым применима концептуальная модель сети, что ничего не говорит о генетической связи между ними.

Кастельс также считает, что сеть сочетает единство цели и гибкость в её достижении [256, р. 2]. Это, впрочем, следует из рассмотренных выше тезисов: если инсталлированная программа предполагает некую цель, то все включённые в сеть узлы и связи между ними «выстраиваются» под неё и, меняя вес каждой связи в процессе обучения, опытным путём находят оптимальную стратегию её достижения. Из этой же теории следует, что единицей является сеть, а не отдельные узлы [256, р. 1]. Ценность узла — в его способности служить достижению целей сети. Цели определяются программой, а программа в целом может быть реализована только на большом множестве узлов и их отношений. В определённой мере это коррелирует с некогда широко пропагандировавшейся в Интернете идеологией Web 2.0: ценность сетевого ресурса для вновь пришедших тем выше, чем больше у него уже зарегистрированных участников — это значит больше контента и коммуникативных возможностей.

А корреляция эта, замечу, не случайна, поскольку технологии Web 2.0, наряду с GSM, GPRS, GPS, Wi-Fi и многими другими, составляющими арсенал человека нынешней эпохи, и являются тем фундаментом, на котором стоит современное сетевое общество, и который обеспечивает ему решающее конкурентное преимущество. Не имея этого преимущества, сетевые организации, согласно Кастельсу, ранее в истории проигрывали вертикально организованным структурам в способности к мобилизации ресурсов для выполнения проектов сверх определённого размера и сложности организации [255, р. 4]. «Так, в исторической перспективе сети локализовались в сфере частной жизни, тогда как мир производства, власти и войн был оккупирован большими вертикальными

организациями, такими как государства, церкви, армии и корпорации, которые могли командовать широкими пулами ресурсов во имя достижения цели, определённой центральной властью» [255, р. 4].

Но, несмотря на временное превосходство вертикальных социальных моделей в прежние исторические эпохи, тем не менее, согласно Кастельсу, картина истории человечества как истории сменяющихся друг друга вертикальных бюрократических образований противоречит фактам. Эта картина представляет собой искажённый образ истории, основанный скорее на этноцентризме и апологии существующих режимов, ищущих своё оправдание в событиях далёких веков, чем на научном исследовании мультикультурного мира в его сложности. На самом деле в древности имела место глобализация своего рода (если перенести это современное понятие в контекст древней географии и возможностей транспортных средств): локальные общества жизненно зависели от включённости их функциональных элементов в глобальные сети [256, р. 3]. Иными словами, у сетевой модели глобализации были крайне невысокие, но всё же реальные шансы осуществиться.

Но, с другой стороны, невнимание к роли сетевых организаций в структуре и динамике обществ может быть отчасти обусловлено реальной подчиненностью сетей «логике» вертикальных организаций, власть которых была запрограммирована в социальных институтах и распространяема в однонаправленных потоках информации и ресурсов [256, р. 3]. Иерархические бюрократии преобладали в истории, поскольку старые информационные и транспортные технологии склоняли к однонаправленному распространению информации. Как пишет Кастельс, «Да, ветряные суда могли создать морские пути сообщения и даже трансокеанскую торговую и военную сеть. И верховые посланники или быстро бегающие гонцы могли поддерживать связь от центра к периферии огромных по территории империй. Но время задержки в цепи обратной связи в процессе коммуникации было таково, что логика системы свелась к одностороннему потоку передачи информации и знаний. В этих условиях, сети были расширением власти, сосредоточенной на верхних этажах вертикальных организаций, которые и оформляли историю человечества: государства, религиозные аппараты, военных стратегов, армии, бюрократии и подчиненные им структуры, отвечающие за производство, торговлю и культуру» [256, р. 4].

Я бы предположил, что, помимо отсталых технологий, на доминирование вертикальных бюрократических структур работала высокая степень военной

опасности, о чём говорят некоторые недавние историко-антропологические исследования (см. [257]). С её повышением империи множились и укреплялись, а когда она снижалась — распадались. Так, вполне сетевое сообщество древнегреческих полисов, несмотря на отсталые транспортные и информационные технологии, демонстрировало устойчивость, до тех пор, пока могло справиться с военной опасностью со стороны Персии. Как только соотношение сил изменилось, а на севере, в Македонии, сформировался имперский центр, на него тут же возник спрос в доселе свободных торговых городах, оформившийся в промакедонские партии в наиболее крупных из них, и триумф Александра был исторически предрешён.

Если это предположение верно, то мы, уже независимо от Кастельса, можем указать на послевоенную сетевую организацию мирового сообщества, включающую множество правительственных, надправительственных и неправительственных организаций с постоянно меняющейся формой отношений между ними и запрограммированную на предотвращение войн, как на один из важнейших «спусковых крючков» современной глобализации, осуществляемой по сетевой модели.

С самого начала, с первых строк его публикаций середины 2000-х, сетевое общество помещается Кастельсом в контекст глобализационных процессов. Он пишет: «...то, что мы называем глобализацией, есть ещё одно обозначение сетевого общества, хотя и более описательное и менее аналитическое, чем то подразумевает концепция сетевого общества. Тем не менее, поскольку сети проявляют избирательность в соответствии с заложенными в них программами, поскольку они могут одновременно вступать и не вступать в коммуникацию, сетевое общество распространяется во всем мире, но не включает в себя всё человечество. На самом деле, сейчас, в начале XXI века, оно исключает большую часть человечества, хотя всё человечество зависит от его логики, и от властных отношений, которые взаимодействуют в глобальных сетях социальной организации» [255, р. 5].

Итак, глобализация, по Кастельсу, – синоним сетевого общества. С политической точки зрения «поскольку сетевое общество глобально, государство сетевого общества не может действовать только или преимущественно в национальном контексте. Оно вынужденно вовлекается в процесс глобального управления, только без глобального правительства» [255, р. 15]». Таким образом, «глобализация есть форма, которую принимает сетевое общество,

распространяясь в планетарном масштабе, в то время как новые технологии коммуникации и перемещения создают необходимую инфраструктуру для процесса глобализации» [255, р. 16].

Появляется новая форма государства, которая заменяет собой национальное государство индустриальной эпохи. Глобализация сетевого общества состоит в формировании сети глобальных сетей, «которая избирательно связывает по всей планете функциональные измерения сообществ. Поскольку сетевое общество глобально, государство сетевого общества не может действовать только или главным образом в национальном контексте. Оно вынуждено вовлекаться в процесс глобального управления без глобального правительства» [255, р. 15]. «Мировое правительство», которым пугают доверчивых, могло бы иметь место при глобализации по имперской модели. В случае сетевой глобализации мир управляется «сетью сетей», которая — дополним Кастельса — может включать в себя некоторые имперские правительства в качестве своих узлов, вынуждая их действовать в соответствии с общесетевой программой и тем самым постепенно меняя их природу.

Формирование глобальной сети происходит через процесс политической децентрализации, запущенный национальными государствами, который вовлекает в себя региональные и местные власти, а также негосударственные организации (НГО), часто ассоциированные с политическим руководством. Поэтому подлинная система управления современным миром не выстраивается вокруг национальных государств, хотя последние никуда не деваются. «Политическое управление осуществляется в сети политических институтов, которая делится своим суверенитетом в различной степени и перестраивает себя в переменной геополитической геометрии» [255, р. 15].

Новые коммуникационные технологии также способствуют деятельности комплексного сетевого государства, но именно способствуют, а не определяют её. Переход от национального к сетевому государству — это организационный и политический процесс, направляемый изменениями в политическом менеджменте, представительстве и доминировании в условиях сетевого общества. Поэтому сетевое общество — это не будущее, которое нас ожидает, не следующая ступень прогресса человечества, связанная с новой технологической парадигмой. Это наше нынешнее общество, представленное в разных точках мира в разной степени и в разных формах, в зависимости от истории и культуры. «Любая политика, любая стратегия, любой гуманитарный проект должны исходить из

этого основополагающего факта. Это для нас не направление движения, а пункт начала движения, к чему бы “мы” ни хотели прийти — будь это рай, или ад, или просто отремонтированный дом» [255, р. 16].

Однако национальные правительства, как утверждает Кастельс, часто относятся к Интернету двусмысленно. Они охотно принимают его преимущества, но, в то же время, боятся потерять контроль над информацией и коммуникацией — на чём всегда держалась власть. Принять демократию в сфере коммуникации — значит принять прямую демократию: то, чего не принимало ни одно государство в истории. Любая попытка переопределить права собственности угрожает средоточию легитимности капитализма. Принимая, что пользователи суть производители технологии, мы бросаем вызов власти эксперта. «Поэтому, инновационная, но прагматическая политика должна будет найти средний путь между тем, что в данном контексте социально и политически осуществимо, и совершенствованием культурных и организационных условий творчества, на котором основаны инновации, а следовательно и власть, благосостояние и культура в сетевом обществе» [255, р. 20].

### 3.2.2 Когнитивные надстройки над социальными сетями

Вместе с тем, нельзя не заметить, что сеть нейронов мозга и социальная сеть структурно аналогичны: обе состоят из элементов, умеющих выполнять некоторый набор функций и взвешивать связи с близлежащими элементами. Обе имеют когнитивные надстройки, которые возникают эмерджентно: в случае с мозгом это так называемый когнитом (термин К. В. Анохина, см. параграф 2.2.3), тогда как в случае с социальными сетями — это когнитивные социальные сети. Если мы имеем машины распределённой сетевой архитектуры, управляемые разными программами, то между ними возможен интерфейс — библиотека функций, переводящих команды одного языка в команды другого. Таким интерфейсом между мозгом и обществом, согласно защищаемой здесь концепции, выступает естественный язык, развитый людьми в процессе социальной эволюции.

То, что мы в нашей культуре называем мышлением, на самом деле, по мысли Л.С. Выготского, представляет собой внутреннюю речь — свёрнутое,



интериоризированное движение смыслов, подчиняющееся явным и неявным правилам языка. Минимальные условия осмысленности языка, как и любой знаковой системы, сформулированы в аристотелевской логике. Это линейное, по своей архитектуре, мышление создает картину, отличную от того, как на самом деле работает мозг, и как на самом деле устроено общество. Именно здесь корень многих трудноразрешимых проблем эпистемологии и социальной философии.

Как писал Эвиатар Зерубавел, один из пионеров когнитивной социологии, когнитивная наука исходит из романтического идеала одинокого мыслителя или из универсалистских концепций; когнитивная социология – из социальных норм и условностей [35].

Социология когнитивных социальных сетей (КСС) изучает, каким образом изменения в информационных последовательностях, количестве источников информации (узлов в информационной сети) и в видах источников (человеческие или технические) могут влиять на доверие к получаемой информации и на процессы принятия решений в сетевой среде. Здесь используются когнитивные модели для прогнозирования поведения идеального человека-исполнителя, измеряется функционирование реального человека в соотнесении с этими идеальными моделями, и в результате определяется, как обратная связь и обучение могут быть использованы для улучшения поведения человека в области принятия решений. Такие исследования финансируются, в частности, Министерством обороны США [41].

Как показано в [39], идеи КСС выросли из современных исследований в рамках анализа социальных сетей (АСС), которые, в свою очередь, развились из различных направлений исследований в области математики, физики и социальных наук. В рамках АСС существует давняя «творческая напряженность» между взглядом на сети как на взаимодействия и как на когниции. При изучении этой границы между когнитивной и поведенческой сферами сетей в центре внимания большинства исследований КСС было сравнение и сопоставление сетевых когниций (т. е., восприятий сетевых взаимодействий) людей с окружающей их реальной социальной сетью.

Ключевым вкладом исследований КСС стало расширение и углубление подходов к социальным сетям с привлечением внимания к когнитивным аспектам социальных сетей. Это обусловлено феноменологической позицией, которую занимают исследователи КСС в отношении того, как мы должны понимать сетевые познания отдельных людей. Традиционно исследования со-

циальных сетей интересовались восприятием людьми своих социальных сетей в той мере, в какой это восприятие является репрезентативным для реальных паттернов взаимодействий. Напротив, исследования КСС рассматривают восприятие сети как самостоятельное явление. Этот подход берет свое начало в теории поля Курта Левина, согласно которой поведение индивидов определяется их субъективным опытом социальной среды.

Когнитивный подход также предлагает иную интерпретацию конструкций, которые измеряются с помощью социальных сетей. Исследователи КСС интересуются субъективным восприятием людьми своего социального окружения. Прося людей описать окружающую их социальную структуру, наука о КСС стремится раскрыть основные когнитивные схемы социальных отношений. Таким образом, исследования когнитивных социальных сетей не делают никаких предположений о том, как восприятие одного человека будет соотноситься с восприятием других. На самом деле, любые различия между восприятием одного человека и восприятием других в той же сети представляют собой теоретически и эмпирически интересный вопрос. Этот подход отличается от электронного поиска в социальных сетях, который рассматривает сообщения людей об окружающих их отношениях как имеющие отношение к реальным паттернам взаимодействия. Напротив, традиционные исследования социальных сетей исходят из того, что восприятие людьми своей социальной структуры должно коррелировать с восприятием той же сети другими.

Наконец, когнитивный подход добавил еще один уровень к нашему пониманию того, как на результаты людей влияют их окружающие социальные сети. С точки зрения АСС, поведение людей ограничивается и поддерживается окружающей их социальной структурой. В рамках этого подхода считается, что люди могут извлекать выгоду из социального капитала, встроенного в их окружающие отношения, в той мере, в какой они занимают благоприятные позиции в этой структуре. Исследования КСС углубили это понимание, подчеркнув важность восприятия людьми своих сетей. Здесь предполагается, что люди не могут мобилизовать социальный капитал, доступный в их окружающих сетях, если они не могут точно понять отношения, в которые он встроены. Таким образом, исследования КСС исходят из того, что преимущество связано с остротой восприятия. Острота восприятия, конечно же, ограничена пределами обработки информации человеческим разумом. Именно эти ограничения, поскольку они

вливают на нашу способность воспринимать социальные сети, являются ключевым направлением исследований КСС.

### 3.2.3 Метасетевой подход к социально-когнитивным процессам

В [70; 258—260] мною была разработана теоретико-методологическая основа междисциплинарной интеграции, которую можно обозначить как «метасетевой подход» или, может быть, метасетевая теория (МСТ)<sup>4</sup>. Эта рабочая рамка представляет собой синтез коннекционизма — нейросетевого подхода в когнитивных науках — и сетевой концепции общества. Она предполагает, что в ходе эволюции обе сети — церебральная и социальная — объединяются в единую суперсистему, давая нашему виду решающее эволюционное преимущество.

Если мы примем метафору нейрокомпьютера, мы должны признать, что у этого устройства нет операционной системы в привычном смысле. Скорее, он функционирует на основе встроенного ПО узлов своей сети. Это встроенное ПО может состоять, например, из примитивных правил порога, после превышения которого возбуждение, полученное от соседних узлов, передаётся дальше, другим соседним узлам. Говоря об «операционной системе» сознания или общества, мы должны понимать, что общество — это не примитивная сеть. Это *метасеть*, состоящая из многих нейрокомпьютеров, т.е. сетей. Поэтому важно подчеркнуть, что метафора операционной системы человеческого мира предполагает, что такая система не программируется и не устанавливается никаким высшим существом. Скорее, она создаётся в режиме обучения в процессе эволюционного взаимодействия человеческого мира со своим окружением. Она включает в себя и то, что обычно называют культурой.

Достоинство моего подхода, вкратце, состоит в том, что он предлагает единую «сетевую» онтологию для когнитивных и социальных наук, что

---

<sup>4</sup>Строго говоря, в указанных публикациях речь шла о «гиперсетевой теории» (ГСТ). Однако этот термин уже используется К. В. Анохиным [20; 21]. Причём, как я позже убедился, в контексте его концепции этот термин более адекватен. Гиперсеть, если опустить тонкости формальных определений, — это дополнительный уровень сетевых связей в рамках одной и той же сети. Метасеть, в общем виде, — это структура, внешняя по отношению к исходной сети (см., напр., [261]). Таким образом, последний термин лучше приспособлен для задач настоящего исследования.

принципиально позволяет использовать один и тот же набор формальных инструментов, осуществляя «сквозные» социально-когнитивные исследования.

Когнитивный мейнстрим связывает когницию с тем, что мы понимаем под словами «помнить», «думать», «рассуждать» и т.п. Однако этот словарь имеет отношение скорее к «народной психологии», чем к науке (см. об этом, напр., [262]). Высказывания «народной психологии» имеют тот же эпистемологический статус, что и «превращённые формы» у Маркса: не будучи истинными *sub specie aeternitatis*, они адекватны в контексте внутренних процессов социальной системы. Однако же, точка зрения науки — это именно точка зрения *sub specie aeternitatis*, пусть даже только в нормативном плане.

С моей точки зрения, когнитивные акты осуществляются как подстройка метасети в соответствии с поведением рационального агента и его последствиями. Паттерн возбуждения определённых нейронов головного мозга, которое имеет место в ходе когнитивных актов, мейнстримом рассматривается как «репрезентация», в то время как, с точки зрения МСТ, это аналог программной библиотеки, которая хранит типичные процедуры для типичных обстоятельств и активируется в соответствующих случаях. Способность к запоминанию — это самая примитивная когнитивная способность. В то же время, мышление и рассуждения — благодаря системе социальных связей и приобретённым социально-когнитивным функциям, к которым относится, например, язык, — надстраиваются над нею в качестве более общей способности видеть свою собственную и чужую память в качестве части социальной задачи или проблемной ситуации. Если некто встречает существ, которые также могут хранить нечто в памяти, он или она склонны описывать такую ситуацию примерно так: «некий факт  $F$  имеет место, и наш собеседник  $a$  знает, что это так». Тогда, «знать» означает «хранить в памяти тот факт, который сохраняется в памяти широкого круга собеседников».

Такая позиция, безусловно, является функционалистской. Более того, она совмещает ментальный функционализм — понимание интенциональных актов и состояний как функций, независимых от природы носителя, — с лингвистическим: в соответствии с нею, такие слова, как «знать» не только не относятся к ничему, кроме функциональных отношений в мире, но их собственная функция в языке связана не с референциальным значением, но исключительно инструментальна. Такое слово не называет внешний предмет или сущность (например,

состояние мозга), не указывает на них, а создаёт то, что можно было бы назвать коммуникативной модальностью: «*p* имеет место, но *a* об этом не знает».

Мы также можем говорить о *коммуникативной интенциональности*: это означает, что ментальное состояние или высказывание, имеющее значение, не только направлено на объект, но также с необходимостью *предполагает собеседника*, который может быть даже несуществующим, но, тем не менее, эпистемически отличаться от говорящего — например, не знать, то, что знает говорящий.

Сети как математические объекты достаточно хорошо изучены, и это даёт надежду на то, что и социология и психология прибавят в точности своих методов, воспользовавшись этой моделью (что в ряде случаев и происходит). Я же надеюсь выяснить, каким образом сознание может быть объяснено как эффект метасетевых взаимодействий. Здесь необходимо пояснить термины «метасеть» и «метасетевое взаимодействие». В данном случае имеется в виду не просто «сеть сетей», а взаимодействие сетей через некоторые интерфейсы. Необходимость интерфейсов возникает постольку, поскольку разные сети реализуют разные алгоритмы, и, следовательно, установление прямых связей между их узлами невозможно. Случай с человеческим сознанием — хороший пример. Некоторые психические явления — например, качественные субъективные состояния (квалиа), — являются функциональным эффектом нейронной сети мозга, тогда как язык и его семантика — функциональным эффектом социальной сети. Прямой обмен между узлами этих сетей невозможен, поскольку в каком-то смысле одна из этих сетей представляет собой узел другой. Поэтому взаимодействие осуществляется через интерфейс — семантически нагруженный язык. Подобным образом интерфейсом между нейросетью мозга и миром физических объектов выступает человеческий чувственный опыт.

В рамках нашей привычной картины мира, несмотря на все искушения, мы не можем представить природу как ещё одну сеть — это было бы слишком решительной революцией в науке<sup>5</sup>. Модель описания и объяснения естественных вещей и явлений сформирована естествознанием, как оно сложилось: мы имеем объекты, относящиеся к немногочисленным категориям, и управляющие ими естественные законы, которые, в отличие от человеческих, невозможно изменить. И, по крайней мере, в неживой природе, нет места программированию и

<sup>5</sup>Хотя целый ряд природных объектов неплохо описывается сетевым лексиконом (см., напр., [31; 92; 263]).

вариативности. Именно поэтому мозг как нейросеть использует чувственность и чувствительность как простой, встроенный (*embedded*) и не подлежащий пере-программированию интерфейс. Напротив, язык представляет собой инструмент сложный и настраиваемый пользователем, изменяемый со временем. Причём изменения эти, как свидетельствует современная лингвистика, вполне изучаемы в русле естественнонаучной парадигмы – как управляемые явно формулируемыми законами, позволяющими, например, обратную реконструкцию архаичных морфем и фонем из ныне имеющих.

Интересным и озадачивающим выглядит тот факт, что мозг, будучи компьютером нейросетевой архитектуры, оказывается инструментом мышления, построенного противоположным образом – как линейное и последовательное логическое или математическое исчисление. Возможным объяснением может быть то, что так понимаемое мышление является функцией не столько мозга, сколько интерфейса между ним и социумом – функцией языка<sup>6</sup>. Поэтому и мыслимый мир, будучи проекцией семантики языка в область, трансцендентную нейросети, оказывается линейно организованным в цепочки причинно-следственных связей. На этом основана вся проблематика эпистемологии: как соотносится линейно организованный язык и производные от него иерархические дедуктивные системы со своими предполагаемыми объективными референтами? Но возможность инсайтов, внезапных озарений, неформализуемого творчества указывает на «другое мышление», которое предположительно может быть понято как внутренняя функция нейросети, случайно формируемая, адаптивная по своей природе и потому недоступная для формализации.

Мышление «от языка» доступно для исследования, поскольку оно изначально овнешнено, будучи функцией интерфейса. «Другое мышление» представляет собой внутреннюю жизнь нейросети и поэтому может быть представлено в форме объекта только при условии резкого изменения исследовательского угла зрения.

Попутно замечу, что, возможно, здесь кроется ключ к правильной интерпретации «сферы возможного опыта» и других кантовских терминов, поскольку очевидно, что чувственный опыт, понятый как интерфейс, прямо отсылает к трансцендентальной эстетике, и то же можно сказать о языковом интерфейсе и трансцендентальной логике. Но это слишком обширная тема, чтобы раскры-

<sup>6</sup>Языка, который, впрочем, активно переформатирует нормальную деятельность мозга, как показывают некоторые эмпирические исследования [264–268]

вать её скороговоркой. Оговорюсь только, что теория метасети не обязательно предполагает скептические выводы.

Для взаимодействия с тем, что не является узлами метасети, мозг использует чувственное восприятие как специальный интерфейс. Чувственные образы аналогичны пиктограммам на рабочем столе компьютера: они не изображают того, с чем связывают нас, но очень удобны для организации взаимодействия с тем, что принципиально скрыто от нашего перцептивного аппарата [104]. Они создаются на основе априорных моделей, поставляемых мозгом, которые многократно уточняются в процессе взаимодействий<sup>7</sup>. Априорную модель можно рассматривать как информационный запрос, требующий ответа «да/нет», изменяющиеся восприятия — как ответ на этот запрос. Поэтому полученный в результате многочисленных уточнений целостный чувственный образ представляет собой то, что стоики называли «*λεκτόν* (лектон)» — значение осмысленного высказывания. Без этого эмпиризм был бы невозможен, поскольку в опыте не содержалось бы знания.

Итак, метасеть понимается здесь как связь нейросети мозга и социальной сети, в которую включён его владелец. Метасеть, согласно этой концепции, является подлинной функциональной основой тех ментальных функций, которые мы относим к специфически человеческому сознанию. Этот теоретический подход возможен благодаря формирующейся на наших глазах «сетевой парадигме»: трансдисциплинарной методологической установке, предполагающей применение математических сетевых моделей для изучения связей нейронов головного мозга, социальных взаимосвязей и новых компьютерных архитектур.

То, что в настоящем исследовании называется «сетевой парадигмой» — использование математических моделей сетевых взаимодействий для исследования различных объективных феноменов — демонстрирует теоретическую и практическую эффективность не только в нейронауке, психологии и философии, но и в социальных науках, где различные теории « сетевого общества », социологические исследования социальных сетей, а в последнее время и социальных когнитивных сетей приводят к интересным результатам. Возникает закономерный вопрос о возможности сквозной методологии, основанной на сетевой парадигме и увязывающей исследования социальных и нейроцеребральных сетей в единую междисциплинарную исследовательскую программу.

<sup>7</sup>Подробнее см. раздел 2.2.5.

Метасеть даёт конкретную модель отношений (близких / опосредованных, сильных / слабых). И если сознание в принципе мыслимо как функция от системы связи нейронов (нейросети), то точно так же оно может мыслиться как функция социальной сети: каждый индивид (узел) обладает частичным знанием и компетенциями, а «всё» знает только сеть в целом. И тогда понятно, что именно социальная сеть, а не нейросеть, ответственна за синтаксис и семантику языка, правила взаимодействия и т. п.

Вместе с тем нужно опасаться неоправданных метафизических обобщений: в духе того, что сети лежат в основе жизни, материи, пронизывают собой всю реальность и т. п. Сетевая модель – это только методология, позволяющая строить эффективные правдоподобные теории относительно, возможно, разнородных объектов.

Сеть мыслится как состоящая из узлов, связей между ними и алгоритмов, управляющих распространением взаимодействий. Узлы – наименее сложный элемент системы. Узел должен обладать достаточными ресурсами для того, чтобы исполнять несложную программу, не тождественную «программе» сети. Так, нейрон искусственной нейросети может находиться в одном из доступных состояний (возбуждения или покоя) и передавать это состояние соседям при определённых условиях. Он также должен эмпирически определять более предпочтительные связи с другими нейронами и менять значения их «весов» в большую сторону. Примерно то же можно сказать об узле любой сети – о пользователе Фейсбука, адепте тайного ордена или члене террористической организации. Они все запрограммированы достаточно простыми функциональными связями между типичными условиями и типичными действиями. Они все знают, что должны делать в определённых типичных ситуациях, число которых ограничено. Идентичность сети определяется этими программными кодами скорее, чем физической природой узлов.

Стандартный пользователь, например, Фейсбука наделён ограниченным набором функций внутри сети: он может публиковать (post), читать (read), выражать одобрение (like), делиться прочитанным (share), присоединиться к группам (join), и т. п. Но даже этого ограниченного набора достаточно для того, чтобы многие из нас проводили в этой сети значительную часть отпущенной нам жизни, со всей серьёзностью относясь к происходящим там процессам. Можно возразить, что полноценное участие в Фейсбуке предполагает не только владение этими нехитрыми функциями, но и знание, как минимум, двух языков



— родного и английского, — погружённость в социальные отношения и политические события, знакомство с общественными условностями и много ещё чего. Но, если разобраться, каждая из этих компетенций может быть смоделирована как сетевая программа, требующая от узлов сети владения ограниченным набором несложных дополнительных функций и превращающая одномерную сеть в гиперсеть (в том смысле, в котором этот термин понимается в теории графов и в рамках концепции когнитома).

И тогда у нас появляется основание обобщить коннекто-когнитомную модель, предложенную К. В. Анохиным<sup>8</sup> до применимости к любым сетям: нейронным, социальным, нейрокомпьютерным. Так, если говорить о социальных сетях, то над сетями, состоящими из взаимодействующих индивидов, надстраиваются когнитивные социальные сети, узлами в которых становятся подсети индивидов, разделяющих определённые убеждения или компетенции. То, что мы на правильном пути в своих рассуждениях, отчасти подтверждается уверенно развивающимися и хорошо финансируемыми зарубежными исследованиями когнитивных социальных сетей.

\* \* \*

Анализ социальных структур когнитивных агентов позволил предложить следующую онтологическую модель. Когнитивные способности агентов обеспечиваются распределённым вычислительным устройством, имеющим сетевую архитектуру и способным к запоминанию, обучению, категоризации, идентификации других агентов. Биологическая необходимость, а также — и, по видимому, в большей степени — стремление к повышению энергоэффективности когнитивных вычислений — объединяет агентов в социальную сеть, которая, как и нейросеть мозга, представляет собой распределённое вычислительное устройство. Обе сети работают как ассоциативные машины, обучающиеся на большом количестве данных.

Однако между агентами в социальной сети эволюционно складывается интерфейс — язык, — который имеет не параллельную, а линейную (серийную) природу, поскольку в его основе лежит исторически случайно обнаруженная у некоторых видов гоминид способность последовательно произносить членораздельные звуки. Таким образом, интерфейс между когнитивной и социальной

---

<sup>8</sup>Подробнее см. 2.2.3.

сетью отличается от них по своей вычислительной архитектуре. С одной стороны, это делает взаимодействия между агентом и социальной сетью более медленными. Но с другой стороны, это обстоятельство порождает целый мир символических систем с их семантикой и прагматикой, даёт начало логическому мышлению, доказательному знанию, а следовательно, философии, науке и, в конечном счёте, всей современной цивилизации, как она сложилась. Если бы эволюция использовала другие способности для коммуникации агентов в сети, вся наша коллективная интенциональность могла бы иметь совершенно другую форму. Единственное, чем мы предопределены мыслить логически — это серийная вычислительная архитектура нашей коммуникации.

### 3.2.4 Вывод из раздела 3.2

Сеть нейронов головного мозга и социальная сеть представляют собой распределённые вычислительные устройства, умеющие рассчитывать вероятностные зависимости между переменными окружения на основе обработки больших массивов данных. Человечество эволюционно выработало интерфейс между этими двумя типами компьютеров сетевой архитектуры — язык, который работает как серийная последовательность команд. Язык не только «переодевает мысли», по словам Витгенштейна [67, с. 18], но и переформатирует под себя мозговую деятельность и социальные отношения. **Социальная сеть, внешняя по отношению к нейросети мозга и связанная с нею языковым интерфейсом, образует метасеть со сложной комплексной вычислительной архитектурой.**

## 3.3 Сетевые и вычислительные подходы в биологии

В предыдущих разделах была сформулирована онтологическая концепция, согласно которой в основе когнитивной и социальной реальности лежат сетевые структуры. В соответствии с трёхуровневой схемой Дэвида Марра<sup>9</sup>,

<sup>9</sup>Подробнее см. раздел 2.3.2.

они являют собой уровень материальных реализаций распределённых вычислений — соответственно, когнитивных и социальных. Однако, наша общенаучная картина мира предполагает, что когнитивные и социальные взаимодействия представляют собой функции живых системам, эволюционно развиваясь из усложняющихся биологических взаимодействий. Если это так, то логично направить исследовательский фокус на такие свойства живой материи, которые генетически лежат в основе соответствующих свойств когнитивных и социальных систем. Иными словами, логика исследования подталкивает к обнаружению распределённых вычислений на биологическом уровне организации материальных систем.

### 3.3.1 Онтологии жизни

В 2002 году Юрий Лазебник, тогда еще ассоциированный профессор лаборатории Колд Спринг Харбор, опубликовал статью [269], в которой в том числе содержалась личная история. При переезде из России в США его жена привезла с собой сломанный радиоприемник, сделанный в СССР. Размышляя над сформулированным им самим парадоксом — чем больше мы знаем эмпирических фактов, тем меньше мы понимаем в биологии — Юрий обратился к этому устройству как к метафоре, объясняющей, чего биологам не хватает в обращении с эмпирическими фактами. По его словам, если бы биолог попытался выяснить принципы работы радио, чтобы починить его, он или она прошли бы долгий путь, классифицируя его мелкие компоненты по цвету и форме, отключая их по одному, чтобы увидеть, как это влияет на звук и т. д. В конце концов такой исследователь придёт к предполагаемой функциональной схеме, состоящей из меток и стрелок, которая будет гипотетически и на качественном уровне представлять, какие части важны для воспроизведения звука. Независимо от того, можем ли мы назвать этот результат новым знанием, данная схема, даже если она в целом окажется случайно истинной, не позволит предсказать какие-либо факты, поскольку в ней не используются формальный язык описания и количественные показатели, подобные тем, которые применяются радиоинженерами. Таким образом, она не сообщает нам, как система на самом деле *настроена* для работы. Лазебник приходит к выводу, что вероятность то-

го, что биолог починит радиоприемник, примерно равна вероятности того, что обезьяна случайно напечатает стихотворение Роберта Бёрнса.

По его мнению, биологии не хватает формального языка, который, подобно языку радиосхем, содержал бы термины для типичных абстрактных элементов и их количественных свойств. Такой язык, если бы он существовал, позволил бы строить объяснительные модели, помещая типовые элементы в различные комбинации и проецируя их функциональные связи, а также указывая числовые значения их свойств. Радио представляет собой яркую метафору того, в чем нуждается биология. Но следует отметить, что не только инженеры владеют таким языком — физика как наука началась, когда Ньютон установил, что мир механики должен содержать только физические тела, которые характеризуются исключительно массой и движутся прямолинейно и равномерно, если на него не действует какая-либо сила. Если сопоставить её с нашим повседневным опытом, эта картина противоречит здравому смыслу. Но она оказалась эффективным инструментом объяснения вместе с соответствующей ей математикой.

Здесь я хотел бы предложить важное различие. То, что Лазебник подразумевает под «формальным языком», не является языком высказываний о мире, т. е. утверждений каких-либо положений вещей. Его аналогия с радиосхемами показывает, что здесь имеется в виду язык, называющий соответствующие типы объектов и их отношений. В настоящем исследовании такая система именования и классификации называется *предметной онтологией*<sup>10</sup>. Кроме этого, наука нуждается ещё в одном формальном языке, построенном на основе первого, для генерации описательных предложений. Этот *язык высказываний*<sup>11</sup>, вероятно, будет математическим — по крайней мере, так было до сих пор, и это положительно сказывалось на эпистемической эффективности науки. Следовательно, если мы принимаем аргумент Лазебника, то проблема с биологией не в недостатке математики в ней, поскольку мы видели многочисленные попытки количественно выразить или формализовать наши знания о жизни: применялась и статистика в постановке экспериментов, и поиск паттернов в биоинформатике, и модели в эволюции, экологии и эпидемиологии [270]. Однако они редко приводили к заметному успеху в том, что можно назвать теоретической интеграцией. Скорее, с этой точки зрения, биологии не хватает единого

<sup>10</sup>Подробнее об этом см. раздел 1.1.3.

<sup>11</sup>Здесь значение термина может отличаться от его значения в логике высказываний. Под «языком высказываний» я имею в виду определённую систему символов, а не набор актуальных высказываний, выражаемых ею.

формального языка, который бы назвал и описал некоторые основные элементы, комбинации которых создают различные конструкции живой материи. Таким образом, проблема скорее в онтологии предметной области.

Есть популярные взгляды, согласно которым онтология предметной области выводится из собственно теории. В частности, она понимается как система экзистенциальных пресуппозиций ( $E$ ), подразумеваемых положениями теории ( $T$ ). Например, если ваша теория утверждает, что живые организмы развиваются в зависимости от наследственности и изменчивости, она подразумевает, что существуют живые организмы<sup>12</sup>. На этом основании часто делается вывод, что если  $T$  истинна и подразумевает определенные  $E$ , то и последние верны. Но вспомним, что, когда в 1824 году Сади Карно предложил цикл своего имени, он был сторонником теории теплорода, т. е. считал, что тепло переносится самоотталкивающей жидкостью. Если истинное описание цикла Карно — это  $T$ , а вера в теплород —  $E$ , и имеет место  $T \vdash E$ , то мы до сих пор должны верить в теплород, что не так. Следовательно, отношение  $T$  и  $E$  — это не отношение вывода, а скорее отношение интерпретации. Т. е., формально говоря, чтобы теоретическое высказывание  $T$  было истинным, составляющие его термины должны интерпретироваться на той или иной *онтологической модели*, которая в примере Лазебника представляет собой набор абстрактных радиоэлементов (конденсатор, резистор и т. д.). Но мы могли бы также попробовать интерпретировать схему его радиоприемника, скажем, на персонажах пьесы Шекспира. Я подозреваю, что при соответствующей количественной настройке этой онтологической модели мы могли бы сохранить это схему как валидную теорию. Другое дело, что такая интерпретация не была бы «честной» в смысле Коплэнда [169].

Следовательно, онтологии предметной области логически независимы от собственно теорий.

Иллюзия того, что  $T \vdash E$  верно для естественных наук, проистекает из наших рассуждений на естественном языке, согласно которым, если «Сегодняшний король Франции лыс», то он существует. Но в научном контексте  $T$  обычно представляют собой формальные выражения, которые, помимо прочего, могут быть истинными, будучи интерпретируемыми на соответствующей модели. Но факт состоит также в том, что для теории  $T$  может существовать несколько релевантных моделей, при которых она сохраняет своё истинностное значение. Это важно для дальнейшего обсуждения.

<sup>12</sup>Моё обсуждение этого вопроса см. в [83].

### 3.3.2 Математические подходы в биологии

Как уже было сказано, до сих пор было много попыток инкорпорировать математику в биологию, и некоторые из них были успешными на локальном уровне, но они принципиально не меняют общей картины, поскольку теоретическая биология все ещё не имеет универсальных принципов и единого формального языка, с помощью которых можно было бы вывести все истинные предложения. Одна из наиболее интересных попыток — это статья Алана Тьюринга [271] о морфогенезе, в которой автор наиболее жизнеспособной теории цифровых вычислений предложил некоторые аналоговые инструменты для объяснения возникновения биологической сложности из начальной однородности. Формальные инструменты сводились в основном к линейным дифференциальным уравнениям с постоянными коэффициентами.

Его теория постулирует два морфогена: один называется «активатором», а второй — «ингибитором». Активатор воспроизводит себя со скоростью, пропорциональной его избытку. Он также производит ингибитор, которым он сам естественным образом подавляется. Хотя они оба распространяются по всему пространству, ингибитор делает это быстрее. Согласно расчетам Тьюринга, на первом этапе первоначальная однородность нарушается небольшими случайными изменениями, а на следующих этапах случайные колебания усиливаются. В конце концов, активатор собирается в несколько пятен, между которыми — пустые участки. Таким образом, в результате серии бифуркаций внутри изначально однородного раствора, структура рождается из единообразия.

Басконт в [272] упоминает математическую теорию детерминированного хаоса, которая вдохновлена в основном биологией и использует нелинейные уравнения, в которых округлённые до определенного числа десятичных знаков представления действительно-числовых данных, полученных из наблюдений сложных неравновесных систем, вызывают со временем неожиданные последствия. Но, согласно Хофмейру в [273], такие системы являются не более чем *симуляциями* объекта, в то время как зрелые науки, такие как физика, достигают уровня *моделирования*. Он объясняет разницу между этими понятиями с помощью простой схемы. Пусть есть естественные объекты ( $N$ ) и причинно-следственные связи ( $C$ ) между ними. И есть предложения внутри формальной теории ( $F$ ) и отношения вывода ( $I$ ) между ними. Теоретик создает словарь

кодирования  $\varepsilon$ , который отображает наблюдаемые в  $N$  входные переменные в  $F$ . А ещё есть словарь декодирования  $\delta$  для обратного отображения каждого  $f \in F$  в определенный  $n \in N$ .

Тогда  $F$  является моделью  $N$ , если и только если (1)  $F = \varepsilon(N)$  и (2)  $I = \varepsilon(C)$ . Если верно только (1), но не (2),  $F$  является симуляцией  $N$ . В последнем случае предсказания не гарантированы, так как не обеспечено  $N = \delta(F)$ .

В качестве примера достижений в области теоретических «моделей» Хофмейр приводит работы Николая Рашевского, чей опыт и знания в области математической физики помогли ему сформулировать ряд плодотворных гипотез в области наук о жизни. Так, например,

булева версия его «двухфакторной теории» возбудимых элементов в руках Уолтера Питтса и Уоррена Маккаллоу привела к разработке нейронных сетей, а также выступила предтечей модели распространения потенциалов действия в нейронах Ходжкина и Хаксли [273, р. 2].

Но вскоре Рашевский понял, что все, что он делал, было моделированием переходов между состояниями в различных режимах живой материи. Тогда он перешел «от компонентов биологических систем к отношениям между ними» [там же], и это дало начало так называемой «реляционной биологии». Используемыми математическими инструментами были топология, теория множеств и логика высказываний. Это предприятие продолжил его аспирант Роберт Розен, который в своем исследовании систем восстановления метаболизма опирался на теорию категорий. По оценке Хофмейра,

Если эпоха анализа характеризовалась невысказанным девизом «разделяй и властвуй», то, возможно, эпоха синтеза направлена на «объединение и господство». Сегодня реляционная биология использует богатый набор математических инструментов: теорию категорий, теорию графов, теорию сетей, теорию автоматов, формальные системы, если говорить о наиболее важных [там же].

### 3.3.3 Информационные пути и сети в живых тканях

Исторически сложилось так, что аналитически построенная математика породила номотетическую науку, открывающую законы, в рамках которой биология до недавнего времени не демонстрировала большого успеха по сравнению с физикой. Далее я попытаюсь добавить некоторые свидетельства в пользу гипотезы, что алгоритмическая или — в некотором смысле — конструктивная точка зрения на предметный мир биологии вполне может способствовать развитию некоторых вычислительных онтологий и исчислений, от которых, вероятно, в выигрыше будет биология, а также когнитивные и социальные науки.

В [274] Бхалла и Айенгар утверждают, что живые системы следует рассматривать как сложные сети путей передачи сигналов. Биологически они регулируются межпротеиновыми взаимодействиями, фосфорилированием протеинов, регуляцией элизиматической активности, продуцированием вторичных мессенджеров и системами передачи сигналов на клеточной поверхности. Они утверждают, что

сигнальные пути взаимодействуют друг с другом, и окончательный биологический ответ формируется взаимодействием между путями. Эти взаимодействия приводят к созданию сетей, которые довольно сложны и могут иметь неинтуитивные свойства [274, р. 381].

Сетевые взаимодействия ответственны за такие новые свойства, как увеличенная продолжительность сигнала, активация петель обратной связи, определение пороговой стимуляции для биологических эффектов и множественные выходные сигналы. Этот анализ свидетельствует о том, что при правильном сочетании простые биохимические реакции могут хранить и обрабатывать информацию, а весь механизм реакций внутри сигнальных путей формирует ещё одну биологическую основу для памяти и обучения. Следовательно, как заключают Бхалла и Айенгар, вычислительный подход хорошо подходит для понимания как сложности множественных сигнальных взаимодействий, так и тонких количественных деталей.

Чете и Дойл [275] исследуют такое очевидное свойство живых систем, как их модульность. Они представляют модули как подсистемы более крупной системы, которые используют интерфейсы (протоколы) для подключения



к другим модулям, могут быть относительно независимо изменены, допускают упрощенные описания с целью более абстрактного моделирования, сохраняют свою локальную идентичность при изоляции или перегруппировке и, наконец, заимствуют дополнительную идентичность у остальной системы. Протоколы — это предписанные интерфейсы между модулями, которые ответственны как за облегчение масштабирования системы, так и за появление новых свойств в целом. Введение этих понятий делает очевидным, что

такие абстракции, как регуляция генов  $\langle \dots \rangle$ , ковалентная модификация, мембранные потенциалы, метаболические пути и пути передачи сигналов, потенциалы действия и даже транскрипция-трансляция, клеточный цикл и репликация ДНК, все могут быть разумно описаны как протоколы  $\langle \dots \rangle$ , с их сопутствующими модульными реализациями в различных активаторах и репрессорах, киназах и фосфатазах, ионных каналах, рецепторах, гетеротримерных протеинах, связывающих гуанин-нуклеотид (G-протеины) и т. д. [275, p. 1666].

Предлагаемая онтология достаточно абстрактна для компьютерного моделирования биологических систем и процессов. В то же время необходимо принимать во внимание, что чем более реалистичные модели биологических сетей нам нужны, тем более запутанными и сложными они должны быть, чтобы включать множественные сигналы обратной связи, нелинейную компонентную динамику, неопределённые параметры, стохастический шум, паразитную динамику и другие нечёткие факторы и входные данные. Здесь мы рискуем, принимая во внимание вышеупомянутый детерминированный хаос: неизбежно неточные значения входных данных могут привести к слишком большим ошибкам в предсказаниях теории. Чете и Дойл утверждают, что математические инструменты в настоящее время развиваются, чтобы в конечном итоге справиться с этими проблемами.

Как показано в [276], протеины сами по себе являются крайне универсальными обработчиками информации. В одноклеточных организмах цепи на основе протеинов заменяют всю нервную систему в роли сети, контролирующей поведение. В клетках растений и животных многочисленные организованные в сети протеины передают информацию от плазматической мембраны к геному. Продолжительное воздействие окружающей среды на концентрацию и активность

множества протеинов в клетке — это реальный механизм сохранения в памяти важных данных об окружающей среде. Взаимодействующие белки архитектурно и функционально похожи на нейронные сети. Они эволюционно обучены определять повторяющиеся шаблоны внешних раздражителей и соответствующим образом реагировать на них. Их коннектом зависит от ограниченных диффузией взаимодействий молекул, что обеспечивает уникальные особенности, отсутствующие в искусственных нейронных сетях.

Таким образом, хотя современные технологии все чаще используют вычислительные устройства, «вдохновлённые биологией», как это принято называть, такие как нейронные сети, эволюционные алгоритмы и мультиагентные системы, сама биология нуждается в том, что я бы назвал вычислительной онтологией и вычислительными формальными инструментами.

### 3.3.4 «Принцип свободной энергии» в биологии

Важным примером применения байесовской статистики к биологическим, когнитивным и социальным онтологиям, связанным с *принципом свободной энергии*, являются исследования моделей *предиктивного процессинга*<sup>13</sup> [200; 277—285]. Это интеллектуальное движение, которое часто называют «парадигмой», распространилось как экстенсивно — завоёвывая новых адептов, так и интенсивно — охватывая новые предметы и области. Основные принципы доктрины были сформулированы применительно к когнитивной сфере, но их предполагаемая объяснительная сила подтолкнула основателей к расширению области применения теории, которая теперь охватывает не-когнитивные предметы психологии, а также предметы био- и социальных наук. Основатели и сторонники этого подхода считают свою теоретическую систему «вычислимым (tractable) руководством» для открытий в биологических, когнитивных и социальных науках [282, p. 1].

Шрёдингер однажды заметил, что живые системы «уникальны среди естественных систем, тем что они, как кажется, сопротивляются второму закону термодинамики, сохраняясь как изолированные, самоорганизующиеся системы во времени» [там же]. Трансдисциплинарная область, известная как теория

<sup>13</sup>Теоретические основы ПП были рассмотрены в разделе 2.2.5.

эволюционных систем (ТЭС), была попыткой справиться с этой проблемой. Принцип свободной энергии — это механистическая версия ТЭС, которая применима к живым системам в целом.

Концепция свободной энергии опирается на ряд опорных понятий и принципов. В целом эта теория рассматривает организмы как воплощающие «ожидания, которые они должны гарантированно реализовать посредством адаптивных действий» [277, р. 196].

**Вариативная свободная энергия** определяется как мера разницы между ожидаемым состоянием окружающей среды и фактическими поступающими данными. Математически это верхняя граница так называемого «удивления», которое отражает то, насколько странным или неожиданным является текущее состояние мира или внутренней части организма в его восприятии.

**Информационная энтропия** - это мера неопределенности, которая остаётся после того, как свободная энергия эффективно устанавливает верхний предел неожиданности.

**Принцип свободной энергии (ПСЭ)** , в отличие от статистической механики, относится к системам в неравновесном стабильном состоянии (НРСС). Это согласуется с термодинамикой, поскольку последнюю можно рассматривать как частный случай ПСЭ при выполнении определенных условий.

**Ограждение Маркова** в статистической сети — это наименьший набор узлов, который делает ограждённый узел условно независимым от всех остальных. Поведение ограждённого узла можно предсказать, зная только состояния этих ближайших узлов. То же самое и с внешними узлами: ограждённый узел бессмыслен для предсказания их поведения. Ограждение Маркова разделяет все важные для организма состояния на внешние и внутренние, сенсорные и активные.

**Активный вывод** — один из двух способов минимизации свободной энергии, заключающийся в адаптивных действиях, снижающих неопределенность или удивление относительно причин ввода данных.

**Генеративная модель** — это вероятностное отображение внешних причин на наблюдаемые организмом входные данные.

Статистические свойства марковских ограждений ответственны за процессы, которые оптимизируют обоснованность байесовской модели, что в конечном

итоге делает последнюю актуальной моделью внешних состояний. Как утверждают Хесп и соавторы,

мы можем описать универсум биологических систем как ограждения Маркова и их внутренние состояния, которые сами состоят из ограждений Маркова и их внутренних состояний [277, p. 201].

Ограждения Маркова способны масштабировать свою структуру. Одним из примеров этого является гипотеза о том, что некоторые органеллы (например, митохондрии и хлоропласты) эукариотических клеток раньше были прокариотическими клетками .

Эта многоуровневая структура ограждений Маркова формирует многомерное фазовое пространство — готовую к формализации метатеоретическую онтологию, которая, благодаря воображению авторов, получила название *вариационная нейроэтология*.

Сказанное выше касается онтологии. А теория<sup>14</sup> утверждает, что

для того, чтобы организм сопротивлялся распаду и сохранялся как адаптивная система, встроенная в более крупную систему, но при этом статистически независимая от неё, он должен воплощать собой вероятностную модель статистической взаимозависимости и закономерностей своего окружения [282, p. 2].

Биологические системы, будучи нелинейными и неравновесными, соотносятся со случайным динамическим аттрактором, который представляет собой набор часто и с высокой вероятностью повторяющихся состояний. Пространство состояний можно представить как ландшафт свободной энергии, в самых нижних точках которого обитают живые организмы. ПСЭ утверждает, что все биологические системы постоянно участвуют в минимизации своей вариационной свободной энергии. Для организма выжить — значит избежать неожиданности, которая соответствует термодинамической потенциальной энергии. Согласно ТЭС, стремление минимизировать неожиданность является результатом естественного отбора, который сам по себе является процессом минимизации свободной энергии.

<sup>14</sup>О различии теории и онтологии шла речь в 3.3.1.

Внутренние состояния организма кодируют распределение вероятностей по внешним состояниям. «Свободная энергия — это функционал (т. е., функция функции), который описывает распределение вероятностей, закодированное внутренними состояниями марковского ограждения» [282, р. 4].

Попытки оспорить концепцию ПСЭ как биологическую метатеорию, по моим данным, немногочисленны. Единственный случай, встретившийся мне, содержится в [286].

### 3.3.5 Вычислительный подход в биологии

Механика Ньютона может описать, объяснить и даже предсказать траекторию летящего камня, но она ни в коем случае не наука о минералах. Когда мы ожидаем, что биология достигнет высоких теоретических стандартов физики, мы думаем о ней как о «науке о жизни», которая, в идеале, должна быть способна вывести все детали органических, генетических, физиологических и эволюционных процессов из набора высоких теоретических принципов и формул. Но эти дедуктивно последовательные теории никогда не объясняют всех фактов в какой-либо эмпирической области. Скорее, они имеют дело с определенным родом фактами, наблюдаемыми в различных областях. Механика ничего не знает ни о химическом составе летящего камня, ни о его геологической истории, но она может предсказать события, которые могут произойти как с камнем, так и с пулей. Точно так же мы вряд ли сможем продвинуть «науку о жизни» из состояния того, что Кант назвал «историей», в отличие от «чистой науки», к желаемому состоянию дедуктивной последовательности. Но мы можем сделать это с наукой об открытых системах в неравновесных состояниях. Эти системы, помимо живых существ, могут включать в себя общества, астрономические объекты и даже сами научные теории. Такая дисциплина, основанная на прозрачной онтологии и оснащенная хорошо построенным формальным языком, может объяснять и предсказывать определенные виды фактов в этих областях, включая биологию. Но биология как таковая, вероятно, останется своего рода «историей» в терминах Канта, где различные формализованные теории хорошо делают свою работу, уживаясь с эмпирическими классификациями и качественными нарративами.

Чтобы «починить радио» жизни, биолог должен начать с первого уровня вычислений Дэвида Марра<sup>15</sup>: то есть определить общую цель функционирования системы. Это может быть определенное сочетание адаптации [281], гомеостаза [280; 287] и энергоэффективности [177]. Т. е., на этом этапе у нас есть набор управляющих переменных. Затем биолог должен разработать комплекс алгоритмов, наиболее подходящих для достижения целей, определенных на первом уровне. Внутри алгоритмов должны быть идентифицированы так называемые вычислительные примитивы [171], т. е., элементарные узлы, комбинации которых образуют основные схемы частей живого тела, целых организмов, их симбиотических и социальных комбинаций. И, наконец, наблюдаемые биологические сущности должны быть сопоставлены с этими примитивами, чтобы алгоритмические шаги последних «кодировали» (а не просто «имитировали») причинные связи первых.

Принимая во внимание принцип множественной реализации, мы можем ожидать появления множества таким образом построенных конкурирующих (и, несомненно, вычислительных) теорий. И от этого наука о жизни только выиграет.

### 3.3.6 Вывод из раздела 3.3

История попыток математизации биологии показывает, что наиболее успешные примеры её связаны с изменением угла зрения на сам предмет этой науки. Мир органических молекул и живых организмов в результате недавних переосмыслений его онтологии представляется как многоуровневая структура вложенных ограждений Маркова — сетеподобных систем, осуществляющих вычислительную обработку информации на основе, скорее всего, статистических алгоритмов. Этот результат позволяет предположить, что когнитивные способности, в основе которых лежат те же механизмы, по сути являются продолжением и усложнением базисного устройства жизни, а биологические, когнитивные и социальные науки закономерным образом должны образовать единый комплекс знаний, связанный предметно и методологически.

---

<sup>15</sup>См. раздел 2.3.2.

### 3.4 Когниции в социальной среде

#### 3.4.1 Общество как на результат когнитивной эволюции

Согласно большинству теорий социологического и социально-философского мейнстрима, общественные отношения отличаются тем, что в их каузальную структуру встраиваются состояния сознания, которые или приводят онтологическую «подкладку» теории к полному индетерминизму, или требуют для собственного объяснения привлечения другой дисциплины с принципиально другой онтологией и теоретическим ядром — в данном случае, очевидно, психологии.

Как пишет Тодд Тремлин, характеризуя отличие «агентов» от «объектов»,

Агенты, с другой стороны, это существа, способные независимо и намеренно инициировать действия на основе внутренних ментальных состояний, таких как убеждения и желания. Наиболее очевидными интенциональными агентами являются животные и люди. Львы убивают антилоп, потому что они чувствуют голод. Женщины украшают свои тела, поскольку считают, что это сделает их более привлекательными [288, p. 76].

Но что значит «инициировать действия на основе внутренних ментальных состояний»? Согласно мнению Муиндара Сингха,

система является разумной, если для научных или интуитивных целей, чтобы охарактеризовать, понять, проанализировать или предсказать её поведение, необходимо приписать ей когнитивные определения, такие как интенции или убеждения [289, p. 1].

Это немного другой угол зрения — более эффективный, на мой взгляд, для получения адекватной онтологии общества. Не интенциональные признаки и состояния определяют некую сущность как «агента» (актера, субъекта), а, напротив, эти теоретические конструкты оказываются лишь «грамматическими» индикаторами сложных систем с линейно непредсказуемым поведением.

Те известные теории, которые пытаются увязать социум и сознание в единую каузально-онтологическую или концептуальную схему (Выготский, некоторые разновидности марксизма), делают это во многом декларативно, не предлагая работающего объяснительного механизма. Далее я намерен предложить концепцию, которая может вывести не только на удовлетворительную онтологию сознания, но и на удовлетворительную социальную онтологию. Предлагаемый подход обладает следующими преимуществами:

- (1) общество и сознание оказываются частями единой онтологии;
- (2) общество понимается как сеть, в каком-то смысле расширяющая возможности сети нейронов головного мозга, надстраивающаяся над ней и использующая её возможности;
- (3) к этой единой реальности применяются однотипные или схожие формализмы, связывающие теорию общества и теорию сознания в единый междисциплинарный проект с хорошими перспективами полной интеграции.

Следующие соображения могут помочь интегрировать предлагаемую концепцию в общенаучную картину мира. Эволюция, начавшаяся с усложнения структур органических молекул и соединения их в комплексы, прошла следующие основные переломные точки: выработка механизма генетического наследования, появление мозга, объединение биологических организмов в социальные организации и появление символических систем.

**Генетический механизм** сделал возможным производство многочисленных экземпляров одного вида. На первой стадии это имело смысл как производство избыточного биологического материала, с тем чтобы адаптироваться к среде через отбор случайно возникших полезных изменений.

**Появление мозга** сделало возможной эффективную адаптацию на уровне и в течение жизни отдельного индивида.

**Появление социальных организаций** изменило функциональную роль индивидов: теперь они уже не столько расходный материал, сколько субстанция единой сложной сети с эмерджентными свойствами.

**Начало использования символических систем** сделало социальные системы в собственном смысле когнитивными — наделёнными общей



памятью, а также другими разделяемыми ресурсами и функциями для обработки информации.

В основе социальной субстанции — механизм генетического наследования, производящий множество однотипных экземпляров. Эмерджентные свойства социальных сетей основаны на когнитивных способностях составляющих их агентов, которые (способности) позволяют выделять себе подобных среди других предметов окружающей среды. Первоначально это распознавание себе подобных необходимо для полового размножения. Поэтому социальные организации организмов, ориентированных на инстинктивное поведение, функционально бедны и ограничиваются, как правило, совместным использованием кормовой базы и собственно размножением, иногда совместной заботой о потомстве. Появление в мозге структур, ответственных за прижизненное обучение — у млекопитающих это неокортекс и «серое вещество» [290] — позволило социальным структурам значительно усложниться, создать социальную иерархию и в целом превратиться в мощное средство совместной адаптации. Создание и переход к использованию символических систем — сначала устных, затем письменных — ознаменовал дальнейшую эволюцию разделяемых когнитивных ресурсов, таких как социальная память. Кроме того, символические системы преобразили когнитивные способности индивидов, потеснив естественные механизмы статистических вероятностных выводов, действующие на основе нейросетевого обучения, в пользу линейно организованных логических умозаключений, основанных на иерархической семантике языков общения.

Таким образом, взгляд на общество как на результат когнитивной эволюции живых существ оказывается более фундаментальным, чем объяснение социальности, исходящее, например, из разделения труда. Во-первых, последнее представляется слишком антропоцентричным, упускающим из вида социальные организации других животных видов. Во-вторых, чтобы додуматься до какого-либо разделения труда, живые организмы должны, с одной стороны, уже заниматься трудом, а с другой — уметь «мыслить социально», т. е. понимать возможность использования ресурсов себе подобных. Напротив, мы обнаруживаем относительно развитые иерархические социальные организации у животных, не занимающихся систематическим производством и не использующих орудий труда и, вместе с тем, развивших нетривиальные сигнальные системы и демонстрирующих сложное совместное поведение. Эти факты поз-

воляют предположить, что *главный смысл социальности — в совместном доступе к разделяемым когнитивным (т. е., вычислительным) ресурсам.*

### 3.4.2 Общество как подсистема когнитивного механизма

При исследовании вопроса о соотношении когнитивного и социального в психологической и философской литературе чаще обсуждались механизмы детерминации когнитивных механизмов социальными структурами. Прогресс в области когнитивных наук в наше время даёт возможность взглянуть на проблему с другой стороны: каким образом когнитивные структуры отдельного индивида влияют на формирование и функционирование социальных связей. Как пишет Рон Сан,

Если посмотреть на проблему с другой стороны: когнитивные науки могут служить базисом наук социальных во многом так же, как физика даёт основание химии, или квантовая механика даёт основание классической. Социальные, политические и культурные силы, хотя, возможно, и “эмерджентные” (как часто считают), действуют как на отдельные сознания, так и через них [245, p. 16].

Тезис о том, что общество является подсистемой некоего когнитивного аппарата, звучит на первый взгляд несколько экстравагантно, но следует оговориться, что я понимаю его не в онтологическом, а в логическом смысле. Очевидно, что если некто включён в социальные взаимодействия, то он/она не может не использовать когнитивные механизмы в этом контексте. Этот тезис можно назвать «слабым» принципом интеграции социального и когнитивного, в противоположность «сильному» принципу.

1. Слабый принцип: некоторые или все когнитивные способности человека необходимо участвуют в социальной жизни. Это значит, что без когнитивного аппарата человека не было бы общества. Но оно есть. На основании *modus tollens* получаем:

$$(1) \forall x(Soc(x) \implies Cog(x)): \text{если некий элемент включён в социальную систему, то он также участвует в когнитивных процессах.}$$

2. Сильный принцип: некоторые или все когнитивные способности человека определяются его социальной организацией. Отсюда следует, что без общества не было бы когнитивного аппарата человека:

$$(2) \forall x(Cog(x) \implies Soc(x)) \text{ (по закону контрпозиции).}$$

В рамках моего подхода я защищаю слабый принцип, поскольку, во-первых, его легче обосновать логически и эмпирически, а во-вторых, он в большей степени соответствует общепринятому пониманию эволюции как развития от простого к сложному. Действительно, можно указать на когнитивные свойства агентов, такие как чувство боли или способность следить за движущимся объектом, которые для своего осуществления не требуют социальной организации. Можно также назвать соответствующие им социально-когнитивные способности: например, распознавание в других чувства боли или состояния слежения за движением. Но вряд ли можно привести пример социального взаимодействия, не предполагающего участия когнитивных способностей, поскольку любое взаимодействие с другим агентом предполагает способность к восприятию, категоризации и коммуникации.

### 3.4.3 Мультиагентные системы

«Понять» не обязательно значит «сделать», но «сделать» непременно предполагает «понять». Если кому-то удалось создать устройство, поведение которого, в полном соответствии с тестом Тьюринга, а также «утиным тестом» поэта Райли, при любых условиях неотличимо от поведения сознательного существа, то вряд ли кто-то сможет убедительно доказать, что понимание принципов его работы не является пониманием сущности сознания. Аналогично, если кому-то удалось построить колонию роботов, компьютерных программ, беспроводных устройств или чего угодно, которая функционирует, самоорганизуется и развивается подобно нормальному обществу, это значит, что её создатель с максимальной степенью точности ответил на вопрос, что такое общество. В первом случае, очевидно, мы говорим об искусственном интеллекте. Во втором случае мы говорим... тоже об искусственном интеллекте, только распределённом. Потому что в основе мультиагентных систем (МАС) как одной из технологий ИИ, равно как и исторически предшествующих («искусственная жизнь») или

сопутствующих («искусственное общество») им форм, лежат математические и программно-вычислительные инструменты, аналогичные или смежные с теми, которые используются в «традиционном» искусственном интеллекте. И потому также, что на выходе в том и в другом случае получается поведение, которое можно назвать сложным и интеллектуальным.

Мультиагентные системы — это программные или программно-аппаратные комплексы, в которых определены агенты и их среда или окружение. Агенты, вне зависимости от их материальной или программной реализации, автономны, преследуют, как правило, эгоистические цели и взаимодействуют друг с другом по простым настраиваемым правилам. Характерной особенностью таких систем является отсутствие централизованного управления; результаты их функционирования сложны и эмерджентны — линейно не предсказуемы и качественно отличны от предшествующих состояний. Эта особенность роднит их с нейросетями — разработчики могут управлять ими с помощью настроек, но не знают доподлинно, что происходит «внутри» и что именно вызывает эмерджентные эффекты. Хотя, безусловно, МАС нейросетями не являются — это просто ещё одна модель саморегулируемой сложности.

МАС используются в совершенно различных, иногда совершенно не пересекающихся, подчас неожиданных, областях: таких как экономика [291; 292], приборостроение [293], техническая стандартизация [294] и даже управление беспроводными сетями [295]. В «Программе фундаментальных научных исследований академий наук на 2013 - 2020 годы» [296], подписанной председателем правительства в 2013 году, МАС упомянуты в следующих контекстах:

- развитие и функционирования энергетических систем
- управление многоуровневыми и распределенными динамическими системами
- управление знаниями и системами междисциплинарной природы
- системы автоматизации, CALS-технологии
- методологии макроэкономических измерений.

В Европейском союзе действует European Social Simulation Association (ESSA), которая имеет партнёрские отношения с Computational Social Science Society of the Americas (CSSSA) и с Pan-Asian Association for Agent-based Approach in Social Systems Sciences (PAAA).

Мультиагентные технологии появились в конце 1980-х годов на стыке достижений в области объектного программирования, параллельных вычислений,

искусственного интеллекта (а также «искусственной жизни» и «искусственного общества»), интернет-технологий и телекоммуникаций. Это направление изначально называли *bio-inspired* («вдохновленное биологией»). Впоследствии оно выросло в новый класс интеллектуальных систем так называемого эмерджентного интеллекта.

Важным понятием в области распределённых интеллектуальных сетей является *swarm intelligence* («интеллект роя»), которое предполагает, что интеллект в таких системах не содержится в каких-то специальных компонентах, а рождается в ходе тонких взаимодействий совершенно простых, но автономных программ (агентов), сами по себе не имеющих развитых интеллектуальных способностей.

В популярной, но очень интересной лекции В.Л. Макарова [297] изложены основные принципы конструирования «искусственных обществ» методами МАС, хотя автор предпочитает термин «агент-ориентированные модели».

По мнению Макарова, начало этому научному направлению положили исследования в области «искусственной жизни» (AL — *artificial life*): ещё докомпьютерные системы клеточных автоматов, а также знаменитая «сахарная модель» (*SUGARSCAPE*), которая была использована в некоторых зарубежных исследованиях, цитируемых мною далее. Во-первых, имитация сложной «жизни» достигается использованием простых, подчас примитивных, правил [Макаров 2009, 9]. Во-вторых, один и тот же эффект может достигаться разным набором правил: одна модель демонстрирует становление социальной иерархии, закладывая положительную обратную связь в предыдущем опыте доминирования или подчинения, другая показывает тот же эффект, заставляя агентов предпочитать «общество себе подобных» [297, с. 15]. Причём тонкой настройкой правил можно точно определять и при необходимости менять количество получаемых «каст». В-третьих, что очень важно, в более сложных моделях, имитирующих социальные сети, приходится опираться на гипотезу ограниченной рациональности агентов, поскольку полный объём информации о состоянии социальной сети, на основе которой должны приниматься решения, превосходит когнитивные способности отдельного агента — такие, как объём памяти [297, с. 18]. Это обстоятельство многое говорит о реальных обществах, поскольку очевидно, что они функционировали бы принципиально по-другому, если бы каждый индивид мог воспринять, обработать и запомнить весь объём информации, необходимой для принятия решений.

Одна идея, высказанная в лекции, — и это в-четвёртых — очень важна для построения социальной онтологии конкретно-научного уровня. Макаров утверждает, что, подобно тому, как физическая материя состоит из конечного набора частиц, социальная материя, чтобы быть научно объяснимой, также должна состоять из конечного числа элементов. Поэтому такими элементами не могут быть люди — это должны быть действия [297, с. 19]. Не могу не отметить, что эта мысль интересно коррелирует с концепцией, изложенной мною в [74]: философская метаонтология общества это теория элементов как «первых сущностей» по Аристотелю, тогда как конкретно-научные онтологии — это или эссенциалистские теории «объектов» как носителей свойств и отношений, или «сетевые» теории, в которых элементам приписываются только отношения, а «свойства» возникают как эмерджентные эффекты сетевых взаимодействий. Учитывая то, что общество есть динамическая система, отношения реализуются именно как действия.

И, наконец, в-пятых — это уже не принцип, а скорее эффект — агент-ориентированные системы, или «искусственные общества», преодолевают кризис в математике, который, по мнению Макарова, наступил в XX веке из-за ограниченности экономических моделей, где невозможно было, например, ввести в модель внеэкономические — скажем, моральные — мотивы предпринимателей [297, с. 22]. Поскольку технология МАС позволяет моделировать какое угодно социальное поведение, её можно рассматривать как средство преодоления этого кризиса. Итак, главной философской импликацией теории и практики МАС можно считать то соображение, что социальные эффекты «программируются» правилами, определяющими действия и взаимодействие рациональных агентов. При этом некоторый набор правил «программирует» определённый срез или аспект социальной жизни, в чём нетрудно увидеть почти полную аналогию с витгенштейновской концепцией «языковых игр»: повторяющиеся паттерны взаимодействий, основанные на «следовании правилу».

Какую пользу может извлечь философия из МАС? Мультиагентные системы представляют собой хорошую функциональную модель для экспликации понятий «интеллектуальный» и «когнитивный». При помощи этой технологии можно провести более или менее определённую границу между когнитивными способностями, обеспечиваемыми мозгом, и способностями, в основе которых лежит правилосообразное социальное взаимодействие, которые я назвал бы собственно интеллектуальными. Распространённый философский предрассудок

состоит в том, что переход от простых когнитивных к собственно интеллектуальным способностям обеспечивается более развитым мозгом. И, поскольку с интеллектом мы связываем способность оперировать символами и, в частности, производить вычисления на их основе, возникает представление о развитом мозге как вычислительной машине.

МАС и теория рефлексивного управления [298] показывают, что сложное интеллектуальное поведение может быть моделировано и объяснено как следствие взаимодействия многих агентов, подчиняющихся одним правилам и обладающим обратной связью. Тогда логично предположить, что когнитивно развитый мозг, встраиваясь в такую систему социальных взаимодействий, развивает дополнительный этаж — функциональный блок, обеспечивающий правилосообразные социальные взаимодействия — вместе со способностью манипулировать ими самостоятельно. Если в ходе этих взаимодействий используются символические системы (которые сами суть результат развития и усложнения правил социального взаимодействия), то мозг получает дополнительные инструменты, делающие его когнитивные возможности ещё более мощными и универсальными.

В литературе по искусственным обществам и МАС часто термины «когнитивный», «рациональный» и «интеллектуальный» употребляются как синонимы. Так, например, Ниази и Хуссейн полагают, что «агент был бы чем-то таким в системе, что может быть рассмотрено и идентифицировано экспертами в любой дисциплине и интересующем субдомене как играющее в этой системе важную, индивидуальную, интерактивную и интересную роль, приводящую к поведению, которое называют интеллектуальным, рациональным, когнитивным или эмерждентным» [299, р. 2–3]. Исследователям, практикующим в этой сфере, которую иначе называют когнитивными распределёнными вычислениями или комплексными адаптивными системами, строгое различие этих терминов может представляться не столь существенным для их задач. Однако философы, в чью миссию входит выработка концептуальных каркасов для построения «большой картины», должны уделить внимание этим различиям, тем более, что мультиагентный сетевой подход способен сделать некоторые аспекты проблемы чуть более заметными.

Итак, отвлекаясь изначально от любой доменной специфики, в реальном или искусственном мире мы имеем системы сетевого типа, состоящие из действующих элементов (агентов), которые (1) обладают внутренним аппаратом,

позволяющим воспринимать и запоминать данные, принимать на их основе решения. Эти агенты (2) подчиняются явно выразимым правилам, упорядочивающим их поведение. В результате их совокупной деятельности (3) система демонстрирует сложное адаптивное поведение, которое выглядит непредсказуемым в рамках линейного причинно-следственного видения, и, тем не менее, ведёт к решению её жизненно важных задач.



Вершины треугольника: 1 —  
нейропсихический процессор (мозг), 2 —  
правила поведения и взаимодействия  
агентов, 3 — сложное адаптивное поведение  
сообщества.

Рисунок 3.1 — Рациональное, когнитивное и интеллектуальное

В этой ситуации, призвав на помощь свою семантическую интуицию, я бы предположил (как показано на рис. 1), что сфера когнитивного располагается в пространстве между (1) и (2), создавая основу для использования возможностей внутренних процессоров агентов для поддержания и использования системы правил, определяющих их поведение в типичных социальных ситуациях. Сфера рационального включает все случаи, когда устойчивая система правил (2) применяется для обеспечения сложного адаптивного поведения (3) системы в целом и её отдельных агентов, находя в нём своё обоснование. И, наконец, всё то, что может быть определено как интеллектуальное, представляет собой «вычислительные» мощности (1), лежащие в основе стратегий, тактик и эвристик (3), обеспечивающих наилучшую адаптацию сообщества агентов к условиям среды. От параметров настройки системы зависит, обнаруживает ли интеллектуальное поведение только «рой» как целое или также и отдельные агенты. Очевидно, что все три термина обозначают подсистемы, которые в реальных системах неотделимы друг от друга и лишь выражают различные аспекты одной целостности.



В этом же контексте немного по-другому предстаёт и традиционная философская проблема интенциональности и интенциональных состояний. В феноменологии и аналитической философии интенция понимается как предметность актов сознания, их свойство «быть о чём-то». Нетрудно заметить, что в этом интенция акта сознания аналогична семантике знака (символа), который тоже «направлен на предмет». Но только ли аналогией исчерпывается это сходство, нет ли здесь существенной связи?

Если принять социальную модель интеллекта, то, несомненно, есть. Во-первых, для того, чтобы осознать свой сознательный акт как направленный на некоторый предмет, нужно уметь этот предмет назвать. Более того, нужно уметь назвать и сам сознательный акт. Если бы заяц умел говорить, он сказал бы: «Я боюсь волка» — и тогда он в явном виде представил интенцию своего страха. Не умея говорить, он просто убегает при виде волка, и, если даже его действие сопровождается ощущением страха, это не делает страх интенциональным актом его сознания.

Почему? Представим себе, что в рамках моделирования «искусственных эмоций» (существует и такое направление) мы обучаем некоторую нейросеть испытывать страх при виде волка. Мы последовательно показываем ей образы разных животных, но только в ответ на продуцирование ею страха перед волком мы даём положительную обратную связь. После некоторого количества итераций возникает функциональная зависимость «образ волка → эмоция страха». В каком смысле здесь можно говорить о какой-то «направленности» данной эмоции на данный образ? Скорее, она просто является его следствием.

Представим более сложный случай. Мы тренируем нейросеть искать морковку. Сначала мы учим её распознавать образ морковки среди многих смежных образов, потом связываем с этим образом желание, которое запускает программу поискового поведения. Тогда функциональная схема выглядит следующим образом: вызов желания → запуск поиска → положительная обратная связь при встрече с нужным образом → прекращение поиска. Конечно, мы можем при этом говорить, что наша нейросеть «ищет морковку», точно так же, как ваш текстовый редактор «ищет и исправляет» слова, но с чисто технической точки зрения она делает это не более, чем бильярдный шар «ищет» лунку.

Если некоторое техническое устройство успешно моделирует поведение зайца, у нас нет основания полагать, что соответствующие функциональные

подсистемы зайца устроены сложнее, чем это устройство — даже если заяц в целом устроен сложнее.

Таким образом, «направленность» акта сознания «на предмет» — не более, чем фигура речи, и, чтобы она была возможна, речь должна иметься в наличии. Не нужно, наверное, специально доказывать, что речь, язык — порождение системы социальных связей, технически несовершенный интерфейс между нейросетями индивидов. И тогда получается, что акт сознания оказывается «направленным на предмет», т.е. становится интенциональным, только будучи «проговоренным», поскольку такова грамматика нашей речи: сам акт обозначается сказуемым (предикатом), которое требует дополнения (объекта) — как бы обладает «идеальной» стрелкой, указывающей на объект. При этом психологическая реальность может быть устроена совершенно по-другому.

В рамках теории МАС М.Е. Сингх [289, р. 11] вводит понятие коммуникативных интенций, что философски двусмысленно: интенция понимается как намерение выразить некий смысл (тогда она эквивалентна имплицатуре, о которой речь пойдёт ниже) или как *aboutness* (направленность на предмет)? Я предлагаю понятие *коммуникативных модальностей* как более подходящее в данной проблемной ситуации, поскольку при помощи него знание, полагание, сомнение, опасение и т. п. могут быть формализованы средствами, например, эпистемической логики, интерпретированной на МАС. Это модальности, описывающие специфические логические структуры различных коммуникативных ситуаций.

Ещё одна традиционная философская тема, которая, на мой взгляд, является бенефициаром исследований в области МАС, это проблема понимания сути человеческой (и не только) социальности. Философы традиционной школы, признавая наличие устойчивых и структурированных, основанных на коммуникации сообществ у некоторых животных, тем не менее полагают, что социальность как таковая каким-то существенным образом связана с человеком, является его видовой особенностью, появившейся в результате некоторого «скачка» в эволюционном развитии, а то и дарованной свыше. При этом в качестве существенных отличий, отделяющих человеческий мир от животного царства, указывают то на «идеальный план», имеющийся в голове самого плохого архитектора в отличие от самой лучшей пчелы [300, с. 189], то на способность к моральному суждению и различению добра и зла.

Вряд ли можно представить себе какой бы то ни было социум, состоящий из членов, лишённых когнитивных способностей. Члены социума должны идентифицировать себя и друг друга как таковых, воспринимать и правильно истолковывать сигналы коммуникации, знать и хранить в памяти основные паттерны социального поведения. Для этого они должны обладать перцептивным аппаратом, блоком обработки информации, памятью, системой принятия решений. Т. е., их физическая способность быть когнитивным, или рациональным, агентом выступает в качестве необходимого условия их социальности. Возникает естественный вопрос: что является её достаточным условием? Технология МАС подсказывает ответ на него — простые правила, основанные на обратной связи: если твой сосед делает *A*, делай *B*, и т. п. Очевидно, что способность действовать по правилу также предполагает наличие развитой системы когнитивных способностей.

Следующий закономерный вопрос: как оказывается возможной саморегулируемая система автономных агентов, основанная на простых правилах, в которых не содержится описание её конечных или желаемых состояний? Мозг это самообучаемая нейросеть, и люди уже научились моделировать некоторые аспекты его работы в технических устройствах. И более или менее понятно, как эта природная нейросеть развивалась эволюционно, давая преимущества особям и видам, наиболее эффективно использующим способность к обучению в целях адаптации. Общество отличается от мозга в одном важном отношении: между его членами нет физических соединений, подобных синапсам нейронов, вся сеть построена на действиях и правилосообразных реакциях на них. Более того, члены социума не привязаны жёстко к своим соседям в пространстве, а могут менять местоположение, взаимодействовать с новыми соседями, меняя физическую конфигурацию сети на уровне индивидов. И теме не менее, на уровне типов или ролей социальная система остаётся, как правило, стабильной и идентифицируется именно благодаря этой стабильности.

Таким образом, в отличие от мозга, протяжённого в пространстве, социальная система протяжена во времени и сохраняет стабильность благодаря памяти и другим когнитивным способностям её элементов. Отсюда возникает наивно-психологическое убеждение, что социальные связи существуют «в головах». Правильнее было бы сказать, что они существуют во временном измерении динамической системы, а «голова» выступают в качестве её необхо-

димого условия. Социум оказывается своего рода производной от когнитивных способностей составляющих его индивидов в рамках динамической системы.

Чтобы предложить правдоподобный ответ на вопрос об эволюционной оправданности социальных систем, нужно сначала согласиться, что эволюция возможна, если и только если существует механизм, закрепляющий случайно найденные преимущества. Если в качестве такого механизма выступают гены, то индивиды с полезной мутацией статистически чаще выживают и дают потомство в условиях, в которых другие гибнут. Если в качестве такого механизма выступает мозг, то к генетическим преимуществам индивида добавляется способность к обучению, которая увеличивает выживаемость на индивидуальном уровне. Может быть, в этом случае эволюция способствует увеличению объёма мозга в определённых условиях или изменению его структуры путём изменения соотношения объёма и роли мезенцефалона, ответственного за инстинктивное наследуемое поведение, и коры, обеспечивающей прижизненное обучение [290, р. 7], в пользу последней.

Однако паттерны поведения, удачно найденные в процессе обучения и адаптивно полезные, не передаются от особи к особи, от поколения к поколению, поскольку не закрепляются в генах. В то же время, очевидно, что популяции, обладающие такой возможностью, получили бы неоспоримое эволюционное преимущество, поскольку скорость накопления полезных изменений возросла бы экспоненциально. Для реализации этого преимущества необходимо два базовых условия: во-первых, особи должны уметь идентифицировать друг друга как члены одной общности. Можно предположить, что такая идентификация изначально основывалась на механизме полового размножения, когда особи противоположного пола рассматриваются как объект влечения, а особи того же пола — как соперники. В любом случае, и те, и другие выделяются из окружающей среды и представителей других видов. Во-вторых, должен появиться какой-либо механизм обратной связи, заставляющий определённым образом реагировать на действия соплеменников. Наиболее примитивный и, видимо, исторически первый такой механизм — это подражание: ты ищешь пищу, когда и где её ищут твои сородичи, и убегаешь, когда они убегают. При этом механизмы обоих условий социального поведения — узнавания своих и подражания им — вполне могут передаваться генетически (зеркальные нейроны), тогда как само социальное поведение эмерджентно возникает в каждой

конкретной ситуации, давая, тем не менее, существенные адаптивные преимущества данной популяции.

И, поскольку адаптивные преимущества приобретены, можно предположить, что срок жизни особей в этой прогрессивной популяции увеличивается, позволяя одновременно существовать не двум, а трём поколениям. В недавнем исследовании, осуществлённом, кстати, с помощью компьютерного мультиагентного моделирования сообщества обезьян, была продемонстрирована роль «института бабушки» не только в прогрессивном увеличении продолжительности жизни членов сообщества, но и в укреплении их социальных связей. Авторы пишут: «Как показывает наша модель, даже отбор только по признаку участия бабушек в уходе за детьми мог бы привести к эволюции нашего постменопаузального долголетия, усилению взаимозависимостей и созданию социального контекста для многих других функций, которые впоследствии развились в рамках нашего вида» [301, р. 5]. Одновременное сосуществование трёх поколений не только дальше увеличивает продолжительность жизни за счёт лучшего ухода за детьми, но и способствует созданию подсистемы социального наследования, в рамках которой результаты индивидуального научения систематически передаются другим особям без участия генного механизма. Кстати, этот факт подтверждается и исследованиями на мультиагентных моделях. Как показывают результаты группы голландских исследователей, «эволюционный подход [к обучению агентов] способен поддерживать более крупные и более стабильные популяции агентов, а также поддерживать более высокий уровень индивидуального успеха, по сравнению с обучением на протяжении жизни [одного агента]» [302, р. 155]. Лишь два эволюционных шага отделяют такое общество от собственно человеческого: появление коммуникации с использованием символических систем и появление письменности как расширения системы социального наследования.

Как мы уже увидели, наука о сложных агент-ориентированных системах способна внести существенный вклад в — по крайней мере — прояснение, если не решение, некоторых фундаментальных философских проблем. Вместе с тем, современные исследования, связанные с мультиагентным моделированием, не только методологически опираются на определённые философские школы, но и инкорпорируют отдельные концепции, развитые в рамках этих философских направлений, в свои теоретические основы.

Так, широкое использование модальных (эпистемических) логик в конструировании МАС помогает лучше понять социально-коммуникативную природу знания. Автор комплексного руководства по МАС Майкл Вулдридж предлагает формализм, представляющий собой традиционную пропозициональную логику первого порядка с добавлением множества одноместных модальных операторов  $K_i$ , которые читаются как «агент  $i$  знает, что...» [303, р. 279]. Он показывает далее, что при интерпретации на МАС эта логика сталкивается с некоторыми семантическими проблемами. Например, при негарантированной надёжности коммуникации такая важная эпистемическая позиция как «общее знание» (common knowledge) — «все знают, что  $p$ , и все знают, что все это знают» — может оказаться недостижимой, потерявшись в бесконечных итерациях взаимных подтверждений. Аналогичным образом может оказаться проблематичным распределённое знание (distributed knowledge), когда агент  $i$  знает, что  $A$ , а агент  $k$  знает, что  $A \implies B$ . В этом случае в системе в распределённом виде содержится знание, что  $B$ . Для формализации такого знания Вулдридж предлагает специальный эпистемический оператор  $D$ , семантическое определение которого предполагает не объединение, как в традиционных модальных логиках, а пересечение эпистемических миров  $w_i$  всех агентов в системе. По его мнению, «ограничения, накладываемые на возможные миры, в общем случае означают увеличение знания» [Wooldridge 2002, 283]. Не менее важно, с его точки зрения, то обстоятельство, что семантика возможных миров, которая обычно применяется для интерпретации эпистемических и других модальных логик, предполагает «совершенного (идеального)» агента на месте рассуждающего агента, который, например, увидит противоречие между  $A$  и  $\neg B$ , если известно, что  $A \implies B$ . Однако известно, что реальные агенты — включая большинство живых людей — часто весьма толерантны к таким неявным противоречиям, что не мешает им жить и функционировать. Поэтому Вулдридж предлагает ограничиться в МАС требованиями слабой непротиворечивости, которое запрещало бы только явные противоречия  $A$  и  $\neg A$  [303, р. 276].

Как мы понимаем, все эти аргументы могут быть приняты в поддержку тезиса ограниченной рациональности агентов, который обсуждался здесь ранее.

Для дальнейшего пояснения этого принципа обратимся к интересной формализации, которая предложена в другом руководстве по МАС [304]. Анализируя понятие общего знания на примере известной логической задачи про трёх игроков в шляпах, каждый из которых видит только цвет шляп двух других, Ни-

кос Влассис вводит следующее определение. Пусть  $S$  — множество всех вообще возможных состояний, из которых только  $s$  — актуальное («истинное» в его терминологии) состояние. Пусть также  $i$  — порядковый номер агента из некоторого конечного множества. Каждый агент  $i$  видит состояние  $s$  через «информационную функцию»  $P_i(s)$ , которая порождает подмножество  $S$ , включающее только те состояния, которые данный агент — учитывая ограниченность имеющейся у него информации — считает возможными. Пусть также  $E$  есть некоторое подмножество  $S$ , которое мы будем называть событием.  $K_i$  — оператор «знания», относимый к определённому агенту  $i$ . Тогда, по определению:

$$K_i(E) = \{s \in S : P_i(s) \subseteq E\} \text{ [304, р. 39],}$$

т.е., агент  $i$  знает  $E$ , если его информационная функция  $P_i$  в истинном состоянии  $s$  содержит  $E$ . Говоря естественным языком, некоторый агент знает некоторое событие, если все состояния, которые из его перспективы представляются возможными, содержат это событие.

Как считает Влассис, предложенное им определение соответствует тому, которое используется в эпистемической логике. Там считается, что агент знает факт  $\varphi$ , если  $\varphi$  имеет место во всех состояниях, которые агент рассматривает как возможные. В рамках событийного подхода агент знает событие  $E$ , если все состояния, которые агент считает возможными, содержат  $E$ . Влассис ссылается на основополагающую<sup>16</sup> работу [305], где показано, что оба подхода — логический и событийный — эквивалентны [304, р. 39].

Применение описанного формализма к задаче про игроков в шляпах хорошо иллюстрирует принцип обратного отношения знания и возможных миров, сформулированный Вулдриджем: каждое добавленное знание уменьшает число возможных состояний в перспективе агента. Однако принцип ограниченной рациональности относится не только к количеству знания, имеющегося у агента, но и к качеству его рассуждений. В модели Влассиса все игроки в шляпах являются «идеальными логиками» в терминах Вулдриджа: они мастерски и последовательно делают выводы, применяя законы непротиворечия и исключённого третьего. В то же время очевидно, что реальные агенты, поведение которых моделируется в МАС, таковыми являются далеко не всегда. Поэтому в теоретические основы распределённого искусственного интеллекта, как ещё называют обсуждаемое нами научное направление, входят и нечёткие логики,

<sup>16</sup>На неё ссылается также и Вулдридж.

и статистические теории, и некоторые другие теоретические средства, позволяющие приблизить модели к нашей непростой реальности.

О значении для МАС такого философского изобретения как теория речевых актов, начало которой положили работы Остина [306] и Сёрла [307], пишут и Сингх [289, р. 126–130], и Вулдридж [303, р. 164–168]. Общая идея состоит в том, что коммуницирующие агенты (не важно, люди или автоматы), как правило, с помощью своих сообщений хотят достичь некоторых целей, которые состоят в изменении ситуации — реальной физической или, по крайней мере, эпистемической. Согласно Остину, речевые (локутивные) акты обладают иллокутивной силой, которая состоит в намерении изменить ситуацию, а также набором перлокуций — возможных последствий совершения речевого акта.

Так, Сингх вводит в свой формализм, описывающий коммуникацию внутри МАС, специальный оператор  $W$  для обозначения состояния «подлинной выполненности» речевого акта и даёт несколько семантических определений этого состояния для речевых актов различной иллокутивной силы:

- *Ассертив* (утверждение) выполнен, если только предложение, его выражающее, истинно в момент произнесения высказывания.
- *Директив* (указание) выполнен только в том случае, если (а) его предложение становится истинным в некоторый момент в будущем относительно его произнесения, и (б) на протяжении всего сценария у слушателя есть ноу-хау, а также намерение этого достичь.
- *Комиссив* (обещание, угроза) выполнен только в том случае, если (а) его предложение становится истинным в некоторый момент в будущем относительно его произнесения, и (б) на протяжении всего сценария у говорящего есть ноу-хау, а также намерение добиться этого. В отличие от директива, условие выполненности комиссива зависит от действий, намерений и ноу-хау только одного агента.
- *Пермиссив* (разрешение) выполнен в сценарии, в котором слушатель выполняет хотя бы одно действие, выполнение которого может привести к состоянию, в котором он не может предотвратить наступление условия, выраженного в произнесённом предложении.
- *Прохибитив* (запрет) выполнен в некотором сценарии в некоторый момент, только если ни одно из действий, совершаемых слушателем по этому сценарию (в будущем), не может привести к состоянию, в кото-



ром слушатель не сможет предотвратить возникновение выраженного в высказывании условия.

- *Декларатив* (заявление, объявление) выполняется только в том случае, если (а) его предложение становится истинным в первый раз в тот момент, когда оно произносится, и (б) все время, пока оратор произносит, он предполагает, что выраженное в нём условие должно наступить, и знает, как сделать так, чтобы это произошло. [289, р. 131–135]

Инженерная проблема для проектировщиков МАС состоит в том, чтобы создать систему, которая позволит реализовать только корректные сценарии. По мнению автора, предложенные им определения дают объективные критерии, позволяющие оценить корректность различных сценариев для мультиагентной системы. Требования, предъявляемые к агентам на основе его формализации теории речевых актов, предписывают им, как коммуницировать, учитывая их убеждения и намерения, таким образом, чтобы могли исполниться только корректные сценарии. Так, семантические определения оператора  $W$  могут быть использованы для обоснования некоторых условий корректности в МАС. Например, сценарий может быть определен как корректный, если все переданные в его рамках сообщения выполнены.

То же можно сказать и о теории импликатур, разработанной философом Г.П. Грайсом [308], которая занимает важное место в методологии МАС. Под импликатурой понимается непрямой смысл высказывания, который не сводится к смыслу составляющих высказывание слов. Например, фраза «Здесь холодно» может означать просьбу закрыть окно. Грайс различал конвенциональную и неконвенциональную импликатуру. Первая относится к случаю, когда то, что подразумевается, определяется конвенциональным значением используемых слов. Во втором случае импликатура следует не непосредственно из конвенционального значения слов, а из контекста или из структуры разговора.

В разговорной импликатуре подразумеваемый смысл основан на том факте, что слушатель предполагает, что говорящий следует максима́м Грайса: принципам уместности и истинности. Для того, чтобы импликатура имела место в диалоге, (1) слушатель должен предполагать, что говорящий следует указанным максима́м, (2) это предположение необходимо для того, чтобы слушатель понял смысл высказывания, содержащего импликатуру, и (3) тот факт, что слушатель может сделать то, что от него ожидает говорящий, должен быть общим знанием. Очевидно, что перечисленные требования хорошо поддаются

формализации, например, семантическим инструментарием Сингха и, таким образом, могут быть имплементированы в МАС на инженерном уровне.

Мультиагентные системы как искусственные квазисоциальные образования демонстрируют в готовом виде то, к чему человечество шло долгим эволюционным путём. Так, ограниченная рациональность агентов соответствует когнитивной ограниченности людей и других социальных животных. Только в последнем случае это объясняется тем обстоятельством, что животные — это мобильные устройства, которые не могут быть подключены к сети питания, и вынуждены экономить энергию своих «аккумуляторов», периодически подзаряжаясь. Поэтому природа должна была создать энергетически экономные когнитивные решения: перцептивные механизмы, дающие минимум информации, достаточный для адаптации и выживания; рациональность, сводящую все объяснения к минимально возможному числу принципов и сущностей; распределённый интеллект, который даёт возможность каждому компенсировать собственную когнитивную ограниченность за счёт привлечения ресурсов социальной сети. В этом последнем и состоит суть социальности биологических организмов, включая людей. Люди, правда, могут похвастаться ещё символическими системами и основанными на них механизмами социального наследования (литература, образование и др.), что, конечно, на порядок повышает эффективность распределённого интеллекта.

Успех МАС как научного направления поддерживает метасетевую теорию сознания (МСТ), согласно которой сознание является эмерджентным эффектом не только сложной сети нейронов головного мозга, но и социальной метасети. МАС — один из путей научной демистификации «жизненного мира» человека: когнитивного и социального.

#### 3.4.4 Взгляд на нейроэтику

В рамках кантианской философии морали правила этики понимались как трансцендентальные и имеющие сверхприродный характер. Этим требованиям нужно свободно и достойно подчиняться, даже если они противоречат интересам и задачам выживания индивида и рода. Все попытки причинного обоснования морали не могли объяснить настоятельности её требований и их

всеобщий, безусловный характер. Казалось, что «нравственный закон внутри нас» доказывает нашу причастность миру ноуменов, миру абсолютной свободы.

Этот философский подход был уместен, пока наука имела дело с миром, жестко и линейно детерминированным, в котором предмет научного знания – природа – исключает возможность разрыва причинно-следственных цепочек, а непризнание свободы, с другой стороны, логически влечёт за собой отрицание ответственности – краеугольного камня социальной управляемости. Такая ситуация предполагала удвоение мира на природу и мир сверхчувственных ноуменов и разделение объектного знания на науку и метафизику.

Ситуация изменилась с появлением, с одной стороны, научных дисциплин, не основывающихся на классическом детерминизме, предполагающих разрыв причинных последовательностей в виде принципа неопределённости, «точек сингулярности» и т.п., а с другой – технологий, позволяющих видеть происходящее в голове в реальном времени. Вследствие этого, мы, с одной стороны, теперь уже не противопоставляем свободу и необходимость в форме абсолютной дихотомии, а с другой – не можем настаивать на абсолютной автономности индивида, поскольку видим реально управляющие им механизмы.

Как следствие, классическая парадигма философии морали – свободный и ответственный индивид, ограничивающий себя правилами при общении с себе подобными – не теряя философской привлекательности, всё менее уживается с новыми научными данными.

Получив техническую поддержку, стала бурно развиваться нейронаука, идущая рука об руку с когнитивной наукой: компьютерные модели последней стало легче интерпретировать на значительно расширившемся эмпирическом материале. Прогресс в этой научной области вдохновил появление нейрофилософии – дисциплины, переосмысливающей и переформулирующей традиционные метафизические проблемы сознания в терминах, соизмеримых с данными нейрофизиологии. Логичным выглядит и появление нейроэтики, поскольку трансцендентальный язык классической философии морали стало возможно перевести на язык научного эксперимента.

Нейрофилософия – по крайней мере, в версии Патриции и Пола Чёрчлендов и близких по взглядам авторов – формировалась в тот период, когда в когнитивной науке бурно развивался, внушая определённые надежды, коннекционистский подход, который рассматривает нейронную сеть как вычислительную архитектуру, альтернативную тьюринговой. Машина Тьюринга

основана на интерпретации и преобразовании линейной последовательности символов в соответствии с формальными правилами. Коннекционистская сеть, напротив, осуществляет параллельную обработку сложных входящих сигналов путём изменения «весов» межнейронных связей, создавая устойчивые паттерны возбуждений участков сети, которые интерпретируются как репрезентации ментальных состояний. Видение нейросети как ещё одной вычислительной архитектуры лежит в основе нейрокомпьютинга – направления в информационной науке, строящего вычисления на параллельной работе огромного количества мелких процессоров, соединённых в сеть. Нейрокомпьютер не нуждается – или нуждается в значительно меньшей степени – во внешнем программировании и способен к самообучению. Он гораздо эффективнее машины Тьюринга в обработке больших массивов «аналоговой» информации – распознавании образов или символов. Однако представляется, что последних прорывов в нейро- и когнитивных науках всё же недостаточно для того, чтобы справиться с традиционными философско-этическими вопросами. Кант был прав в том отношении, что моральные отношения возникают в пространстве между индивидами, а не в церебральных сетях. Поэтому подлинный прогресс нейроэтики можно ожидать в области междисциплинарного сотрудничества когнитивных и социальных наук. Но для этого те и другие должны говорить примерно на одном языке.

Надежду вселяет, в частности, то, что параллельно внедрению компьютерных метафор в науку о мозге аналогичный процесс шёл в социальных науках. Те же или подобные математические инструменты применяются в социологии социальных сетей. В рамках компьютерных симуляций деятельности рациональных агентов большое распространение получили мультиагентные системы. Сформировалось целое научное направление, постепенно входящее в социологический мейнстрим – компьютерная социология, имеющая прямое отношение к так называемым «искусственному обществу» (*Artificial Society*) и «искусственной жизни» (*Artificial Life*). Одна из задач настоящего исследования — поиск и анализ теоретических, эмпирических и квазиэмпирических (на уровне компьютерных моделей) результатов конкретных наук об обществе и сознании, которые помогли бы сформировать методологическую основу для междисциплинарного взаимодействия этих наук при решении проблем, традиционно относимых к философской этике.

В нашу эпоху, когда стремительный технический и методологический прогресс когнитивных и социальных наук позволяет им подбираться к тра-

диционным «трудным» философским проблемам, в том числе этическим, эффективное управление самыми разными сферами общественной жизни нуждается в новом теоретическом обеспечении, которое возможно на основе новых форм интеграции философии и конкретных наук о человеке.

\* \* \*

Теории, лежащие в основе мультиагентных систем, имеют важное значение для коррекции некоторых традиционных философских идей. Так, многократно критиковавшейся концепции знания как обоснованного истинного убеждения (ЖТВ) можно противопоставить понимание, в значительной степени навеянное эпистемической логикой и хорошо сочетающееся с принципом коммуникативного функционализма (см. раздел 1.3.2): агент *убеждён* в некотором положении вещей, если оно содержится во всех состояниях (возможных мирах), которые допустимы из перспективы этого агента. Назовём это множество состояний *эпистемическим миром* агента. Однако, состав допустимых миров из перспективы агента может отличаться от пересечения эпистемических миров агентов, участвующих в коммуникации. Назовём последнее *распределённым знанием* системы. Тогда, агент *знает* некое положение дел, если оно содержится во всех состояниях его эпистемического мира и в распределённом знании.

Эта концепция, во-первых, предлагает более ясное формальное различие знания и убеждения и, во-вторых, учитывает различные социокультурные практики «сертификации» знания, которые не всегда предполагают его обоснование в привычном нам смысле. Недостатком этой концепции является то обстоятельство, что знание социологизируется: агент знает некий факт, если он убеждён в нём, и это убеждение разделяется его сообществом. Однако, если мы исходим из того, что социальные структуры имеют цель наилучшей адаптации к среде обитания и достигают этой цели путём обучения, то такое понимание знания кажется вполне естественным.

### 3.4.5 Вывод из раздела 3.4

Социальная организация возникает и существует как система разделяемого (shared) доступа индивидов к когнитивным ресурсам друг друга, а также

общим (распределённым) когнитивным ресурсам. **Мультиагентные системы как разновидность ИИ моделируют социальные взаимодействия, основанные на правилах, и распределённый (distributed) интеллект, основанный на эпистемических логиках. Мультиагентная система, состоящая из обучаемых нейросетевых агентов, создающих и использующих символический язык для коммуникации, была бы наиболее реалистической моделью человеческого общества.**

### 3.5 Люди как (ир)рациональные агенты

#### 3.5.1 Мыслим ли мы рационально?

То, какие когнитивные устройства мы, люди, собой представляем, и как эти устройства работают, поможет понять научно-фантастическая притча, рассказанная в статье [309], цит. по [30]. Представим себе, что команда океанографов снарядила батискаф, автоматически управляемый с плавучего судна, для исследования затонувшего «Титаника». Однако в «Титанике» поселилось злобное и очень умное морское чудовище, которое сразу же переподключило кабели, идущие от батискафа к пульту управления, таким образом, что акустические сигналы стали подаваться на видеомониторы, и наоборот. Учёные сидят в растерянности перед ставшими бессмысленными показаниями приборов, и всё, что им остаётся — это искать регулярности в непонятных мерцаниях и шипениях, сравнивать их между собой и предлагать правдоподобные гипотезы, объясняющие эти регулярности.

То, что чувства нас обманывают, открыли ещё древнегреческие философы, предложив Логос в качестве панацеи или альтернативного способа созерцания истины. Дальнейшая история философии, в анамнезе которой столько различных картин мира, сколько философов, показала, что разум нас обманывает тоже. Поэтому, анализ и сопоставление потоков ненадёжной информации (эмпирического и рационального) с целью выявления закономерностей и порождения правдоподобных гипотез есть *modus operandi* как отдельного индивида, так и науки в целом.

Как показывают современные исследования по поведенческой экономике и восприятию, проблемы рациональности встают уже на уровне перцепции. Как считают Тешо Фелин с соавторами, «<a>ргументы о восприятии непреднамеренно вплетаются в литературу о рациональности через использование визуальных иллюзий, метафор и задач в качестве примеров предвзятости, ограниченности и слепоты» [310]. Когнитивная наука как бы заново открывает аргументы Парменида против «мира во мнении». И если в исследованиях по поведенческой экономике молча предполагается, что есть некая единственная объективная реальность и некая единственная соответствующая ей рациональность, а иллюзии и предвзятости восприятия и суждения трактуются как девиации, то более, на мой взгляд, философски корректный подход подсказывает нам, что ни у одного из имеющихся у нас «окон в мир» нет эпистемических привилегий. И поэтому, в условиях недостатка информации мы выбираем одну из возможных рациональных интерпретаций реальности в надежде, что именно она позволит нам выжить и адаптироваться.

Как представляется, Фелин с соавторами претендует на своего рода кантианско-коперниканский переворот в науке: «То, что мы отстаиваем, — это принципиально иной взгляд на когницию. Осознание и восприятие — это, скорее, функция воспринимающего (а не воспринимаемых данных — И.М.), вопросов, проб и теорий, которые любой из нас навязывает даже самым простым визуальным сценам или окружению, или же реальности в целом» [310]. Так, авторы обсуждают широко известный эксперимент с «невидимой гориллой» [311]. В кадре — две команды, игроки которых одеты, соответственно, в белые и чёрные майки. Они параллельно передают волейбольные мячи, каждый — игрокам своей команды. Испытуемым — зрителям — даётся задание посчитать, сколько раз передают мяч игроки в белых майках. Поскольку задание несложное, практически все дают правильный ответ. Однако только половина испытуемых замечает, что посреди игры на поле выходит человек, переодетый в гориллу, бьёт по-обезьяньи себя в грудь и покидает кадр. И никто из испытуемых не обращает внимания на то, что в этот же момент один из игроков в чёрной майке покидает игру, а занавес на фоне меняет цвет с красного на золотистый.

С точки зрения более традиционного, «реалистического» подхода, этот эффект говорит о некоей неполноценности нашего восприятия, неспособного отразить подлинную реальность. Однако, как справедливо указывают авторы цитируемой статьи, «подлинная» реальность состоит из потенциально беско-

нечного числа свойств и отношений (цвет волос и расовый состав игроков, скорость их движения, буквы на занавесе и т.д.), которые в сумме так или иначе недоступны никакому восприятию, кроме божественного (а в последнем случае можно ли вообще говорить о «восприятии»?). Поэтому направленность и ограниченность реального восприятия реальных людей говорит именно о его рациональности, а не наоборот, — рациональности, которая в этом случае очевидно многовариантна.



Рисунок 3.2 — Иллюзия Понцо

С этой же — «кантианско-когнитивной» — точки зрения можно рассмотреть и иные известные в науке визуальные иллюзии. Так, иллюзия Понцо (рис. 2), как и иллюзия Эббингауза (рис. 3), демонстрирует «ложное» восприятие сравнительных размеров идентичных фигур, помещённых в переменный контекст.



Рисунок 3.3 — Иллюзия Эббингауза

Однако ложным это восприятие оказывается, если у нас есть альтернативный доступ к реальности, или альтернативный способ её репрезентации. Например, линейка. В культуре, где линейка является центром важной практики, более рациональным выглядит решение скорректировать с её помощью представление о размерах фигур — и то только в том случае, если возникнет прагматически оправданное подозрение. Кроме того, по крайней мере, в случае с иллюзией Понцо, «ложной» они оказываются только постольку, поскольку фигуры наложены на изображение, а не находятся в изображаемой реальности. Если бы это были не белые отрезки, а забытые на рельсах или потерянные



кем-то деревянные брусья, у нас и сомнений не возникло бы в разнице их размеров. И такая репрезентация была бы, несомненно, рациональной.

Из этого можно заключить, что наше перспективное видение — когда дальние предметы кажутся меньше — представляет собой эволюционно сформировавшуюся экономную, а следовательно, рациональную, форму визуальной репрезентации сложных пространственных отношений, достаточную для нашей ориентации и адаптации. Да, «обмануть её несложно», но именно потому что природа методом отбора вырабатывает наиболее экономные и энергетически эффективные способы представления данных, нужные для статистически достаточного выживания особей данного вида.

Для нашей темы принципиальную важность представляют комментарии, данные в руководстве по кросс-культурной психологии [312, р. 205–206]. Согласно данным кросс-культурных исследований, наиболее подвержены иллюзиям типа иллюзии Понцо представители «западной» культуры и жители больших городов, в отличие от обитателей «третьего мира» и сельской местности. Авторы приводят ряд гипотез, объясняющих эту разницу, которые все так или иначе сводятся к эмпирическому научению. Однако в последнем по времени переиздании этой книжки появилась ссылка на работу индийских врачей, которые нашли хирургический способ восстановить зрение у слепых от рождения детей. На удивление, излеченные дети, которым практически сразу же предъявлялась картинка Понцо, и у которых не было времени на культурное научение, регулярно находили верхний отрезок большим. Авторы указывают, что общепринятого объяснения этих новых данных пока нет. На мой взгляд, некоторый свет на

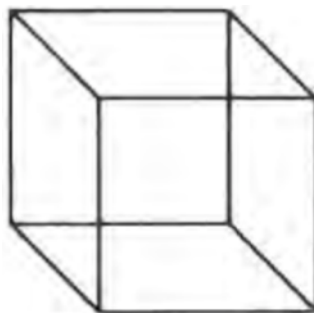


Рисунок 3.4 — Куб Неккера

возможные универсальные объяснения визуальных иллюзий может пролить замечание Дэвида Марра, классика вычислительной нейронауки, высказанное им в его основополагающей книге ‘Vision’ по поводу, пожалуй, самой старой и известной из визуальных иллюзий — куба Неккера (рис. 4). Это изображение

характерно тем, что зритель может свободно «переключаться» между двумя возможными аспектами: в одном из них ближняя к зрителю грань видится как расположенная ниже, а в другом — как расположенная выше дальней. С иллюзией Понцо его роднит то, что обе иллюзии возникают при двухмерной проекции трёхмерной картины. Соответственно, можно предположить, что это как-то связано с особенностями обработки объёмных изображений соответствующими нашими процессинговыми центрами. Марр предположил, что «часть объяснения его (куба — И.М.) перцептивного изменения должна быть связана с бистабильной нейронной сетью (то есть с двумя отчетливыми стабильными состояниями) где-то внутри мозга. . . » [172, р. 26]. Иными словами, архитектура многочисленных взаимосвязанных и соподчинённых нейронных сетей, образующих мозг, предполагает наличие многих одинаково стабильных состояний, в которых эти сети могут зафиксироваться, получив данные определённого рода. Получение дополнительных данных может переместить репрезентацию из одного стабильного состояния в другое, меняя, например, видимый аспект образа. Но система в целом настроена на работу в условиях недостаточных данных, когда одно из возможных и, в принципе, одинаково рациональных стабильных состояний выбирается случайно, или вероятность того или иного выбора подкреплена научением, обусловленным предшествующим опытом, культурным контекстом или физиологическими особенностями. С такой интерпретацией согласны Фелин и его соавторы, полагая, что «восприимчивость человека к праймингу и чувствительность к более заметным сигналам не являются очевидными доказательствами иррациональности, а скорее свидетельствуют об этой (когнитивной — И.М.) мультстабильности. Независимо от того, имеем ли мы дело с восприятием или рассуждением, в недостаточно информативных и неоднозначных ситуациях люди используют любые имеющиеся свидетельства или сигналы. . . для вынесения суждений» [310].

Таким образом, сравнивая психологические данные и предложенные психологические и нейрофизиологические интерпретации, мы можем с определённой вероятностью заключить, что устройство когнитивного аппарата мультстабильно и, следовательно, создаёт возможность множественных «рациональных» восприятий и суждений. Никто не предопределён фатально к определённому перцептивному стилю и определённому визуальным решениям в условиях неоднозначности данных, хотя некоторое влияние могут оказывать прошлый опыт или культурный контекст, повышая вероятность выбора тех или

иных форм репрезентаций по сравнению с другими. Но главный вывод состоит в том, что рациональность выбора восприятий, мыслей или решений принципиально вариативна и относительна.

### 3.5.2 Логично ли мышление?

Весьма распространённое и многими некритически принимаемое убеждение состоит в том, что логика является имманентной формой мышления и одновременно его описательной (а не нормативной) теорией, поскольку, мысля, мы следуем тем или иным непреложным правилам. А если таковые правила разнятся, то, следовательно, мы имеем дело с разными логиками. Чтобы уверенно утверждать это, необходимо показать, что реальное мышление людей подчиняется декларируемому в данной культуре логическим правилам.

Однако обратимся к известным экспериментам<sup>17</sup>. В эксперименте Уэйсона 1968 года испытуемым предлагался набор карт, на одной стороне которых были нанесены буквы, на другой — цифры. Было также сформулировано правило: если на одной стороне карты гласная буква, то на другой непременно будет чётная цифра. Испытуемым предлагались четыре лежащие карты, на «верхних» сторонах которых имелись символы, соответственно: А, К, 2, 7. В качестве задания нужно было ответить на вопрос: какие карты нужно перевернуть, чтобы убедиться в том, что правило выполняется. Согласно логическим правилам *modus ponens* и *modus tollens*, перевернуть нужно карты «А» и «7». Однако распределение (в %) реальных ответов реальных испытуемых выглядело следующим образом: «А» — 89, «К» — 16, «2» — 62, «7» — 25. Очевидно, что люди в массе своей руководствовались не декларируемыми в их культуре правилами логики, а какими-то другими соображениями.

В эксперименте Гигеренцера (1998) испытуемым предлагалось следующее условие: предположим, что у 0,3 % людей есть рак толстой кишки. Существует 50-процентная вероятность, что тест на рак толстой кишки обнаружит заболевание, и 3-процентная вероятность, что он покажет, что рак имеет место, когда его на самом деле нет. Вопрос: какова вероятность того, что у человека с положительным тестом рак? Согласно правилу Байеса, правильный ответ —

<sup>17</sup> Данные взяты из [313, р. 12771–12772]

4,8%. Гигеренцер — обратите внимание — опрашивал профессиональных медиков, которых трудно заподозрить в недостатке релевантного опыта и научения. И тем не менее, медианный ответ опрошенных — 47%, почти в 10 раз больше математически правильного.

Эти эксперименты спровоцировали интенсивное обсуждение в литературе и породили некоторое количество объяснительных гипотез<sup>18</sup>. Так, сам автор обзорной статьи считает, что «выполнение карточных заданий у Уэйсона (1968), статистического задания Гигеренцера (1998) и другие эксперименты демонстрируют, что люди на самом деле иррациональны по сравнению с принятыми (в их культурах — И.М.) стандартами логического вывода» [313, р. 12772]. Тверски и Канеман считают, что в этих экспериментах люди переносят эвристические и эмпирические правила, успешные в повседневной жизни, в лабораторные условия, где они не применимы. В этих интерпретациях проглядывает убеждение, что всякое отклонение от единственно возможной рациональности является иррациональным — убеждение, с которым полемизируют Фелин и соавторы.

Напротив, Джон Андерсон, классик когнитивной науки в её классическом — символическом — воплощении, полагает, что критический момент того, что можно было бы назвать «рациональным анализом», состоит в том, чтобы выяснить, какой показатель может оптимизировать когнитивная система, и на основе этого делать прогнозы о том, как люди будут вести себя в конкретных экспериментальных задачах. Как бы конкретизируя эту мысль, Оуксфорд и Чейтер подчёркивают, что люди не рассматривают задачу как логический тест, а скорее пытаются установить причинно-следственную связь между двумя событиями. Не имея соответствующего опыта в отношении карточек, люди интерпретируют объявленное Уэйсоном правило как причинно-следственную связь, подобную тем, которые они видели в прошлом. В задаче Уэйсона, по их мнению, мы имеем ситуацию, в которой поведение нерационально по отношению к законам дедукции, но может быть понято как рациональное в контексте того, как люди обычно ищут информацию.

Я бы отметил то важное для понимания природы рациональности обстоятельство, что причинно-следственная связь является ассоциативной, а не логической. Как считают Бендер и Беллер, «в текущих исследованиях всё ещё обсуждается «модифицированная версия [концепции Леви-Брюля — И.М.]»

<sup>18</sup>Они в основном цитируются по [313, р. 12772–12775], если не указано иное.

согласно которой два способа мышления все же различаются — иногда их называют “основанное на правилах” в отличие от “ассоциативного”, “рефлексивное” в отличие от “интуитивного”, или “абстрактное” в отличие от “контент-специфичного” — хотя и сосуществуют они, как предполагается, во всех культурах (курсив мой — И.М.)» [314, р. 2]. И достаточно важное уточнение было сделано самим Гигеренцером. Если испытуемым медикам та же задача предлагалась не в процентно-статистических, а в простых арифметических терминах — 30 из каждых 10000 человек имеют рак толстой кишки; из этих 30, у 15 тесты будут положительными; из оставшихся 9 970 человек без рака у 300, тем не менее, будет положительный результат — то 67 процентов испытуемых отвечали правильно, по сравнению с 4 процентами в том эксперименте, когда данные были представлены по правилам теории вероятности. Из этого факта можно сделать вывод, что успешность решения задачи зависит не только от содержания и сложности данных, но и от соответствия формы её репрезентации повседневной практике людей в поиске информации.

### 3.5.3 Вывод из раздела 3.5

Как я полагаю, дискуссии, представленные в этом разделе, делают вполне обоснованным следующее обобщение.

**Изначально мы — люди — как и другие животные, не логические машины, а ассоциативно-статистические, поскольку имеем «на борту» не серийный, а параллельный (нейросетевой) компьютер. Следовать дедуктивным правилам нас принуждают наука, юриспруденция и другие социальные институты, осуществляющие и регулирующие социальные вычисления. Логика, будучи важным порождением этих институтов, тем не менее, очень мало знает о реальном человеческом мышлении.**

## 3.6 Как возможна когнитивная теория общества?

### 3.6.1 Когнитивность как методологическое обязательство

В большинстве известных социальных теорий возникает метафизическая проблема целого и частей, которыми в конечном счёте оказываются составляющие общество индивиды. Однако существование индивидов наглядно, а общество как таковое дано в абстракциях, и не обязательно научных или философских, а тоже как бы реально существующих где-то рядом с индивидами. Возникает платоновская проблема умопостигаемого бытия, разделаться с которой здесь нельзя так же просто, как в естественных науках.

Здесь должен возникнуть естественный вопрос: а чем может помочь когнитивная теория общества? Ведь она, скорее всего, предложит выводить социальное из психологического, т. е., до- и внесоциального начала, что, вообще говоря, уже практиковалось. Здесь нужно разобраться со смыслом термина «когнитивный». Этот термин может относиться как к предмету, так и к методу науки. «Когнитивной психологией» традиционно называют раздел науки о душе, который посвящён памяти, вниманию, воображению и т. п., какими бы методологическими предпочтениями ни направлялись эти исследования. Когда же мы говорим о «когнитивной науке», или «когнитивных науках», мы имеем в виду комплекс дисциплин, различающихся по предмету или ведомственной принадлежности — как, например, лингвистика и искусственный интеллект — но объединённых общим методологическим подходом. В его основе лежит то, что не так давно называлось «компьютерной метафорой». Теперь, я полагаю, его можно назвать компьютеризмом. Этот термин следует понимать в самом широком смысле — как признание того, что сложные процессы в нелинейных системах реально представляют собой или, по крайней мере, могут быть представлены как вычисления. Вычисления могут быть тьюринговы или нетьюринговы, линейные или распределённые, серийные или параллельные, цифровые или аналоговые, натуральные или искусственные и т. п. Главное, что их объединяет, это, во-первых, возможность описать их — на каком-то уровне — как исполнение алгоритмов и, во-вторых, их относительная независимость от конкретного физического воплощения: любая материальная система должна

быть способна реализовать данное вычисление, если она обладает достаточным для него числом возможных состояний или, что то же самое, степеней свободы.

Как мы видим, «вычисление» не связано непременно ни с «числом», ни с собственно человеческими целями и свойствами. В общем виде его можно было бы определить как формальные операции над структурами, осуществляемые как в искусственных, так и в естественных системах.

Таким образом, если мы занимается не просто когнитивной психологией, но когнитивной наукой о психике, то мы понимаем свой предмет как вычислительную систему той или иной архитектуры, обладающую определёнными состояниями и осуществляющую определённые процедуры, которые реализуются и сменяют друг друга по определённым алгоритмам. Но никто не мешает нам и социальные системы понимать как вычислительные со своими процедурами и алгоритмами. И тот факт, что общество состоит из индивидов, выглядит для нас таким образом, что некая вычислительная система состоит из элементов, каждый из которых — тоже вычислительная система или процессор. И тогда естественным образом возникает мысль о социальных вычислениях как о параллельных распределённых вычислениях. Иными словами, в общественном суперкомпьютере вычислительные задачи распределяются между многочисленными процессорами. По мнению Е. Г. Драгалиной-Чёрной, «социальные алгоритмы представляют собой статусные функции, исполнение которых зависит от коллективной интенциональности – институциональных актов согласованного приписывания и исполнения этих функций» [315, с. 150].

Соответственно, в каждом элементарном процессоре выделяются ресурсы под решение социальных задач и, наряду с этим, сохраняются ресурсы, решающие собственные индивидуальные задачи. И тогда перед когнитивной теорией общества встают две основные задачи: одна более простая, другая — более сложная. Простая состоит в том, чтобы научиться однозначно различать индивидуальную и социальную части когнитивного аппарата. Сложная состоит в том, чтобы показать необходимость и способы превращения когнитивного в социальное.

### 3.6.2 Образы когнитивно-социальных наук

Далее я попытаюсь дать обзор проектов и результатов исследований тех или иных когнитивных механизмов социальных взаимодействий, которые представлены в англоязычной научной литературе последних двух десятилетий. Некоторые из этих проектов выглядят величественными метафизическими построениями, другие, наоборот, утопают в измерениях и экспериментах. Но тем объёмнее картина проблемного поля.

В книге «Когнитивные измерения социальной науки» [316] Марк Тёрнер проводит достаточно любопытную аналогию между когнитивными исследованиями социальных отношений и классической риторикой. Именно риторы, по его мнению, с древних времён обучались знаниям о когнитивных ограничениях людей и искусству использовать их в своих целях. Он иронически призывает «бросить горсть земли на памятную плиту такой дисциплины как риторика — мир праху её — и предпочесть вместо неё более современное название для нашего проекта, возможно, что-то вроде “когнитивной социальной науки”» [316, p. 154].

Уподобляя человеческий мир театру в согласии с заезженной метафорой, Тёрнер говорит о закулисных когнициях (backstage cognition), которые, по его мнению, должны стать «зонтиком», под которым объединятся когнитивная наука и социальные науки. К закулисным когнициям, более конкретно, относятся исследования порождения значений, мышления, выбора, изменения и формирования понятий, «поскольку всё это тесно сопряжено с человеческой нейробиологией и исполняется на сцене, образованной всем многообразием человеческих сообществ и культур» cite[p. 154–155]turner2001.

Рон Сан, известный исследователь социально-когнитивных сюжетов, в «Пролегоменах» к коллективной монографии «Обосновывая социальные науки когнитивными науками» пишет, что когнитивные науки могут служить основой для социальных наук, во многом таким же образом, как физика даёт обоснование для химии или квантовой механики, или как квантовая механика даёт обоснование для классической механики [245, p. 5].

Социально-когнитивные взаимодействия, по его мнению, могут быть исследованы на разных уровнях, которые могут быть представлены как совокупность смежных дисциплин, от самых макроскопических до самых



микроскопических. Эти различные уровни включают социологический, психологический, компонентный и физиологический уровни. Первый, второй и четвёртый уровни достаточно очевидны. Но компонентный уровень достаточно интересен. На этом уровне мы пытаемся понять сознание с точки зрения его компонентов, применяя язык определенной теоретической парадигмы. Этот уровень «может включать концептуальные, вычислительные и / или математические структурные спецификации, такие как определение общей вычислительной архитектуры сознания и его компонентов» [245, р. 7]. Учитывая сказанное мною вначале, становится ясно, что именно на этом (а не на традиционно-психологическом) уровне в наши построения входит когнитивное в собственном смысле слова.

И далее начинается дальнейшая экспансия когнитивного видения. Несмотря на то, что, на первый взгляд, компонентный уровень в основном касается внутриагентных процессов, концептуальные, вычислительные или математические модели, разработанные на нем, могут быть использованы для понимания процессов, происходящих на более высоких уровнях, включая взаимодействия на социологическом уровне, в которых участвуют многие индивиды [245, р. 7]. Это значит, что когнитивный (в определённом мною смысле) подход может и стать парадигмой не только психологического, но и социального знания.

Первый факт, который открывается из этой перспективы состоит в том, что социальная реальность определяется когнитивным устройством образующих его индивидов, которое выступает не только усиливающим, но и ограничивающим её фактором.

Когнитивные особенности индивидов «могут повлиять на то, какая социальная система или институт лучше всего в каждом случае. Социокультурная изменчивость существует в результате активной человеческой деятельности и человеческой психологии, а не просто как некое предписание сверху для человеческих сознаний» [245, р. 16].

Автор отмечает важность в этом контексте понятие интериоризации, введенного Л. С. Выготским. Советский психолог считал социальное взаимодействие важной детерминантой в развитии мышления индивидов. Интериоризованные социокультурные знаки и символы позволяют людям развивать сложные репрезентации, в том числе сформированные социокультурно и исторически. Сан считает, что на основе понятия интериоризации могут быть разработаны детальные вычислительные модели. Он пишет: «В качестве

примера здесь можно привести когнитивную архитектуру CLARION. Интериоризация может быть выполнена в CLARION посредством процесса “сверху вниз” <...>, который хорошо соответствует феноменологической характеристике интериоризации» (Sun 2012,19).

В целом, делает вывод автор, социальная реальность является, в некотором роде, совокупным продуктом действий, подпрограмм, навыков, знаний, решений и мыслей индивидов, каждый из которых имеет непосредственное, содержательное взаимодействие со своим миром. Структуры социального макроуровня, независимо от того, являются ли они осязаемыми или нет (то есть физическими или нефизическими), ограничивают индивидуальное поведение только в том случае, если существуют индивиды — структуры макроуровня не являются независимыми от отдельных лиц в этом смысле. Более того, они имеют значение только в том случае, если люди воспринимают, интерпретируют, распознают, запоминают и реагируют на них.

Итак, согласно Сану, в конечном счете, они существуют внутри и через людей и их когнитивных способностей.

Пол Тегард [317] спорит против крайностей методологического индивидуализма и постмодернистского холизма, предлагая концепцию взаимного конституирования общества и индивидов. По его мнению, в психологии культуры восторжествовал более богатый взгляд на динамическую взаимозависимость «я»-систем и социальных систем: психологические и культурные взаимосвязи конституируют друг друга и должны быть проанализированы и поняты вместе. Вместе с тем, с точки зрения естественных наук, идея взаимного конституирования выглядит странной, поскольку конституирование в физических системах является однонаправленным, асимметричным отношением части и целого [317, р. 35-36].

Взаимное конституирование состоит в том, что социальные связи в основном психологичны и возникают из-за того, что у индивидов в группе есть ментальные представления, в том числе такие, которые отличают их как членов данной группы. Однако процесс связывания частей целым не является чисто психологическим, поскольку он может также включать различные виды физических взаимодействий, которые являются по сути социальными, лингвистическими или иными, такими как участие в ритуалах или в юридических договорах, или даже простой зрительный контакт. Эти взаимодействия объединяют людей в группы, когда и если они порождают ментальные репрезентации

(как аффективные, так и когнитивные), через которые индивидуумы осознают себя как часть группы.

Тегард предлагает взгляд на личность как на многоуровневую систему, действующую на социальном, индивидуальном, нервном и молекулярном уровнях. Каждый из этих уровней может быть использован для объяснения эмоций, сознания и других важных аспектов личности [317, р. 38].

Вместе с тем, уже в течение нескольких десятилетий проводятся исследования в рамках области, обозначаемой как *когнитивная социология*. Э. Зерубавель и другие классики этого направления рассматривают его как область исследования, которая изучает аспекты познания, не являющиеся ни когнитивными универсалиями, ни специфичными для отдельных лиц. Эта промежуточная точка между когнитивным универсализмом и когнитивным индивидуализмом важна для понимания социальных процессов выстраивания границ и создания смысла. Согласно этому подходу, люди не думают ни как универсальные социальные акторы, ни как индивиды, но как участники социального и культурного контекста.

В коллективной статье [318], посвящённой проблеме расовых предрассудков, высказывается мысль, что когнитивная социология занимается классификациями, коллективной памятью, социальным невниманием и вниманием, коллективным смыслообразованием и социальным созданием групп идентичности посредством социальной маркировки. Вместе с тем, авторы обнаруживают важные отличия когнитивной социологии от когнитивной социальной психологии. Так, например, когнитивная социология фокусируется на индивидуальном познании в непосредственном социальном контексте и занимается тем, как индивиды выбирают из нескольких стратегий обработки информации на основе целей, мотивов, потребностей, чувств и ситуационного контекста. Когнитивная социальная психология, напротив, рассматривает человека, с точки зрения его подверженности социальному влиянию, но обычно не рассматривает коллективное познание на уровне сообществ как единицу анализа.

Когнитивная социальная психология и когнитивная социология также различаются, по мнению авторов, методологически. Когнитивная социальная психология обычно использует экспериментальные методы, тогда как когнитивная социология основывается на исторических, этнографических и других социологических подходах, в которых социальная жизнь анализируется за пределами экспериментальной ситуации.

Важным направлением исследования когнитивных оснований социальности является *психология совместных действий*. Обзор идей и исследований в этой области содержится в статье [319] другого коллектива авторов.

Психологическое исследование концентрируется на структуре совместных действий и взаимном использовании его участников когнитивных ресурсов или маркеров друг друга. Так, выполняя совместную задачу, её участники обычно контролируют ход своих задач, чтобы определить, согласовано ли текущее состояние совместных действий и желаемые результаты. Важным источником информации о том, что видят другие люди, и об их ментальных состояниях являются движения их глаз. Не менее важной выступает способность известная в традиционной психологии как моторное предсказание — опережающее понимание актором вероятного результата своего физического действия. И хотя оно в основном изучается в рамках исследования индивидуальных действий, некоторые данные свидетельствуют о том, что оно поддерживает и совместные действия, обеспечивая точную временную координацию, и что оно модулируется прошлым опытом собственных действий субъекта.

Соответственно, участники совместного действия могут корректировать его кинематические особенности (например, скорость или высоту движения), чтобы сделать свои собственные действия более предсказуемыми для другого человека. Например, если дверь нужно открыть при переноске диван в другую комнату, это сделает актер, который ближе к двери, а другой на мгновение возьмет на себя больший вес, чтобы обеспечить поддержку.

Спорным мне представляется понимание авторами культуры. Как сказано в статье, «Культура и является продуктом широкомасштабных действий, таких как торжества или протесты, и в то же время она существенным образом определяет, как люди подходят к совместным действиям в небольших межличностных взаимодействиях» [319, р. 5]. Если принять эту точку зрения, встанет проблема объяснения природы и причин этих самых «широкомасштабных действий», которое уже явно будет за пределами теории культуры. Более рациональным, на мой взгляд был бы подход с точки зрения восходящей причинности, в какой-то мере согласующийся с концепциями Сана и Тэгарда. Культуру в этом случае следует понимать как результат тонких когнитивно обусловленных взаимодействий на микроуровне, из которых ткётся ткань большого целого.

С точки зрения «механистических» объяснений и эмпирических подтверждений интерес представляет область исследования, определяемая как

*когнитивная социальная нейронаука*. Так, Натан Эмери [320] обращает внимание на распространённое среди исследователей мнение, что жизнь в социальной группе и прогнозирование поведения представителей одного животного вида требует беспрецедентных уровней когнитивной обработки, которые не обнаруживаются у неприматов. Эта «гипотеза социального интеллекта», как она известна в этих профессиональных кругах, была предложена в качестве альтернативы более традиционным кандидатам на объяснение эволюции приматов и человеческого интеллекта: таким как использование инструментов, охота, расширенная пространственная память или добывающие промыслы. Из-за антропоцентрического смещения тех, кто работал в этой области в то время, и кто сосредоточился исключительно на эволюции человеческого интеллекта, экспериментальная и теоретическая работа в этой области действительно касалась только сравнения нечеловеческих и человеческих приматов.

Была, в частности, предложена идея, что есть преимущества в понимании и запоминании предыдущих взаимодействий и взаимоотношений представителей одного вида, и что эта информация может использоваться для прогнозирования или манипулирования их поведением в будущем. Ещё в 1980-х гг. было обнаружено, что существует явное сходство между сложным социальным поведением шимпанзе и политическими махинациями, которые распространены в государственных делах людей, что впоследствии было названо «политикой шимпанзе», или даже был предложен более яркий термин — «макиавеллианский интеллект». Если говорить о приматах, то размер их мозга и, более конкретно, объём неокортекса в отношении к объёму остальной части мозга, коррелируют со средним размером группы; то есть, приматы с относительно большим мозгом обычно образуют более крупные социальные группы. В то же время, нет никакой корреляции между социальным обучением, инновациями и размером группы у приматов, хотя существует значительная корреляция между социальным обучением и размером неокортекса.

В статье [321], посвящённой функциональной анатомии социального познания, рассказывается о так называемом тесте Салли-Энн. Салли и Энн — это две куклы. Ещё имеются две коробочки, свечка и маленькая девочка — собственно, испытуемая. Одна из кукол удаляется из комнаты, и свечка перепрячивается под другую коробочку. Когда кукла возвращается, у девочки спрашивают: под какой коробочкой Энн будет искать свечку? Экспериментально установлено, что репрезентация «ложных убеждений» развиваются у детей

в возрасте 4 лет. Т. е. только начиная в среднем с 4 лет ребёнок начинает понимать, что кто-то другой может иметь убеждения, отличные от его собственных. До этого момента, по его мнению, вернувшаяся кукла будет искать свечку под той коробочкой, под которой она лежит сейчас, а не под той, под которой она лежала на момент удаления куклы из комнаты.

В социальной когнитивной нейронауке традиционно существенную роль играет понятие «теория сознания» (Theory of Mind, ToM), которое также обозначается как ментализация, метарепрезентация или вторичной репрезентация. Под этим понимается способность понимать психологические или ментальные состояния других людей, такие как их убеждения, желания и знания. Различные формы ToM подразделяются на три класса: перцептуальная «теория сознания» (понимание зрения и внимания), мотивационная (понимание желаний, целей и намерений) и информационная (понимание знаний и убеждений). Таким образом, социальное познание интерпретируется как «обработка любой информации, которая завершается точным восприятием предрасположенностей и намерений других людей» [321, p. 157].

Эмпирические исследования роли отдельных секторов мозга в социальном познании приоткрывают интересные факты. Так, стало ясно, что миндалевидное тело не просто важно для распознавания потенциальной угрозы от соответствующих эмоциональных выражений лиц и других стимулов, но также имеет решающее значение для признания несколько более тонкой связанной с угрозой функции, «надежности».

Вентролатеральные и дорсолатеральные префронтальные области, возможно, в большей степени справа, модулируют ответы миндалин на расовые лица вне группы [321, p. 171]. Левые префронтальные области, подобные тем, которые были выявлены в исследованиях имитации и активности «зеркальных нейронов», по-видимому, важны для атрибуции признаков личности по крайней мере в определенных контекстах [321, p. 168]. Медиальные префронтальные области более активно участвуют в назначении характеризующих слов людям, чем в приписывании дескриптивных прилагательных объектам. Наконец, орбитофронтальные области могут быть важны, по крайней мере, для одного типа социальной категоризации [321, p. 173].

Нейрокогнитивные основы «Я» и самосознания исследуются в [322]. Эмпирические данные дают возможность утверждать, что особая роль в этом отношении принадлежит задней теменной коре, соответствующие свойства ко-

торой только начинают становиться предметом исследования самосознания. Задней теменной коре, как правило, приписывались функции обслуживания рабочей памяти и пространственной обработки. Однако этот отдел мозга может быть местом, где несимвольные, параллельные, распределённые репрезентации переводятся в символические, последовательные, локальные репрезентации.

На это обстоятельство я хотел бы обратить особое внимание. На мой взгляд именно здесь проходит водораздел между индивидуальными и социальными когнициями (см. определение термина на с. 306). Если первые целиком зависят от нейросетевой архитектуры мозга и поэтому характеризуются параллельностью и распределённостью, то вторые формируются в ходе социальной коммуникации и зависят от архитектуры языка — линейной и последовательной. И если задняя теменная кора, действительно, есть «дом» самосознания, то это обстоятельство вполне можно рассматривать как эмпирическое подтверждение понимания «Я» как социального конструкта.

Либерман и Пфайфер указывают, что существует соблазн думать о «Я» как об объекте со стабильными атрибутами. На самом деле, это искушение не только для ученых, но и для всех людей, которые ценят чувство собственного достоинства и независимости. Однако, как показывают нейрокогнитивные исследования, «Я», по крайней мере, частично строится и реконструируется с течением времени как функция ситуационных и межличностных ограничений [322, р. 223].

Необходимо отметить, что проблема коллективной интенциональности широко обсуждается в том числе на философском уровне, хотя, по мнению Брюса Хюбнера, постановка этой проблемы в рамках аналитической философии сознания традиционно связана с представлением, что *лучшее объяснение* коллективной интенциональности — это объяснение, основанное на интенциональности индивидов. [323, р. 73]. Действительно, основываясь на наших онтологических интуициях, естественно полагать, что коллективная интенциональность супервентна (см. определение термина на с. 306) по отношению к индивидуальной. Но одна из целей настоящего исследования состоит в том, чтобы показать непосредственное наличие и операциональное участие элементов коллективной интенциональности в когнитивном «устройстве» индивида.

### 3.6.3 Проект когнитивной науки об обществе

Как мы видим, в основе всех многообразных когнитивных подходов к социальности лежит представление о существенной важности опознания когнитивным агентом себе подобных и взаимного контроля акторов в социальных ситуациях. Оставаясь на эволюционных позициях и рассматривая мир через оптику традиционного естествознания, несложно понять, как это стало возможно, но достаточно трудно найти объяснение, почему это в какой-то момент стало необходимо. Но мы окажемся ближе к пониманию этого, если предположим, что возрастающая структурная сложность — главный способ для сложных нелинейных систем противостоять энтропии. Их деятельная забота о повышении степени структурности реализуется в процессах, которые мы называем вычислительными. Вычислительная система, обладающая памятью и обратной связью, становится когнитивной и естественным образом стремится к дальнейшему повышению уровня сложности путём использования чужих когнитивных ресурсов. Результат — объединение когнитивных систем в социальную суперсистему. Если такое видение имеет реальные теоретические перспективы, то я бы выделил три теоретико-методологические основы будущей когнитивной теории общества.

Это, во-первых, **компьютационализм**. Компьютационализм состоит в понимании церебрально-когнитивных и социальных систем как вычислительных устройств разного уровня и архитектуры. За пределами вычислительного подхода достаточно трудно понять, зачем системам наужны когнитивные свойства вообще, и в частности, зачем они постоянно усложняются в ходе эволюции.

Во-вторых, это **когнитивизм** — переход к пониманию социальных систем как этапа развития церебрально-когнитивных систем. В традиционной концептуализации когнитивное — это характеристика мозга и тела индивида, а взаимодействия между индивидами — это сфера социального, то в моей концепции социальное есть этап развития когнитивных свойств и функций. Именно поэтому становится возможным рассматривать когнитивные науки как основание социальных.

Наконец, **коммуникативный функционализм**<sup>19</sup> требует от исследователя интерпретировать «высшие» когнитивные способности как функции

<sup>19</sup>Коммуникативный функционализм как принцип рассмотрен в разделе 1.3.2.



социально-когнитивных систем. Особенно это касается собственно человеческих способностей, выделяющих их носителей из животного мира: рационального мышления и языка. И то, и другое существенным образом отличается по своей архитектуре от досоциальных когнитивных механизмов: если те реализуются в распределённых репрезентациях и параллельных вычислениях, то язык и рациональное мышление линейны и серийны, в идеале стремятся к последовательности и однозначности. Это свойства, необходимые для социальной коммуникации на том этапе развития человечества, на котором оно находится сейчас.

Я бы поддался искушению назвать описанный подход принципом трёх «К», но из-за очевидных и нежелательных ассоциаций с ку-клукс-кланом, пожалуй, откажусь от этой идеи. Как бы то ни было, этот подход находится в русле тенденций последнего времени, когда мы видим появление вычислительной биологии, вычислительной экономики и даже вычислительной астрономии. Он предлагает возможный унифицированный ответ на самые интересные вопросы относительно нашего биологического вида: как возможно, что мы умеем думать и знать, и как получается, что мы умеем объединяться и взаимодействовать.

### **3.6.4 Социальная организация как естественный параллельный компьютер**

Один из наиболее известных сегодня пропонентов возрождения социальной онтологии как особого предмета исследования, Тони Лоусон, видит элементарную единицу социальной онтологии в социальной практике. Он определяет её как «способ действия, который (неявно) имеет статус (коллективно) принятого в сообществе» [84, р. 34]. Там же он уточняет, что социальные практики — суть принятые или признанные наблюдаемые способы ведения дел, которые определяют методы, которым люди следуют в рамках определенного сообщества [84, р. 34]. Можем ли мы предположить, что «способ ведения дел» сродни алгоритму? Что-то похожее на: «если вы хотите  $x$ , то выполните шаги  $a$ ,  $b$ ,  $c$ , и если произойдёт  $d$ , то выполните  $e$ , в противном случае выполните  $f$ ». Безусловно, социальные практики могут определяться не только выбором

возможных действий, но и некоторыми эстетическими аспектами, такими как форма одежды. Но выбор подходящего костюма также может быть представлен как «логический вентиль» в алгоритме. Такой взгляд на социальную действительность может столкнуться с некоторыми сложностями. Онтология имеет смысл, если она сводится к некоторым элементарным объектам определенной реальности, которые нельзя представить как сложные. Алгоритм, или, как говорит Лоусон, «способ ведения дел», не может быть представлен таким образом, поскольку «вещи» должны быть названы и описаны, чтобы идентифицировать алгоритм, который имеет к ним отношение. Кто-то может подумать, что сама суть вещей, обрабатываемых социальными алгоритмами (или практиками, в терминах Лоусона), и есть то, что отличает социальные алгоритмы от естественных. Но в чём их отличие? Если я пойду на рыбалку только со своими собственными инструментами ручной работы, мой способ ведения дел будет определяться их физическими свойствами. Но что является физическим в этом случае? Этот термин означает только то, что мне не нужно думать, скажем, о марках производителей удочек и крючков или о том, оплатил ли я лицензию на рыбалку. Я просто использую производительные свойства вещей, не имеющие общественного происхождения. Напротив, если мне приходится обращать внимание на некоторые «нефизические» свойства вещей, то они оказываются определенными токенами, так же распознаваемыми по определенным физическим свойствам — например, форма и цвет логотипа или бумага с чернилами. Принципиальное отличие этих вещей заключается не в их собственной «природе», а в том, как они обрабатываются. В одном случае борода на лице человека или его выбритые щёки ничего не значат, кроме его предпочтений в стиле и личной гигиене. Но в ином случае тот или иной вариант внешности может означать его профессиональную принадлежность. Соответственно, люди, встречающие этого человека, или вообще игнорируют наличие или отсутствие бороды у него, или, наоборот, принимают важные решения о вступлении в некоторые отношения с ним. Борода остается «физической» в любом случае, меняется алгоритм. Поэтому необходимо выяснить, что определяет разницу в алгоритмах такого рода. Проще говоря, если мы воспринимаем бумагу с чернилами как рыболовную лицензию, мы должны помнить об определенном учреждении, которое выдает такие лицензии. Если мы рассматриваем бороду человека как знак его принадлежности, мы, скорее всего, помним об устойчивой группе людей, которые обычно используют бороду в качестве способа идентификации. В целом, что-

бы иметь возможность выполнять алгоритмы такого рода, нужно (1) обладать некоторыми когнитивными способностями — такими, как память, различие и категоризация — и (2) участвовать в длительных отношениях с этими людьми. И тогда мы в состоянии подойти к правильному пониманию социальной онтологии. Как уже было сказано, исследование социальной реальности должно дойти до конечных объектов, которые не могут быть далее делимы без потери их сущности. Наиболее распространенными кандидатами на эту роль обычно были люди или их отношения. Но в обоих случаях нам необходимо некоторое дополнение, чтобы сделать эти объекты социальными, так как человек может выпасть из социальной структуры, а межличностные отношения могут оказаться биологическими, а не социальными (как в случае случайного секса или каннибализма). Я полагаю, что в поисках социальных атомов мы должны обратиться к составляющим частям алгоритмов (или «практик», по Лоусону) как таковым, поскольку мы установили, что именно они хранят секрет социального. Первое, что необходимо здесь отметить, — это то, что социальные алгоритмические процессы являются вычислительными. Хотя здесь возможны возражения. Например, если кто-то копает канаву — а это последовательность операций, которые могут быть выполнены машиной — то этот труд должен называться вычислением в соответствии с моим определением, что сильно противоречит интуиции. И это возражение справедливо на данном уровне сложности. Но представим себе, что землекоп находится в среде, где он может наткнуться или на египетскую мумию, или на неразорвавшуюся бомбу времён войны, или на сокровище пиратов. Он / она может действовать одинаковым образом в каждом из этих случаев, но результат вряд ли будет удовлетворительным. Чтобы землекоп действовал «умнее», ему / ей нужен набор правил «если—то—иначе», который является не чем иным, как простым примером алгоритма: в случае А вызывай археологов, в случае В — сапёров, в случае С звони в полицию, ничто из перечисленного — продолжай копать. Если землекоп — человек, то то, что он / она делает в соответствии с таким алгоритмом, есть вычисление в собственном смысле. Но, насколько я знаю, в наше время машины могут реализовывать такие и даже гораздо более сложные алгоритмы, включая визуальное распознавание, категоризацию, статистическое обучение и многое другое. И человек, и машина вычисляют, проходя через цепочку альтернатив. Тогда вычисление — это деятельность машины, способной регулярно изменять свое поведение в соответствии с изменяющимися входными данными. Занима-

ясь вычислениями, мы действуем как машины. Важно отметить, что до сих пор мы говорили о вычислениях как о линейных (или последовательных) действиях. Но, как мы теперь знаем благодаря некоторым новым технологиям, вычисления могут быть организованы параллельно, выполняя сравнительно простые алгоритмы на множестве взаимосвязанных процессоров (скажем, нейронов), что приводит к сложным эмерджентным результатам. И то, что я называю социальными вычислениями, имеет очень похожую архитектуру. Вторым пунктом моего онтологического анализа касается элементарных единиц «социальных практик», рассматриваемых как вычисления. Поскольку ранее было установлено, что социальные атомы должны быть единицами, которые не могут быть далее делимы, не переставая быть социальными в каком-то принципиальном отношении, а люди и их отношения не соответствуют этому условию, мы должны более внимательно взглянуть на структуру социальных алгоритмов. Специалистам по компьютерам известно, что алгоритмы состоят из некоторых типичных элементарных операций, выстраиваемых в различных комбинациях. Их обычно называют вычислительными примитивами [171], которые должны служить строительными блоками более сложных алгоритмов. Это приводит нас к идее идентификации подобных примитивов в социальных вычислениях, которые я для краткости назову социальными примитивами. Их обнаружение — это подлинное искусство воссоздания реальной структуры социальных алгоритмов, что отчасти похоже на так называемый «обратный инжиниринг» (т.е., воссоздание изначального чертежа реально работающего устройства), в основе которого — угадывание логики его создателей. Что касается анализа общества с этой точки зрения, это, конечно, большой исследовательский проект, для которого очевидно тесны рамки небольшой статьи. Но можно попытаться представить, какими могут быть социальные примитивы. Если мы говорим о сетевой структуре, которая является средой для параллельных распределенных вычислений, то здесь алгоритмы выполняются на уровне отдельных нейронов (агентов), в то время как эмерджентный результат проявляется на уровне целой сети. Тогда внутренние способности агентов, способных обрабатывать потоки данных в социальной сети, могут включать способности к следующим различиям:

- друг / враг
- разрешено / запрещено
- одобряется / осуждается

– сеньор / вассал, а также способность к распознаванию значений знаков, и так далее.

Важно отметить, что, когда дело доходит до научной теории в подлинном смысле, любые философские попытки дать некое дискурсивное определение в стиле «осуждение — это...» или «значение знака — это...» бесполезны. Все примитивы должны быть определены чисто функционально: что происходит в сети, когда что-то осуждается, когда определяется значение знака и т. д. Очевидно, что только такого рода методология позволит создать действительно надёжную социальную науку без плохо обоснованных несоизмеримых «заявлений» (claims), которые Лоусон упоминает не раз как один из главных недостатков этой науки в наши дни [84, р. 31]. Общество в этом контексте можно понять как вторичную или производную когнитивную систему, использующую сеть когнитивных аппаратов своих членов. Согласно концепции макрокогниции Б. Хюбнера, «коллективные сообщества должны рассматриваться как распределенные когнитивные системы, если они состоят из высокоинтегрированной сети механизмов и интерфейсов, которые принимают входные данные из окружающей среды и выполняют вычислительные процессы таким образом, что могут порождать поведение на уровне системы, чувствительное к непредвиденным обстоятельствам окружения» [323, р. 256]. Что касается интерфейсов «мозг–социум», то важнейшим из них является естественный человеческий язык, в основе которого лежит способность человеческого когнитивного аппарата к образованию предикативных связей: «когнитивная способность мыслить предикативно сконструирована в контексте социального взаимодействия из совокупности разнородных ресурсов, большинство из которых не являются когнитивными и обычно даже внутренними для сознания ребёнка, и чьё основание и функции специфичны и порождаются уникальными и быстро проходящими параметрами онтогенеза человека» [324, р. 46]. Общество, как и мозг, осуществляет параллельные (сетевые) вычисления для поддержки когнитивных функций. Работа этой системы зависит от таких свойств сети, как связность (показатель равномерного распределения связей между узлами). При недостатке связности возникают хабы (распределительные узлы), которые ответственны за появление социальной иерархии. В эпохи медленных информационных технологий иерархическое (хабизированное) общество эффективнее более связных. В эпоху быстрых информационных обменов сеть с более равномерно распределёнными связями оказывается эффективнее. В современном мире идут процессы

децентрализации (де-хабизации) не только политической власти, но и экономических систем (блокчейн и криптовалюты). Победит не тот, кто пытается противостоять этим тенденциям, а тот, кто сумеет их оседлать.

### 3.6.5 Вывод из раздела 3.6

Итак, когнитивная наука об обществе есть наука, основанная на вычислительном подходе. Этот подход завоевывает всё больше сторонников в академической среде, и такие новые социальные дисциплины, как когнитивная социология, психология совместных действий и когнитивная социальная нейронаука, в разной степени основываются на нём. Таким образом, **когнитивно обоснованная социальная теория необходимо должна быть вычислительной, видеть свой предмет как результат эволюции нейроцеребральных вычислений и основываться на ранее сформулированном принципе коммуникативного функционализма.**

### 3.7 Вывод из главы 3

В соответствии с «вычислительной метафорой», благодаря которой когнитивная наука появилась на свет, когнитивный аппарат живого организма представляет собой вычислительное устройство. Развитие когнитивных наук и теории искусственного интеллекта в последние десятилетия свидетельствует в пользу распределённой (сетевой) архитектуры этого устройства, которая реализует вероятностные алгоритмы для поиска ассоциативных связей. Эволюция в направлении повышения эффективности вычислений при снижении их энергозатратности приводит к объединению когнитивных возможностей живых организмов, что выражается в появлении их социальных организаций, также воплощающих сетевую архитектуру. Интерфейсом для такого объединения выступают сигнальные системы, позже — символические системы (языки). Последние в какой-то степени переформируют познание и основанное на нём

социальное поведение в направлении от ассоциативно-вероятностного к вербально-категориальному (рациональному).

Таким образом, во-первых, **когнитивные способности, выработанные в ходе эволюции, имеют ассоциативно-вероятностный характер, в то время как вербально-категориальное мышление, на безальтернативности которого построены классическая философия и наука, ограничено сферой коммуникации и предназначено для обеспечения сетевых взаимодействий когнитивных агентов. Во-вторых, сама социальная метасеть возникает как — и по сути представляет собой — расширение когнитивных возможностей живых организмов за счёт распределения вычислительных задач между ними.**

## Заключение

Вычислительный сетевой подход имеет все основания претендовать на роль трансдисциплинарной методологии, применяемой в исследовании сознания и общества. Вместе с подразумеваемой им онтологией он раскрывает конструктивный и инструментальный характер любых онтологий, включая свою собственную, поэтому при желании его можно рассматривать как нестандартную стратегию концептуального анализа, право на который всегда остаётся за философией.

Задача построения эффективной научной теории предполагает создание в том числе простой и непротиворечивой онтологии, доступной для формализации и, желательного, математизации. Логические и/или математические формализмы, непротиворечивым образом интерпретируемые на некоторой корректно построенной предметной области, представляют собой мощный инструмент объяснения и прогнозирования. И, если такая теория состоялась, дело остаётся за малым – чтобы её выводы и прогнозы соответствовали фактам. Из всех предложенных вариантов теорий выбирается тот, который лучше работает в этом направлении. Разработанное в рамках данного исследования метасетевое объяснение когнитивных и социальных свойств человека может предложить не только удовлетворительную онтологию сознания, но и удовлетворительную социальную онтологию.

Во-первых, в рамках предложенного видения две «реальности» — ментальная и социальная — связываются единой онтологией и — в перспективе — единым формальным (математическим) аппаратом.

Во-вторых, в этой интегрированной картине исследуемой реальности новое место занимает язык: он оказывается эволюционно развитым интерфейсом между нейросетью мозга и сетью социальных связей.

В ходе исследования было показано, что подлинные онтологические индивиды как таковые не образуют иерархий, поскольку последние предполагают именно различия в соизмеримых признаках, которые отсутствуют у простых объектов (логических индивидов). Возможные отношения между простыми элементами могут быть только сетевыми.

Исследование показало также, что сеть нейронов мозга и социальная сеть структурно аналогичны: обе состоят из элементов, умеющих выполнять неслож-



ные функции и оценивать «вес» своих связей с близлежащими элементами. Обе имеют когнитивные надстройки: соответственно, когнитом и когнитивные социальные сети. Если мы имеем компьютеры одной и той же нейронной архитектуры, управляемые разными программами, то между ними возможен интерфейс — библиотека функций, переводящих команды одного программного языка в команды другого. Таким интерфейсом между мозгом и обществом выступает язык.

Классический компьютерно-информационный подход известен хорошо разработанным инструментарием для исследования овладения языком, и его научные заслуги многочисленны. Его слабым местом является то, что, объясняя, как люди развили эту способность, он исходит из врожденности некоторых языковых структур, располагающихся на когнитивном уровне мозга. Напротив, коннекционизм разработал концептуальную основу, которая в особенности подходит для объяснения гибкости процесса обучения, а также приобретения способностей к грамматике и категоризации. Но эта концептуальная схема хуже справляется с объяснением некоторых аналитических свойств семантики человеческих языков. Учитывая эти соображения, философ оказывается перед сложным выбором.

В ходе исследования было продемонстрировано, что сочетание коннекционизма с лингво-прагматизмом в духе позднего Витгенштейна ведёт к выходу из узких рамок пропозиционального понимания мышления. Основным результатом этого исследования является интерпретация языковых значений как точечных воздействий социальной среды, которое, таким образом, создаёт основу для представления всего процесса как линейного и дискретного. Эти точечные воздействия надстраиваются над волнами активации нейронов головного мозга как социально контролируемое свойство психики.

Благодаря Декарту, европейская философия долгое время исходила из того, что в основе осознания субъектом чего бы то ни было лежит сознающее себя «я». У Канта необходимо сопутствующее всякой мысли представление «я мыслю» оказалось трансцендентальным условием связи понятий в суждении. В настоящем исследовании было показано, что философское «я», в отличие от обыденного, есть продукт употребления местоимения первого лица в *de facto* третьем лице квазинаучного описания.

Были рассмотрены возможные значения «я» в двух аспектах: коммуникативном и квалитативном. Загадка коммуникативного «я» тривиальна и

решается эмпирически. Загадка качественного «я» относится к сфере мистического, невыразимого, — в витгенштейновском смысле — и средствами известных нам науки и философии не решается. Но, как бы то ни было, и что бы по этому поводу ни говорилось в текстах классической философии, «я» как концепт не имеет теоретического отношения к постановке и решению проблемы осознанности.

Как показал анализ литературы, понимание степени осознанности как степени интегрированности нейронных возбуждений в различных отделах мозга подтверждается интересными эмпирическими исследованиями. Это означает, что интенциональные состояния, которые могут быть описаны как функции социальных отношений (коммуникации), должны быть описаны таким образом — как коммуникативные интенции, — а описание остальных должно сместиться на уровень «ниже» — в нейросеть головного мозга. Таким образом, мы получим два типа интенциональных состояний и актов: супервентные по отношению к социальной коммуникации и супервентные по отношению к нейронным связям мозга. При этом для их научного объяснения достаточно в одном случае онтологии, понимающей социальную ткань как системное взаимодействие простых индивидов, а в другом — онтологии, понимающих нейроцеребральную сеть как трансляцию возбуждений между простыми узлами. Ни в одном из этих случаев не требуется никакой дальнейшей редукции к теориям более низкого уровня.

Согласно большинству теорий социологического и социально-философского мейнстрима, общественные отношения отличаются тем, что в их каузальную структуру встраиваются состояния сознания. Те известные теории, которые пытаются увязать социум и сознание в единую каузально-онтологическую или концептуальную схему (Выготский, некоторые разновидности марксизма), делают это во многом декларативно, не предлагая работающего объяснительного механизма. Предложенный в диссертации подход обладает следующими преимуществами:

- общество и сознание оказываются частями единой онтологии, при том, что общество понимается как сеть, в каком-то смысле расширяющая сеть нейронов головного мозга, надстраиваемая над ней и использующая её возможности;
- к этой единой реальности применяются однотипные формализмы, реально превращая теорию общества и теорию сознания в единую науку.

При исследовании вопроса о соотношении когнитивного и социального в психологической и философской литературе чаще обсуждались механизмы детерминации когнитивных механизмов социальными структурами. Прогресс в области когнитивных наук в наше время даёт возможность взглянуть на проблему с другой стороны: каким образом когнитивные структуры отдельного индивида влияют на формирование и функционирование социальных связей. Тезис о том, что общество является подсистемой некоего когнитивного аппарата, звучит на первый взгляд несколько экстравагантно, но следует оговориться, что здесь он понимается его не в онтологическом, а в логическом смысле: если  $A \Rightarrow B$ , то  $A$  является подмножеством  $B$ . Действительно, можно указать на когнитивные свойства агентов, такие как чувство боли или способность следить за движущимся объектом, которые для своего осуществления не требуют социальной организации. Можно также назвать соответствующие им социально-когнитивные способности: например, распознавание в других чувства боли или состояния слежения за движением. Но вряд ли можно привести пример социального взаимодействия, не предполагающего участия когнитивных способностей, поскольку любое взаимодействие с другим агентом предполагает способность к восприятию, категоризации и коммуникации.

В ходе диссертационного исследования были получены следующие **результаты**:

1. Была обоснована роль научных онтологий в структуре научного знания как необходимого элемента, не сводимого ни к теоретическому, ни к эмпирическому компонентам.
2. Была разработана номиналистическая философская метаонтология, которая легла в основу когнитивно-социальной онтологии сетевых структур.
3. Была предложена когнитивно-социальная концепция метасети, построенная на номиналистической онтологии, обладающая следующими преимуществами:
  - 3.1. Она создаёт возможность использования строгих формальных языков, поскольку описывает когнитивно-социальную метасеть как взаимно увязанные системы распределённых (параллельных) вычислений, взаимодействующие через линейный (серийный) интерфейс — язык и другие символические системы.

- 3.2. Она допускает эволюционное объяснение, в соответствии с которым когнитивные и социальные системы складывались путём естественного объединения отдельных узлов (нейронов или агентов) в сеть с целью увеличения вычислительной эффективности системы.
  - 3.3. Она определяет подлинное место место вербально-категориального мышления и основанного на нём доказательного знания как совокупности средств (протоколов) коммуникации когнитивных агентов в социальной метасети.
4. Был предложен общий подход к будущей теории вычислений, которая могла бы охватить нейрофизиологические и социальные процессы, и которая описала бы их как систему многоуровневых распределённых статистических вычислений.
  5. Была обоснована эволюционная необходимость объединения естественных нейросетей в социальные метасети с целью увеличения вычислительной и энергетической эффективности путём распределения вычислительных задач между многочисленными процессорами, связанными единым сетевым протоколом — языком.

Полученные в ходе исследования результаты открывают **перспективы дальнейших исследований** по двум магистральным направлениям: философскому и конкретно-научному. В свою очередь, философские исследовательские перспективы можно разделить на относящиеся к философии сознания и те, которые имеют отношение к философии науки. Перспективы конкретно-научных исследований просматриваются, соответственно, в области когнитивных наук и в области социальных наук.

Для **философии сознания** результаты настоящего исследования имеют следующие последствия:

1. Получены дополнительные аргументы в пользу функционалистского подхода к сознанию. Если мозг и биосистемы в целом рассматриваются как распределённые системы статистических вычислений, а вычислительные задачи реализуемы многими различными способами (разными алгоритмами и в различных физических средах), то функции, связываемые с сознанием, могут быть как реализованы в искусственных системах

различной природы и архитектуры, так и «улучшены» в естественных биосистемах (людях и других животных). Такое видение открывает перспективы философского анализа эпистемологических, антропологических и этических последствий внедрения технологий управления сознанием.

2. Обоснованный в диссертации принцип коммуникативного функционализма предполагает онтологическое «разведение» ассоциативно-вероятностной и рационально-логической сфер сознания. Классический спор эмпиризма и рационализма, постоянно воспроизводящийся в новых терминологических обличьях, теряет свою актуальность. Как «малая» сеть — мозг, — так и «большая» (мета)сеть — общество, — производят вероятностные предсказания и вероятностные ассоциации повторяющихся данных. Но управляемые мозгом индивиды связываются в метасеть с помощью линейно и логически организованного языка, структура которого оформляет «этажи» человеческой психики, ответственные за коммуникацию, как сферу «рационального мышления». Таким образом, философский репрезентационализм, основанный на семантическом анализе, теряет основания. Возможно, более актуальным окажется какое-нибудь движение в стиле «назад к Юму». Логические же свойства мышления придётся рассматривать как привнесённые в психику извне, из сферы информационных обменов.
3. В качестве побочного результата диссертационного исследования можно указать на дополнительные аргументы в пользу концепции «ментальной краски» — иначе говоря, предположения о не-интенциональном характере квалиа. Если качественные характеристики ментальных состояний сами не транслируют содержание, а представляют собой лишь способ его проекции, то философские предположения об их собственной каузальной роли теряют поддержку, и путь решению «трудной проблемы сознания» выглядит более коротким и реалистичным.

В числе открывающихся перспектив исследований в области **философии науки** можно назвать следующие:

1. Тезис о независимости предметных (доменных) онтологий от научных теорий порождает некоторые новые дискуссионные вопросы. Например, нужно ли говорить о их полной независимости, или можно указать на

какие-либо формы мягкой, не обязывающей взаимозависимости онтологических презумпций и теоретических утверждений? Возможна ли строгая демаркация онтологических и теоретических высказываний в реальных научных теориях? Эти и подобные вопросы могут стать предметами будущих обсуждений.

2. В настоящем исследовании получены дополнительные доводы в пользу предположения о вычислительном образе науки как о некоторой новой исторической форме научных теорий, появившейся после спекулятивной и номологической формы. Соответственно, открываются перспективы исследований способов объяснения в вычислительных теориях, специфических взаимозависимостях между онтологиями, моделями (схемами), обобщениями и эмпирическими данными в этом типе научного познания.
3. Особое внимание может быть уделено проблеме демаркации философии сознания и когнитивных наук. Результаты новейших исследований, в том числе и представленные в настоящей диссертации, проливают дополнительный свет на то, какие проблемы сознания являются строго концептуальными и по праву остающимися за философией, а какие могут быть разрешены в ходе эмпирических исследований — и тогда философам нужно определиться, как их аналитические компетенции могут помочь когнитивным исследователям.

В области **КОГНИТИВНЫХ НАУК** можно предвосхитить следующие тенденции:

1. Выводы настоящего исследования недвусмысленно говорят о бесперспективности когнитивных моделей, основанных на квази-лингвистических аналогиях, вроде «языка мысли», и вообще каких-либо представлений о когнитивной деятельности как обработке символов, нагруженных семантикой. Исследователям стоит сосредоточиться на субсимвольных статистических моделях — в диапазоне между коннекционизмом и предиктивным процессингом. Объяснения языковых способностей следует искать в области социальных когниций. Философский анализ говорит в пользу большей эвристичности именно этих подходов.

2. Можно с некоторой степенью вероятности предположить, что территория традиционной «высокоуровневой» психологии будет всё больше осваиваться вычислительными моделями на базе нейронауки, а внутри этих моделей будет происходить поиск подсистем и алгоритмов, ответственных за социальное взаимодействие. Тенденции последних лет в меняющейся тематике научных англоязычных публикаций подтверждают это предположение.

Перспективы исследований в области **социальных наук** я вижу следующим образом:

1. Концепция «сетевого общества» должна преодолеть неизбежную спекулятивную стадию в своём развитии, представленную в работах таких авторов, как Мануэль Кастельс, и представить «механистические» модели, готовые к алгоритмическим описаниям социальных процессов.
2. Гипотеза общества как когнитивного расширения мозга также должна быть сформулирована в измеряемых терминах, чтобы превратиться в нормальную фальсифицируемую теорию, готовую ответить на вопрос: какие количественно измеряемые преимущества обретают отдельные когнитивные агенты, будучи объединены в распределённую вычислительную сеть? Возможно, не все объединяемые и распределяемые таким образом вычислительные процессы окажутся когнитивными в собственном смысле слова. Но в любом случае эта умозрительная гипотеза должна пройти положенный ей путь до состояния эмпирически проверяемой гипотезы.

В заключение хотелось бы **поблагодарить** всех тех, без кого эта работа не была бы выполнена. Автор выражает искреннюю признательность научному консультанту академику В. А. Лекторскому за неоценимый вклад в подготовку и проведение настоящего исследования, за поддержку, помощь, обсуждение результатов. Также автор благодарит проф. М. С. Киселёву и весь сектор методологии междисциплинарных исследований человека Института философии РАН за создание творческой и благоприятной атмосферы в процессе осуществления настоящего исследования. Особенно автор признателен заведующему сектором логики ИФ РАН д. ф. н. В. И. Шалаку за регулярные жаркие дискуссии по теоретическим вопросам, которые во многом помогли уточнить и улучшить представленные здесь позиции и выводы. За такую же поддержку, а

также за помощь в оформлении диссертации автор благодарит старшего научного сотрудника ИФ РАН А. В. Родина и авторов свободно распространяемого шаблона \*Russian-Phd-LaTeX-Dissertation-Template\*.

Трудно переоценить вклад жены диссертанта, которая, не будучи частью профессионального сообщества, тем не менее, своей поддержкой и моральным давлением помогла автору настоящего исследования преодолеть некоторые препятствия психологического характера. Автор также выражает признательность множеству неназванных здесь людей из профессиональной и иных сфер, которые сделали настоящую работу возможной.



**Список сокращений и условных обозначений**

- КСП** качественно-субстанциальный подход (с. 110)
- КФП** количественно-функциональный подход (с. 110)
- МАС** мультиагентная система (с. 251)
- МСТ** метасетевая теория (с. 227)
- ПГ** парадокс гомункула (с. 183)
- ПП** предиктивный процессинг (с. 145)
- ПФЭ** парадокс философской эгологии (с. 61)
- ФСМ** модель формально-символьных манипуляций (с. 136)
- ФСС** модель физических символьных систем (с. 136)
- Df** определение (с. 49)

## Словарь терминов

**Аффорданс** : Аспект или свойства предмета окружающей среды, создающие возможность его использования с той или иной целью.

**Гипостазирование** : Приписывание реального существования значениям абстрактных терминов.

**Когниция** : Множество ментальных процессов, участвующих в восприятии, мышлении, получении знаний, операциях с памятью, суждениях, решении задач, языковой деятельности, воображении и планировании.

**Процессинг** : Алгоритмическая обработка данных или информации в ходе вычисления.

**Супервентность** : Отношение обусловленности отношений элементов одной системы отношениями элементов другой системы. Система  $B$  супервентна к системе  $A$ , если и только если различию свойств любых элементов  $(b_n, b_m) \in B$  необходимо соответствует различие свойств некоторых элементов  $(a_n, a_m) \in A$ .

**Эмерджентность** : Появление или наличие у системы свойств, не являющихся свойствами её элементов, или существенно отличающихся количественно от суммы соответствующих свойств элементов.

## Список литературы

1. *Лекторский В. А.* Философия, познание, культура. [Текст] / Лекторский В. А. — Москва : Канон+РООИ Реабилитация, 2012. — 383 с.
2. *Лекторский В. А.* Человек и культура. Избранные статьи [Текст] / Лекторский В. А. — Санкт-Петербург : Санкт-Петербургский гуманитарный университет профсоюзов, 2018. — 640 с.
3. *Юлина, Н. С.* Головоломки проблемы сознания. Концепция Дэниела Деннета [Текст] / Н. С. Юлина. — Москва : Канон+, 2004. — 544 с.
4. *Васильев, В. В.* Трудная проблема сознания [Текст] / В. В. Васильев. — Москва : Прогресс-Традиция, 2009. — 272 с.
5. *Иванов, Д. В.* Природа феноменального сознания [Текст] / Д. В. Иванов. — Москва : Либроком, 2013. — 240 с.
6. *Смирнов А. В.* Сознание. Логика. Язык. Культура. Смысл. [Текст] / Смирнов А. В. — Москва : Языки славянской культуры, 2015. — 701 с.
7. *Смирнов, А. В.* Процессуальная логика [Текст] / А. В. Смирнов, В. К. Солондаев. — Москва : ООО «Садра», 2019. — 160 с.
8. *Касавин И. Т.* Социальная эпистемология. Фундаментальные и прикладные проблемы [Текст] / Касавин И. Т. — Москва : Издательский Дом "Альфа-М", 2013. — 557 с.
9. *Михайлов, И. Ф.* Человеческий мозг и сознание: биология или вычисления? [Текст] / И. Ф. Михайлов // Философские проблемы информационных технологий и киберпространства. — 2018. — Т. 15, № 2. — С. 92—110. — URL: [http://cyberspace.pglu.ru/issues/detail.php?ELEMENT%7B%5C\\_%7DID=264207](http://cyberspace.pglu.ru/issues/detail.php?ELEMENT%7B%5C_%7DID=264207).
10. *Труфанова Е. О.* Субъект и познание в мире социальных конструкций [Текст] / Труфанова Е. О. — Москва : Канон+, 2018. — 320 с.
11. *Дубровский Д. И.* Сознание, мозг, искусственный интеллект [Текст] / Дубровский Д. И. — Москва : Стратегия-Центр, 2007. — 272 с.

12. *Бажанов В. А.* Мозг - культура - социум: кантианская программа в когнитивных исследованиях [Текст] / Бажанов В. А. — Москва : Канон+ РООИ «Реабилитация», 2019. — 288 с.
13. *Князева, Е. Н.* Энактивизм: новая форма конструктивизма в эпистемологии [Текст] / Е. Н. Князева. — Москва, Санкт-Петербург : Центр гуманитарных инициатив: Университетская книга, 2014. — 352 с.
14. *Baryshnikov, P.* Language, brain and computation: from semiotic asymmetry to recursive rules [Текст] / P. Baryshnikov // RUDN Journal of Philosophy. — 2018. — Vol. 22, no. 2. — P. 168—182. — URL: <http://journals.rudn.ru/philosophy/article/view/18712>.
15. *Барышников, П.* Метафорические основания компьютеризации в когнитивных науках и философии сознания [Текст] / П. Барышников // Философия науки и техники. — 2018. — Сент. — Т. 23, № 2. — С. 61—72. — URL: [https://iphras.ru/uplfile/root/biblio/ps/ps23%7B%5C\\_%7D2/61-72.pdf](https://iphras.ru/uplfile/root/biblio/ps/ps23%7B%5C_%7D2/61-72.pdf).
16. *Барышников, П.* Феноменальное и вычислимое в структурах сознания [Текст] / П. Барышников // Вестник РУДН. Серия: ФИЛОСОФИЯ. — 2017. — Т. 21, № 2. — С. 229—239. — URL: <http://journals.rudn.ru/philosophy/article/view/16182>.
17. *Барышников, П.* Эвристика компьютерных аналогий в свете аналитической философии сознания [Текст] / П. Барышников // История и философия науки в эпоху перемен. Сборник научных статей: в 6 томах. / под ред. И. Касавин [и др.]. — Москва : «Русское общество истории и философии науки», 2018. — С. 89—92.
18. *Алексеев, А. Ю.* Общезначимый концепт искусственной потребности как основа общего искусственного интеллекта [Текст] / А. Ю. Алексеев // Философские науки. — 2019. — Дек. — Т. 62, № 11. — С. 111—124. — URL: <https://www.phisci.info/jour/article/view/2855>.
19. *Винник, Д. В.* Квантовые свойства в физической организации мозга: амплификация или нивелировка? [Текст] / Д. В. Винник // Философия науки. — 2020. — № 1. — С. 96—118. — URL: [http://sibran.ru/journals/issue.php?ID=178606&ARTICLE\\_ID=178612](http://sibran.ru/journals/issue.php?ID=178606&ARTICLE_ID=178612).

20. *Анохин, К. В.* Коды вавилонской библиотеки мозга [Текст] / К. В. Анохин // В мире науки. — 2013. — № 5. — С. 82—89.
21. *Анохин, К. В.* Когнитом: в поисках общей теории когнитивной науки [Текст] / К. В. Анохин // Шестая международная конференция по когнитивной науке. — Калининград, 2014. — С. 26—28.
22. *Черниговская, Т. В.* Чеширская улыбка кота Шредингера: язык и сознание [Текст] / Т. В. Черниговская. — Москва : Языки славянской культуры, 2013. — 447 с.
23. *Аллахвердов В. М.* Сознание как парадокс [Текст] / Аллахвердов В. М. — Санкт-Петербург : Издательство ДНК, 2000. — 528 с.
24. *Назаров, А.* Обобщенная модель познавательной деятельности индивида [Текст] / А. Назаров // Психологическая наука и образование. — 2000. — Т. 5, № 3. — С. 40—60.
25. Downregulation of the Posterior Medial Frontal Cortex Prevents Social Conformity [Текст] / V. Klucharev [et al.] // The Journal of Neuroscience. — 2011. — Aug. — Vol. 31, no. 33. — P. 11934—11940. — URL: <http://www.jneurosci.org/content/31/33/11934.abstract>.
26. *Ключарев, В. А.* Нейробиологические механизмы социального влияния [Текст] / В. А. Ключарев, И. П. Зубарев, А. Шестакова // Экспериментальная психология. — 2014. — Т. 7, № 4. — С. 20—36.
27. *Фаликман, М. В.* Парадоксы зрительного внимания : эффекты перцептивных задач [Текст] / М. В. Фаликман. — Москва : Издательский дом ЯСК, 2018. — 252 с.
28. *Utochkin, I. S.* Independent storage of different features of real-world objects in long-term memory. [Текст] / I. S. Utochkin, T. F. Brady // Journal of Experimental Psychology: General. — 2020. — Vol. 149, no. 3. — P. 530—549.
29. *Люсин, Д. В.* Трехмерная модель структуры эмоциональных состояний, основанная на русскоязычных данных [Текст] / Д. В. Люсин // Психология. Журнал Высшей школы экономики. — 2019. — Т. 16, № 2. — С. 341—356.

30. *Спиридонов, В. Ф.* Воплощенное познание (embodied cognition): основные направления исследований [Текст] / В. Ф. Спиридонов, Н. И. Логинов // Вестник СПбГУ. Психология и педагогика. — 2017. — Т. 7, № 4. — С. 343—364.
31. *Олескин, А. В.* Сетевые структуры в биосистемах и человеческом обществе [Текст] / А. В. Олескин. — Москва : УРСС, 2019. — 304 с.
32. *Олескин, А. В.* Социальность микроорганизмов и взаимоотношения в системе микробиота-хозяин: роль нейромедиаторов [Текст] / А. В. Олескин, Б. А. Шендеров, В. С. Роговский. — Москва : Издательский дом МГУ, 2019. — 286 с.
33. *Кузнецов, О. П.* Теория ресурсных сетей [Текст] / О. П. Кузнецов, Л. Ю. Жилиякова. — Москва : РИОР, 2017. — 283 с.
34. *Cicourel, A. V.* Cognitive sociology : language and meaning in social interaction [Текст] / A. V. Cicourel. — New York : Free Press, 1974. — 189 p.
35. *Zerubavel, E.* Social mindscapes : an invitation to cognitive sociology [Текст] / E. Zerubavel. — Cambridge, Mass. : Harvard University Press, 1997. — 164 p.
36. *Cerulo, K.* Cognitive sociology [Текст] / K. Cerulo // Encyclopedia of social theory / ed. by G. Ritzer. — Thousand Oaks, Calif. : SAGE Publications Inc., 2005. — P. 107—111.
37. *DiMaggio, P.* Culture and Cognition [Текст] / P. DiMaggio // Annual Review of Sociology. — 1997. — Aug. — Vol. 23, no. 1. — P. 263—287. — URL: <http://www.annualreviews.org/doi/10.1146/annurev.soc.23.1.263>.
38. The Cognitive Social Network in Dreams: Transitivity, Assortativity, and Giant Component Proportion Are Monotonic [Текст] / H. J. Han [et al.] // Cognitive Science. — 2016. — No. 40. — P. 671—696.
39. *Brands, R. A.* Cognitive social structures in social network research: A review [Текст] / R. A. Brands // Journal of Organizational Behavior. — 2013. — July. — Vol. 34, S1. — S82—S103. — URL: <https://doi.org/10.1002/job.1890>.
40. *Brands, R. A.* The Leader-in-Social-Network Schema: Perceptions of Network Structure Affect Gendered Attributions of Charisma [Текст] / R. A. Brands, J. I. Menges, M. Kilduff // Organization Science. — 2015. — Aug. — Vol. 26, no. 4. — P. 1210—1225. — URL: <http://pubsonline.informs.org/doi/10.1287/orsc.2015.0965>.

41. Social Cognitive Networks Academic Research Center [Текст]. — 2010. — URL: <http://scnarc.rpi.edu> (visited on 06/17/2020).
42. *Cacioppo, J.* Social Neuroscience and its Relationship to Social Psychology [Текст] / J. Cacioppo, G. Berntson, J. Decety // Social cognition. — 2010. — Dec. — Vol. 28. — P. 675–685.
43. *Amodio, D. M.* Meeting of minds: the medial frontal cortex and social cognition [Текст] / D. M. Amodio, C. D. Frith // Nature Reviews Neuroscience. — 2006. — Vol. 7, no. 4. — P. 268–277. — URL: <https://doi.org/10.1038/nrn1884>.
44. *Gazzaniga, M. S.* The Social Brain: Discovering the Networks of the Mind [Текст] / M. S. Gazzaniga //. — New York : Basic Books, 1985. — P. 219.
45. *Rizzolatti, G.* Mirrors in the brain : how our minds share actions and emotions [Текст] / G. Rizzolatti, F. Anderson, C. Sinigaglia. — Oxford : Oxford University Press, 2008. — 242 p.
46. *Churchland, P. M.* Matter and consciousness [Текст] / P. M. Churchland. — Third edit. — Cambridge, Massachusetts : MIT Press, 2013. — 288 pages.
47. *Searle, J. R.* The Rediscovery of the Mind (Representation and Mind) [Текст] / J. R. Searle. — Cambridge, MA : MIT Press, 1992. — 286 p.
48. *Dennett, D. C.* Consciousness Explained [Текст] / D. C. Dennett. — New York ; Boston ; London : Little, Brown, Co., 1991. — 511 p.
49. *Dennett, D. C.* The Part of Cognitive Science That Is Philosophy [Текст] / D. C. Dennett // Topics in Cognitive Science. — 2009. — Vol. 1, no. 2. — P. 231–236.
50. *Chalmers, D. J.* The Character of Consciousness [Текст] / D. J. Chalmers. — New York : Oxford University Press, 2010. — 625 p.
51. *Chalmers, D. J.* Constructing the World [Текст] / D. J. Chalmers. — Oxford : Oxford University Press, 2012. — 521 p.
52. *Clark, A.* Mindware: An Introduction to the Philosophy of Cognitive Science [Текст] / A. Clark. — New York, NY, US : Oxford University Press, 2000. — 112 p.
53. *Clark, A.* Surfing Uncertainty [Текст] / A. Clark. — New York : Oxford University Press, 2016. — 412 p.

54. *Fodor, J. A.* The Language of Thought [Текст] / J. A. Fodor. — Harvard University Press, 1975. — 214 p. — (Language and thought series). — URL: <https://books.google.co.il/books?id=XZwGLBYLbg4C>.
55. Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1: Foundations [Текст] / ed. by D. E. Rumelhart, J. L. McClelland, C. PDP Research Group. — Cambridge, MA, USA : MIT Press, 1986. — 547 p.
56. Parallel Distributed Processing: Explorations in the Microstructure, Vol. 2: Psychological and Biological Models [Текст] / ed. by J. L. McClelland, D. E. Rumelhart, C. PDP Research Group. — Cambridge, MA, USA : MIT Press, 1986. — 611 p.
57. *Churchland, P. M.* A Neurocomputational Perspective: The Nature of Mind and the Structure of Science [Текст] / P. M. Churchland. — Cambridge, Mass. : MIT Press, 1989. — 339 p.
58. *Churchland, P. S.* Touching a Nerve: The Self as a Brain [Текст] / P. S. Churchland. — W. W. Norton & Co., 2013. — 307 p.
59. *Churchland, P. S.* The Computational Brain [Текст] / P. S. Churchland, T. J. Sejnowsky. — Bradford Book, 0216. — 569 p.
60. *Varela, F. J.* The Embodied Mind: Cognitive Science and Human Experience [Текст] / F. J. Varela, E. Rosch, E. Thompson. — MIT Press, 1992. — 328 p. — (The MIT Press). — URL: <https://books.google.co.il/books?id=QY4RoH2z5DoC>.
61. *Hohwy, J.* The Predictive Mind [Текст] / J. Hohwy. — Oxford University Press, 2014. — 237 p.
62. *Clark, A.* A nice surprise? Predictive processing and the active pursuit of novelty [Текст] / A. Clark // Phenomenology and the Cognitive Sciences. — 2018. — Vol. 17, no. 3. — P. 521—534. — URL: <https://doi.org/10.1007/s11097-017-9525-z>.
63. *Friston, K.* Does predictive coding have a future? [Текст] / K. Friston // Nature Neuroscience. — 2018. — Vol. 21, no. 8. — P. 1019—1021.
64. *Мухайлов, И. Ф.* Прошло ли время философии? [Текст] / И. Ф. Михайлов // Вопросы философии. — 2019. — № 1. — С. 15—25.



65. *Витгенштейн, Л.* О достоверности [Текст] / Л. Витгенштейн // *Философские работы. В 2-х частях. Часть I* / под ред. М. С. Козлова. — Москва : Гнозис, 1994. — С. 321—406.
66. *Витгенштейн, Л.* Философские исследования [Текст] / Л. Витгенштейн // *Философские работы. В 2-х частях* / под ред. М. С. Козлова. — Москва : Гнозис, 1994. — С. 75—320.
67. *Витгенштейн, Л.* Логико-философский трактат [Текст] / Л. Витгенштейн // *Философские работы. В 2-х частях* / под ред. М. С. Козлова. — Москва : Гнозис, 1994. — С. 1—74.
68. *Куайн, У. В. О.* О том, что есть [Текст] / У. В. О. Куайн // *С точки зрения логики.* — Томск : Издательство Томского университета, 2003. — С. 7—23.
69. *Almeida, M. B.* Revisiting ontologies: A necessary clarification [Текст] / M. B. Almeida // *Journal of the American Society for Information Science and Technology.* — 2013. — Aug. — Vol. 64, no. 8. — P. 1682—1693. — URL: <https://doi.org/10.1002/asi.22861>.
70. *Михайлов, И. Ф.* Человек, сознание, сети [Текст] / И. Ф. Михайлов. — Москва : ИФРАН, 2015. — С. 196.
71. *Рассел, Б.* Об обозначении [Текст] / Б. Рассел // *Язык, истина, существование.* — Томск : Изд-во ТГУ, 2002. — С. 7—22.
72. *Rosen, G.* Nominalism, Naturalism, Epistemic Relativism [Текст] / G. Rosen // *Noûs.* — 2001. — Oct. — Vol. 35, s15. — P. 69—91. — URL: <https://doi.org/10.1111/0029-4624.35.s15.4>.
73. *Tarski, A.* Logic, Semantics, Metamathematics. 2nd edition. [Текст] / A. Tarski. — Indianapolis : Hackett, 1983. — 506 p.
74. *Михайлов, И. Ф.* К общей онтологии когнитивных и социальных наук [Текст] / И. Ф. Михайлов // *Философия науки и техники.* — 2017. — № 2. — С. 103—119.
75. *Matti, E.* Metaontology [Текст] / E. Matti, M. Eklund // *Philosophy compass.* — 2006. — May. — Vol. 1, no. 3. — P. 317—334.
76. *Berto, F.* Ontology and Metaontology [Текст] / F. Berto, M. Plebani. — London ; New York : A Contemporary Guide, 2015. — 250 p.

77. *Van Inwagen, P.* Meta-Ontology [Текст] / P. Van Inwagen // Erkenntnis. — 1998. — Vol. 48, no. 48. — P. 233—250.
78. *Стёпин, В. С.* Философия науки. Общие проблемы [Текст] / В. С. Стёпин. — Москва : Гардарики, 2006. — 384 с.
79. *Стёпин, В. С.* Теоретическое знание [Текст] / В. С. Стёпин. — Москва : Прогресс-Традиция, 2003. — 744 с.
80. *Стёпин, В. С.* Цивилизация и культура [Текст] / В. С. Стёпин. — Санкт-Петербург : СПбГУП, 2011. — 408 с.
81. *Мирский, Э. М.* Междисциплинарные исследования [Текст] / Э. М. Мирский // Новая философская энциклопедия / под ред. В. С. Стёпин. — Москва : Мысль, 2010. — С. 518.
82. *Касавин, И. Т.* Междисциплинарные исследования и социальная картина мира [Текст] / И. Т. Касавин // Философия науки и техники. — 2014. — Т. 19. — С. 9—26.
83. *Михайлов, И. Ф.* Социальная онтология: время вычислений [Текст] / И. Ф. Михайлов // Вестник Томского государственного университета. Философия. Социология. Политология. — 2020. — № 55. — С. 36—46.
84. *Tony Lawson.* A conception of social ontology [Текст] / Tony Lawson // Social Ontology and Modern Economics / ed. by S. Pratten. — London ; New York : Routledge, 2015. — P. 19—52.
85. *Стёпин, В. С.* Научная картина мира в культуре техногенной цивилизации [Текст] / В. С. Стёпин, Л. Ф. Кузнецова. — Москва : ИФРАН, 1994. — 274 с.
86. *Аристотель.* Метафизика [Текст] / Аристотель // Сочинения в 4 т. Т. 1. — Москва : Мысль, 1976. — С. 63—398.
87. *Аристотель.* Категории [Текст] / Аристотель // Сочинения в 4 т. Т. 2. — Москва : Мысль, 1978. — С. 51—90.
88. *Витгенштейн, Л.* Логико-философский трактат. Пер. и параллельный филос.-семиотич. коммент. В. Руднева [Текст] / Л. Витгенштейн // Логос. — 1999. — Т. 8, № 18. — С. 68—87.
89. *Смирнова, Е.* Возможные миры и понятие «картин мира» [Текст] / Е. Смирнова // Вопросы философии. — 2017. — № 1. — С. 39—49.

90. *Алберт, Д.* Квантовая механика угрожает теории относительности [Текст] / Д. Алберт, Р. Галчен. — 2011. — URL: <http://www.modcos.com/articles.php?id=49> (дата обр. 16.04.2017).
91. *Gruber, T.* A translation approach to portable ontologies [Текст] / Т. Gruber // Knowledge Acquisition. — 1993. — Vol. 5, no. 2. — P. 199—220.
92. Complex Networks [Текст] / ed. by E. Ben-Naim, H. Frauenfelder, Z. Toroczkai. — Berlin, Heidelberg : Springer-Verlag, 2004. — 650 p.
93. Идеальное. — Словари и энциклопедии на Академике [Текст]. — 2014. — URL: <http://dic.academic.ru/dic.nsf/enc3p/136070/> (дата обр. 28.12.2014).
94. *Smart, J. J. C.* The Mind/Brain Identity Theory [Текст] / J. J. C. Smart // The Stanford Encyclopedia of Philosophy (Winter 2014 Edition) / ed. by E. N. Zalta. — 2014. — URL: <http://plato.stanford.edu/archives/win2014/entries/mind-identity/>.
95. *Иванов, Д. В.* Функционализм. Метафизика без онтологии [Текст] / Д. В. Иванов // Эпистемология и философия науки. — 2010. — № 2. — С. 95—111.
96. *Kim, J.* Philosophy of Mind [Текст] / J. Kim. — Colorado : Brown University, Westview Press A Subsidiary of Perseus Books, L.L.C, 1998. — 153 p.
97. *Putnam, H.* Psychological Predicates [Текст] / H. Putnam // Art, Mind, and Religion / ed. by W. H. Capitan, D. D. Merrill. — University of Pittsburgh Press, 1967. — P. 37—48.
98. *Tye, M.* Functionalism and Type Physicalism [Текст] / М. Tye // Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition. — 1983. — Sept. — Vol. 44, no. 2. — P. 161—174. — URL: <http://www.jstor.org/stable/4319625>.
99. *Кант, И.* Критика чистого разума [Текст] / И. Кант // Собрание сочинений в 8-ми томах. Т. 3 / под ред. А. В. Гулыги. — Москва : ЧОРО, 1994. — С. 741.
100. *Russell, B.* The Philosophy of Logical Atomism [Текст] / В. Russell // Logic and Knowledge: Essays 1901—1950 / ed. by R. C. Marsh. — London, 1956. — P. 177—281.

101. *Parikh, R.* Topology and Epistemic Logic [Текст] / R. Parikh, L. Moss, C. Steinsvold // Handbook of Spatial Logics / ed. by M. Aiello, I. Pratt-Hartmann, J. Van Benthem. — Dordrecht : Springer, 2007. — P. 299—341.
102. *Nagel, T.* What is it like to be a bat? [Текст] / T. Nagel // Philosophical Review. — 1974. — Vol. 83, October. — P. 435—450.
103. *Jackson, F.* Epiphenomenal Qualia [Текст] / F. Jackson // The Philosophical Quarterly. — 1982. — Vol. 32, no. 127. — P. 127—136.
104. *Hoffman, D. D.* The Interface Theory of Perception: Natural Selection Drives True Perception To Swift Extinction [Текст] / D. D. Hoffman // Object Categorization: Computer and Human Vision Perspectives / ed. by S. Dickinson [et al.]. — Cambridge : Cambridge University Press, 2009. — P. 148—265.
105. *Champagne, M.* The Semiotic Mind: A Fundamental Theory of Consciousness [Текст] : PhD thesis / Champagne Marc. — York University, Toronto, Ontario, 07/2014.
106. *Михайлов, Ф. Т.* Загадка человеческого Я [Текст] / Ф. Т. Михайлов. — Москва : Политиздат, 1976. — 287 с.
107. *Godwin, D.* Breakdown of the brain’s functional network modularity with awareness [Текст] / D. Godwin, R. L. Barry, R. Marois // Proceedings of the National Academy of Sciences. — 2015. — Mar. — P. 1—21. — URL: <http://www.pnas.org/content/early/2015/03/09/1414466112.abstract>.
108. *De Quinsey, C.* Switched-on Consciousness. Clarifying What It Means [Текст] / C. De Quinsey // Journal of Consciousness Studies. — 2006. — Vol. 13, no. 4. — P. 7—12.
109. *Arp, R.* Consciousness and Awareness. Switched-On Rheostats: A Response to de Quincey [Текст] / R. Arp // Journal of Consciousness Studies. — 2007. — Vol. 14, no. 3. — P. 101—106.
110. A theoretically based index of consciousness independent of sensory processing and behavior [Текст] / A. G. Casali [et al.]. — 08/2013.
111. Intelligence is associated with the modular structure of intrinsic brain networks [Текст] / K. Hilger [et al.] // Scientific Reports. — 2017. — Vol. 7, no. 1. — P. 16088. — URL: <https://doi.org/10.1038/s41598-017-15795-7>.

112. *Arsiwalla, X. D.* Measuring the Complexity of Consciousness [Текст] / X. D. Arsiwala, P. Verschure // *Frontiers in Neuroscience*. — 2018. — June. — Vol. 12, JUN. — arXiv: [1801.03880](https://arxiv.org/abs/1801.03880). — URL: <https://www.frontiersin.org/article/10.3389/fnins.2018.00424/full>.
113. *Brentano, F.* Psychology from an Empirical Standpoint [Текст] / F. Brentano. — London : Routledge, Kegan Paul, 1995. — 350 p.
114. *Чизолм, Р. М.* Формальная структура интенциональности: Метафизическое исследование [Текст] / Р. М. Чизолм // *Логос*. — 2002. — Т. 33, № 2. — С. 40—46.
115. *Гуссерль, Э.* Логические исследования. Т. I: Прологомены к чистой логике [Текст] / Э. Гуссерль. — Москва : Академический проект, 2011. — 256 с.
116. *Фреге, Г.* О смысле и значении [Текст] / Г. Фреге // *Логика и логическая семантика*. — Москва : Аспект Пресс, 2000. — С. 220—246.
117. *Kripke, S.* Naming and Necessity [Текст] / S. Kripke. — Harvard University Press, 1980. — 217 p.
118. *Kaplan, D.* Dthat [Текст] / D. Kaplan // *Syntax and Semantics*. Т. 9 / под ред. P. Cole. — Academic Press, 1978. — Гл. kaplan1978. С. 221—243.
119. *Meinong, A.* On the theory of objects (translation of 'Über Gegenstandstheorie', 1904) [Текст] / A. Meinong // *Realism and the Background of Phenomenology* / ed. by R. Chisholm. — Free Press, 1960. — P. 76—117.
120. *Parsons, T.* Nonexistent Objects [Текст] / T. Parsons. — Yale University Press, 1980. — 280 p.
121. *Mally, E.* Gegenstandstheoretische Grundlagen der Logik und Logistik [Текст] / E. Mally. — Barth, 1912. — 87 p. — (Ergänzungsheft zu Bd 148 der Zeitschrift für Philosophie u. philosophische Kritik). — URL: <https://books.google.co.il/books?id=g65LAAAAYAAJ>.
122. *Zalta, E. N.* Intensional logic and the metaphysics of intentionality [Текст]. Vol. 57 / E. N. Zalta. — Cambridge, Mass., London, Bradford books. The MIT Press, 1988. — 256 p.
123. *Куайн, У. В. О.* Слово и объект [Текст] / У. В. О. Куайн. — Перевод с. — Москва : Логос, Праксис, 2000. — 386 с.

124. *Chisholm, R. M.* Perceiving: A Philosophical Study [Текст]. Vol. 9 / R. M. Chisholm. — Cornell University Press, 1957. — 214 p.
125. *Dennett, D. C.* Content and Consciousness [Текст]. Vol. 20 / D. C. Dennett. — Routledge, 1986. — 264 p.
126. *Dretske, F. I.* The intentionality of cognitive states [Текст] / F. I. Dretske // Midwest Studies in Philosophy. — 1980. — Vol. 5, no. 1. — P. 281–294.
127. *Grice, H. P.* Meaning [Текст] / H. P. Grice // Philosophical Review. — 1957. — Vol. 66, no. 3. — P. 377–388.
128. *Millikan, R. G.* Varieties of Meaning: The 2002 Jean Nicod Lectures [Текст] / R. G. Millikan // Philosophy and Phenomenological Research. — 2007. — Vol. 75, no. 3. — P. 674–681.
129. *Anscombe, G. E. M.* Intention [Текст]. Vol. 57 / G. E. M. Anscombe. — Harvard University Press, 1957. — 106 p.
130. *Searle, J.* Intentionality [Текст]. Vol. 20 / J. Searle. — Oxford University Press, 1983. — 292 p.
131. *Chomsky, N.* New Horizons in the Study of Language and Mind [Текст] / N. Chomsky. — Cambridge University Press, 2000. — 250 p.
132. *Rorty, R.* Philosophy and the Mirror of Nature: Thirtieth-Anniversary Edition [Текст] / R. Rorty, M. Williams, D. Bromwich. — Princeton University Press, 2008. — 472 p.
133. *Block, N.* On a confusion about a function of consciousness [Текст] / N. Block // Behavioral and Brain Sciences. — 1995. — June. — Vol. 18, no. 2. — P. 227–247. — URL: [https://www.cambridge.org/core/product/identifier/S0140525X00038188/type/journal%7B%5C\\_%7Darticle](https://www.cambridge.org/core/product/identifier/S0140525X00038188/type/journal%7B%5C_%7Darticle).
134. *Block, N.* Consciousness, Accessibility, and the Mesh between Psychology and Neuroscience [Текст] / N. Block // Behavioral and Brain Sciences. — 2007. — Vol. 30, no. 5. — P. 481–548.
135. *Деннет, Д. С.* Виды психики: На пути к пониманию сознания [Текст] / Д. С. Деннет. — Москва : Идея-Пресс, 2004. — 184 с.
136. *Goodman, N.* Languages of Art [Текст] / N. Goodman. — Indianapolis : Hackett Publishing Co, Inc, 1976. — 291 p.

137. *Chalmers, D. J.* The representational character of experience [Текст] / D. J. Chalmers // The Future for Philosophy / ed. by B. Leiter. — Oxford University Press, 2004. — P. 153—181.
138. *Harman, G.* The intrinsic quality of experience [Текст] / G. Harman // Philosophical Perspectives. — 1990. — Vol. 4. — P. 31—52.
139. *Block, N.* Mental paint and mental latex [Текст] / N. Block // Philosophical Issues. — 1996. — Vol. 7. — P. 19—49.
140. *Searle, J. R.* Minds, brains, and programs [Текст] / J. R. Searle // Behavioral and Brain Sciences. — 1980. — Vol. 3, no. 3. — P. 417—457.
141. *Jackson, F.* What Mary Didn't Know [Текст] / F. Jackson // Journal of Philosophy. — 1986. — Vol. 83, no. 5. — P. 291—295.
142. *Михайлов, И. Ф.* «Искусственный интеллект» как аргумент в споре о сознании [Текст] / И. Ф. Михайлов // Эпистемология и философия науки. — 2012. — Т. 2. — С. 107—122.
143. *Stoljar, D.* Physicalism [Текст] / D. Stoljar. — London ; New York : Routledge, 2010. — 264 p.
144. *Block, N.* Comparing the major theories of consciousness [Текст] / N. Block // The Cognitive Neurosciences / ed. by M. S. Gazzaniga. — IV0th ed. — MIT Press, 2009. — Chap. 77. P. 1111—1122.
145. *Nath, R.* Experience and Expression: The Inner-Outer Conceptions of Mental Phenomena [Текст] / R. Nath, M. Panda // Indian Philosophical Quarterly. — 2014. — Vol. 1, no. 1—4. — P. 417—457.
146. *Кант, И.* Метафизические начала естествознания [Текст] / И. Кант // Собрание сочинений в 8-ми томах. Т. 4. — Москва : ЧОРО, 1994. — С. 247—372.
147. *Величковский, Б.* Когнитивная наука: Основы психологии познания. В 2 т. Т. 1 [Текст] / Б. Величковский. — Москва : Смысл: Издательский центр «Академия», 2006. — 448 с.
148. *Михайлов, И. Ф.* К онтологии жизненного мира человека: современный взгляд [Текст] / И. Ф. Михайлов // Вопросы социальной теории: научный альманах. 2016 – 2017. — 2017. — Т. IX, № 1/2. — С. 200—211.

149. Augmented Social Cognition [Текст] / Е. Н. Chi [et al.] // AAAI Spring Symposium 2008 — Social Information Processing. Stanford University, CA, USA. AAAI, 2008. — Stanford, 2008. — P. 11–17.
150. *MacLennan, B. J.* Transcending Turing computability [Текст] / В. J. MacLennan // Minds and Machines. — 2003. — Vol. 13, no. 1.
151. *MacLennan, B. J.* Natural computation and non-Turing models of computation [Текст] / В. J. MacLennan // Theoretical Computer Science. — 2004. — Vol. 317, no. 1. — P. 115–145. — URL: <http://www.sciencedirect.com/science/article/pii/S0304397503006352>.
152. *Михайлов, И. Ф.* Как в наше время возможна когнитивная теория общества? [Текст] / И. Ф. Михайлов // Электронный философский журнал Vox. — 2018. — № 25.
153. *Ramsey, W.* Do Connectionist Representations Earn Their Explanatory Keep? [Текст] / W. Ramsey // Mind & Language. — 1997. — Март. — Т. 12, № 1. — С. 34–66. — URL: <https://doi.org/10.1111/j.1468-0017.1997.tb00061.x>.
154. *Barbey, A. K.* Network Neuroscience Theory of Human Intelligence [Текст] / A. K. Barbey // Trends in Cognitive Sciences. — 2018. — Vol. 22, no. 1. — P. 8–20. — URL: <http://www.sciencedirect.com/science/article/pii/S1364661317302218>.
155. *Granovetter, M. S.* The Strength of Weak Ties [Текст] / М. S. Granovetter // Social Networks / ed. by S. B. T. S. N. Leinhardt. — Elsevier, 1977. — P. 347–367. — URL: <http://www.sciencedirect.com/science/article/pii/B9780124424500500250> % 20<https://linkinghub.elsevier.com/retrieve/pii/B9780124424500500250>.
156. Круглый стол «Вычисления в науках о мозге и сознании: атавизм, метафора или эвристика?», 19 декабря 2019 г. [Текст] / В. И. Шалак [и др.]. — 2019. — URL: <https://www.youtube.com/watch?v=Oy7r3ZQuzHY> (дата обр. 14.04.2020).
157. *Михайлов, И. Ф.* Когнитивные вычисления и социальная организация [Текст] / И. Ф. Михайлов // Вопросы философии. — 2020. — № 11. — С. 125–128.



158. *Родин, А. В.* Вычисления в природе и природа вычислений [Текст] / А. В. Родин // Вопросы философии. — 2020. — № 11. — С. 129—132.
159. *Шалак, В. И.* Алгоритмические явления в природе: модель объяснения [Текст] / В. И. Шалак // Вопросы философии. — 2020. — № 11. — С. 120—124.
160. *Гемпель, К. Г.* Логика объяснения [Текст] / К. Г. Гемпель. — Москва : Дом Интеллектуальной Книги, 1998. — С. 237.
161. *Аристотель.* Метеорология [Текст] / Аристотель // Аристотель. Сочинения в четырёх томах. — Москва : Издательство "Мысль", 1981. — С. 441—558.
162. *Einstein, A.* Ether and the Theory of Relativity [Текст] / A. Einstein. — URL: [http://mathshistory.st-andrews.ac.uk/Extras/Einstein\\_ether.html](http://mathshistory.st-andrews.ac.uk/Extras/Einstein_ether.html) (дата обр. 12.04.2020).
163. *Kargon, R.* The Decline of the Caloric Theory of Heat: A Case Study [Текст] / R. Kargon // Centaurus. — 1964. — Т. 10, № 1. — С. 35—39.
164. *Ladyman, J.* Structural realism versus standard scientific realism: The case of phlogiston and dephlogisticated air [Текст] / J. Ladyman // Synthese. — 2011. — Т. 180, № 2. — С. 87—101. — URL: <https://www.jstor.org/stable/41477546>.
165. *Schurz, G.* Structural correspondence, indirect reference, and partial truth: Phlogiston theory and Newtonian mechanics [Текст] / G. Schurz // Synthese. — 2011. — Т. 180, № 2. — С. 103—120. — URL: <https://www.jstor.org/stable/41477547>.
166. Contribution by Lazare and Sadi Carnot to the caloric theory of heat and its inspirative role in thermodynamics [Текст] / J. Šesták [и др.] // Journal of Thermal Analysis and Calorimetry. — 2009. — Август. — Т. 97, № 2. — С. 679—683.
167. *Turing, A. M.* On computable numbers, with an application to the entscheidungsproblem. a correction [Текст] / A. M. Turing // Proceedings of the London Mathematical Society. — 1938. — Vol. s2—43, no. 1. — P. 544—546.
168. *Pierce, C. S.* Logic as Semiotic: the Theory of Signs [Текст] / C. S. Pierce // Philosophical Writings of Peirce / под ред. J. Buchler. — New York : Dover Publications, Inc., 2011. — С. 98—119.

169. *Copeland, B. J.* What is computation? [Текст] / B. J. Copeland // Synthese. — 1996. — Vol. 108, no. 3. — P. 335—359. — URL: <https://doi.org/10.1007/BF00413693>.
170. *Voevodsky, V. A.* Mathematics and the outside world [Текст] / V. A. Voevodsky. — 2003. — URL: [https://www.math.ias.edu/vladimir/sites/math.ias.edu.vladimir/files/Maths\\_and\\_the\\_outside\\_world.pdf](https://www.math.ias.edu/vladimir/sites/math.ias.edu.vladimir/files/Maths_and_the_outside_world.pdf) (дата обр. 07.04.2020).
171. *Marcus, G.* The atoms of neural computation [Текст] / G. Marcus, A. Marblestone, T. Dean // Science. — 2014. — Oct. — Vol. 346, no. 6209. — 551 LP —552. — URL: <http://science.sciencemag.org/content/346/6209/551.abstract>.
172. *Marr, D.* Vision: A Computational Investigation into the Human Representation and Processing of Visual Information [Текст] / D. Marr. — The MIT Press, 2010. — 432 p. — URL: <https://mitpress.universitypressscholarship.com/10.7551/mitpress/9780262514620.001.0001/upso-9780262514620>.
173. *Poggio, T.* The Levels of Understanding Framework, Revised [Текст] / T. Poggio // Perception. — 2012. — Jan. — Vol. 41, no. 9. — P. 1017—1023. — URL: <https://doi.org/10.1068/p7299>.
174. *Craver, C.* Mechanism [Текст] / C. Craver, W. Bechtel // The Philosophy of Science: An Encyclopedia / ed. by J. Pfeifer, S. Sarkar. — Psychology Press, 2006. — P. 469—478.
175. *Miłkowski, M.* Beyond Formal Structure: A Mechanistic Perspective on Computation and Implementation [Текст] / M. Miłkowski // The Journal of Cognitive Science. — 2011. — Vol. 12. — P. 359—379.
176. *Miłkowski, M.* A Mechanistic Account of Computational Explanation in Cognitive Science and Computational Neuroscience [Текст] / M. Miłkowski //. — 01/2016. — P. 191—205.
177. The thermodynamic efficiency of computations made in cells across the range of life [Текст] / C. P. Kempes [et al.] // Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. — 2017. — Dec. — Vol. 375, no. 2109. — P. 20160343. — URL: <https://doi.org/10.1098/rsta.2016.0343>.

178. *Hutto, D. D.* Representation Reconsidered [Текст] / D. D. Hutto // *Philosophical Psychology*. — 2011. — Feb. — Vol. 24, no. 1. — P. 135–139. — URL: <https://doi.org/10.1080/09515089.2010.529261>.
179. *Hutto, D. D.* Radicalizing Enactivism: Basic Minds Without Content [Текст] / D. D. Hutto, E. Myin. — MIT Press, 2013. — 206 p.
180. *Smith, L. B.* Development as a dynamic system [Текст] / L. B. Smith, E. Thelen // *Trends in Cognitive Sciences*. — 2003. — Vol. 7, no. 8. — P. 343–348. — URL: <http://www.sciencedirect.com/science/article/pii/S1364661303001566>.
181. *Thelen, E.* Connectionism and dynamic systems: are they really different? [Текст] / E. Thelen, E. Bates // *Developmental Science*. — 2003. — Vol. 6, no. 4. — P. 378–391. — URL: <http://dx.doi.org/10.1111/1467-7687.00294>.
182. *Гоббс, Т.* Левиафан [Текст] / Т. Гоббс. — Москва : Мысль, 2001. — 478 с.
183. *Turing, A. M.* On computable numbers, with an application to the entscheidungsproblem [Текст] / A. M. Turing // *Proceedings of the London Mathematical Society*. — 1937. — Vol. s2–42, no. 1. — P. 230–265.
184. *Newell, A.* Computer Science as Empirical Inquiry: Symbols and Search [Текст] / A. Newell, H. A. Simon // *Commun. ACM*. — New York, NY, USA, 1976. — Mar. — Vol. 19, no. 3. — P. 113–126. — URL: <https://doi.org/10.1145/360018.360022>.
185. *Pylyshyn, Z. W.* Computation and cognition: Toward a foundation for cognitive science. [Текст] / Z. W. Pylyshyn. — Cambridge, MA, US : The MIT Press, 1986. — 292 p.
186. *Anderson, J. R.* The Architecture of Cognition [Текст] / J. R. Anderson. — Hillsdale, NJ, US : Harvard University Press, 1983. — 345 p. — (Cognitive science series.)
187. *Fresco, N.* The Explanatory Role of Computation in Cognitive Science [Текст] / N. Fresco // *Minds and Machines*. — 2012. — Vol. 22, no. 4. — P. 353–380. — URL: <http://dx.doi.org/10.1007/s11023-012-9286-y><http://link.springer.com/article/10.1007%7B%5C%7D2Fs11023-012-9286-y><https://doi.org/10.1007/s11023-012-9286-y>.

188. *Поспелов, Д. А.* Моделирование рассуждений. Опыт анализа мыслительных актов [Текст] / Д. А. Поспелов. — Москва : Радио и связь, 1989. — 184 с.
189. *Horgan, T. E.* Connectionism and the Philosophical Foundations of Cognitive Science [Текст] / Т. Е. Horgan // *Metaphilosophy*. — 1997. — Vol. 28, no. 1/2. — P. 1—30.
190. *Михайлов, И. Ф.* Социально-сетевые аспекты проблемы выбора [Текст] / И. Ф. Михайлов // Человек перед выбором в современном мире: проблемы, возможности, решения. Материалы Всероссийской научной конференции 27–28 октября 2015 г. ИФ РАН, Москва. — Москва : Научная мысль, 2015. — С. 48–57.
191. *Clark, A.* Radical predictive processing [Текст] / A. Clark // *Southern Journal of Philosophy*. — 2015. — Sept. — Vol. 53, S1. — P. 3–27.
192. *Friston, K.* Prediction, perception and agency [Текст] / K. Friston. — 02/2012.
193. *Friston, K.* A Duet for one [Текст] / K. Friston, C. Frith // *Consciousness and Cognition*. — 2015. — Нояб. — Т. 36. — С. 390–405. — URL: <https://linkinghub.elsevier.com/retrieve/pii/S105381001400230X>.
194. *Hohwy, J.* New directions in predictive processing [Текст] / J. Hohwy // *Mind & Language*. — 2020. — Apr. — Vol. 35, no. 2. — P. 209–223. — URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/mila.12281>.
195. *Helmholtz, H. von.* Treatise on Physiological Optics [Текст] / H. von Helmholtz. — Dover Publications, 2013. — 175 p. — (Dover Books on Physics). — URL: <https://books.google.co.il/books?id=cSjEAgAAQBAJ>.
196. The free energy principle for action and perception: A mathematical review [Текст] / C. L. Buckley [et al.]. — 2017. — arXiv: [1705.09156](https://arxiv.org/abs/1705.09156).
197. The Helmholtz Machine [Текст] / P. Dayan [et al.] // *Neural Computation*. — 1995. — Sept. — Vol. 7, no. 5. — P. 889–904. — URL: <https://doi.org/10.1162/neco.1995.7.5.889>.
198. *Egiazaryan, G. G.* Theory of functional systems in the scientific school of P.K. Anokhin. [Текст] / G. G. Egiazaryan, K. V. Sudakov // *Journal of the history of the neurosciences*. — 2007. — Т. 16 1–2. — С. 194–205.

199. *Витяев, Е.* Формализация когнитива [Текст] / Е. Витяев // Нейроинформатика. — 2016. — Т. 9, № 126. — С. 26—36.
200. Action and behavior: A free-energy formulation [Текст] / К. J. Friston [et al.] // Biological Cybernetics. — 2010. — Mar. — Vol. 102, no. 3. — P. 227—260.
201. *Friston, K.* Action understanding and active inference [Текст] / К. Friston, J. Mattout, J. Kilner // Biological Cybernetics. — 2011. — Feb. — Vol. 104, no. 1/2. — P. 137—160.
202. Active inference and learning [Текст] / К. Friston [et al.]. — 09/2016.
203. *Friston, K.* The free-energy principle: a rough guide to the brain? [Текст] / К. Friston // Trends in Cognitive Sciences. — 2009. — July. — Vol. 13, no. 7. — P. 293—301. — URL: <https://linkinghub.elsevier.com/retrieve/pii/S136466130900117X>.
204. *Metzinger, T.* Philosophy & Predictive Processing [Текст] / Т. Metzinger, W. Wiese. — URL: <https://predictive-mind.net> (visited on 05/30/2020).
205. *Friston, K. J.* Active inference, Communication and hermeneutics [Текст] / К. J. Friston, C. D. Frith // Cortex. — 2015. — July. — Vol. 68. — P. 129—143.
206. *Фритт, К.* Мозг и душа: Как нервная деятельность формирует наш внутренний мир. — Пер. с англ. П. Петрова [Текст] / К. Фритт. — Москва : Астрель: CORPUS, 2010. — 340 с.
207. *Анохин, П. К.* Принципиальные вопросы общей теории функциональных систем [Текст] / П. К. Анохин // Философские аспекты теории функциональной системы. — Москва : Издательство "Наука", 1978. — С. 49—106.
208. Information Entropy As a Basic Building Block of Complexity Theory [Текст] / J. Gao [и др.] // Entropy. — 2013. — Авг. — Т. 15, № 12. — С. 3396—3418. — URL: <http://www.mdpi.com/1099-4300/15/9/3396>.
209. *Vityaev, E.* Cognitive architecture based on the functional systems theory [Текст] / E. Vityaev, A. Demin // Procedia Computer Science. — 2018. — Vol. 145. — P. 623—628. — URL: <https://linkinghub.elsevier.com/retrieve/pii/S1877050918323603>.

210. *Wolfram, S.* A New Kind of Science [Текст] / S. Wolfram. — Wolfram Media Inc., 01/2002. — P. 1192.
211. *Piccinini, G.* Neural Computation and the Computational Theory of Cognition [Текст] / G. Piccinini, S. Bahar // Cognitive Science. — 2013. — Vol. 37, no. 3. — P. 453—488.
212. *McDermott, D. V.* Mind and Mechanism [Текст] / D. V. McDermott. — MIT Press, 2001. — 262 p. — (Bradford Bks). — URL: <https://books.google.com/books?id=aSm4BhlmHYEC>.
213. *Marr, D.* From Understanding Computation to Understanding Neural Circuitry [Текст] / D. Marr, T. Poggio // Neurosciences Research Program Bulletin. — 1979. — Vol. 15, no. 3. — P. 470—488.
214. *Goldstein, L.* Wittgenstein, semantics and connectionism [Текст] / L. Goldstein, H. Slater // Philosophical Investigations. — 1998. — Vol. 21, no. 4. — P. 293—314.
215. *Hogarth, M.* Non-Turing Computers and Non-Turing Computability [Текст] / M. Hogarth // PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association, vol. 1: Contributed Papers. — 1994. — T. 1994. — C. 126—138.
216. *Eberbach, E.* Turing's Ideas and Models of Computation BT - Alan Turing: Life and Legacy of a Great Thinker [Текст] / E. Eberbach, D. Goldin, P. Wegner // / ed. by C. Teuscher. — Berlin, Heidelberg : Springer Berlin Heidelberg, 2004. — P. 159—194. — URL: [https://doi.org/10.1007/978-3-662-05642-4\\_7B%5C\\_%7D7](https://doi.org/10.1007/978-3-662-05642-4_7B%5C_%7D7).
217. *Copeland, M.* What's the Difference Between Artificial Intelligence, Machine Learning, and Deep Learning? [Текст] / M. Copeland. — 2016. — URL: <https://blogs.nvidia.com/blog/2016/07/29/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai/> (visited on 08/05/2018).
218. *Fodor, J. A.* Connectionism and cognitive architecture: a critical analysis [Текст] / J. A. Fodor, Z. W. Pylyshyn // Cognition. — 1988. — T. 28, № 1/2. — C. 3—71. — URL: <http://www.ncbi.nlm.nih.gov/pubmed/2450716>.

219. *Smolensky, P.* On the proper treatment of connectionism [Текст] / P. Smolensky // Behavioral and Brain Sciences. — 1988. — Mar. — Vol. 11, no. 1. — P. 1–23. — URL: [https://www.cambridge.org/core/product/identifier/S0140525X00052432/type/journal%7B%5C\\_%7Darticle](https://www.cambridge.org/core/product/identifier/S0140525X00052432/type/journal%7B%5C_%7Darticle).
220. *Hutto, D. D.* Looking beyond the brain: Social neuroscience meets narrative practice. [Текст] / D. D. Hutto, M. D. Kirchhoff // Cognitive Systems Research. — Hutto, Daniel D.: School of Humanities, Social Inquiry, University of Wollongong, Building 19, Northfields Avenue, Wollongong, NSW, Australia, 2522, ddhutto@uow.edu.au, 2015. — Т. 34/35. — С. 5–17.
221. The Cognitive Basis of Computation: Putting Computation in its place [Текст] / D. D. Hutto [et al.] // The Routledge Handbook of the Computational Mind / ed. by M. Sprevak, M. Colombo. — London : Routledge, 2018. — P. 272–282.
222. *Gärtner, K.* Enactivism, Radical Enactivism and Predictive Processing: What is Radical in Cognitive Science? [Текст] / K. Gärtner, R. W. Clowes // Kairos. Journal of Philosophy & Science. — Berlin, 2017. — Apr. — Vol. 18, no. 1. — P. 54–83. — URL: <https://content.sciendo.com/view/journals/kjps/18/1/article-p54.xml%20http://content.sciendo.com/view/journals/kjps/18/1/article-p54.xml>.
223. *Chomsky, N.* A review of B. F. Skinner's Verbal Behavior [Текст] / N. Chomsky // Language. — 1959. — Vol. 35, no. 1. — P. 26–58.
224. *Fodor, J. A.* LOT 2: The Language of Thought Revisited [Текст] / J. A. Fodor. — Oxford : Oxford University Press, 2008. — 240 p.
225. *Attneave, F.* In defense of homunculi [Текст] / F. Attneave // Sensory communication: Contributions to the symposium on principles of sensory communication, July 19-Aug. 1, 1959, Endicott House. — Cambridge, MA : MIT Press, 1961. — P. 777–782.
226. Reconstructing speech from human auditory cortex [Текст] / B. N. Pasley [et al.] // PLoS Biology. — 2012. — Jan. — Vol. 10, no. 1. — e1001251.
227. *Сахаров, Д. А.* Биологический субстрат генерации поведенческих актов [Текст] / Д. А. Сахаров // Журнал общей биологии. — 2012. — Т. 73, № 5. — С. 324–348.

228. *Базян, А.* Кодирование эмоциональных состояний мозгом животных: молекулярно-химический код [Текст] / А. Базян // Известия ТРТУ. — 2005. — № 47. — С. 164–169.
229. *Сварник, О. Е.* Активность мозга: Специализация нейрона и дифференциация опыта [Текст] / О. Е. Сварник. — Москва : Изд-во "Институт психологии РАН", 2016. — 190 с.
230. *Cash, S. S.* The Emergence of Single Neurons in Clinical Neurology [Текст] / S. S. Cash, L. R. Hochberg // Neuron. — 2015. — Apr. — Vol. 86, no. 1. — P. 79–91. — URL: <https://doi.org/10.1016/j.neuron.2015.03.058>.
231. *Middlebrooks, P. G.* Studying metacognitive processes at the single neuron level [Текст] / P. G. Middlebrooks, Z. M. Abzug, M. A. Sommer // The cognitive neuroscience of metacognition / ed. by S. M. Fleming, C. D. Frith. — Berlin, Heidelberg : Springer-Verlag Publishing, 2014. — P. 225–244. — URL: [https://doi.org/10.1007/978-3-642-45190-4%7B%5C\\_%7D10](https://doi.org/10.1007/978-3-642-45190-4%7B%5C_%7D10).
232. *Proskurkin, I. S.* Experimental verification of an opto-chemical “neuro-computer” [Текст] / I. S. Proskurkin, P. S. Smelov, V. K. Vanag // Physical Chemistry Chemical Physics. — 2020. — Vol. 22, no. 34. — P. 19359–19367. — URL: <http://xlink.rsc.org/?DOI=D0CP01858A>.
233. *Bhalla, U.* Emergent properties of networks of biological signaling pathways [Текст] / U. Bhalla, R. Iyengar // Science. — 1999. — No. 283. — P. 381–387.
234. Free-energy minimization in joint agent-environment systems: A niche construction perspective [Текст] / J. Bruineberg [и др.] // Journal of Theoretical Biology. — 2018. — Т. 455.
235. The Active Inference Approach to Ecological Perception: General Information Dynamics for Natural and Artificial Embodied Cognition [Текст] / A. Linson [et al.] // Frontiers in Robotics and AI. — 2018. — Mar. — Vol. 5, MAR. — URL: <http://journal.frontiersin.org/article/10.3389/frobt.2018.00021/full>.
236. *Goodman, N.* Steps toward a constructive nominalism [Текст] / N. Goodman, W. V. Quine // Journal of Symbolic Logic. — 1947. — Dec. — Vol. 12, no. 4. — P. 105–122. — URL: [https://www.cambridge.org/core/product/identifier/S002248120007540X/type/journal%7B%5C\\_%7Darticle](https://www.cambridge.org/core/product/identifier/S002248120007540X/type/journal%7B%5C_%7Darticle).



237. *Putnam, H.* Mathematics and the existence of abstract entities [Текст] / H. Putnam // Philosophical Studies. — 1956. — Dec. — Vol. 7, no. 6. — P. 81—88. — URL: <http://link.springer.com/10.1007/BF02221758>.
238. *Henkin, L.* Nominalistic Analysis of Mathematical Language [Текст] / L. Henkin // Studies in Logic and the Foundations of Mathematics. T. 44. — 1966. — С. 187—193. — URL: <https://linkinghub.elsevier.com/retrieve/pii/S0049237X09705853>.
239. *Stegmüller, W.* Ontology and Analyticity [Текст] / W. Stegmüller // Collected Papers on Epistemology, Philosophy of Science and History of Philosophy. — Dordrecht : Springer Netherlands, 1977. — P. 213—238. — URL: [http://link.springer.com/10.1007/978-94-010-1129-7%7B%5C\\_%7D6](http://link.springer.com/10.1007/978-94-010-1129-7%7B%5C_%7D6).
240. *Minogue, B. P.* Numbers, properties, and Frege [Текст] / B. P. Minogue // Philosophical Studies. — 1977. — June. — Vol. 31, no. 6. — P. 423—427. — URL: <http://link.springer.com/10.1007/BF01857033>.
241. *Stegmüller, W.* The Problem of Universals Then and Now [Текст] / W. Stegmüller // Collected Papers on Epistemology, Philosophy of Science and History of Philosophy. — Dordrecht : Springer Netherlands, 1977. — P. 1—65. — URL: [http://link.springer.com/10.1007/978-94-010-1129-7%7B%5C\\_%7D1](http://link.springer.com/10.1007/978-94-010-1129-7%7B%5C_%7D1).
242. *Resnik, M. D.* Ontology and logic: Remarks on hartry field's anti-platonist philosophy of mathematics [Текст] / M. D. Resnik // History and Philosophy of Logic. — 1985. — Jan. — Vol. 6, no. 1. — P. 191—209. — URL: <http://www.tandfonline.com/doi/abs/10.1080/01445348508837082>.
243. *Resnik, M. D.* How nominalist is Hartry Field's nominalism? [Текст] / M. D. Resnik // Philosophical Studies. — 1985. — Mar. — Vol. 47, no. 2. — P. 163—181. — URL: <http://link.springer.com/10.1007/BF00354144>.
244. *Goodman, N.* A study of qualities [Текст] : PhD thesis / Goodman Nelson. — 1990. — P. ix, 657. — URL: [http://wwwlib.umi.com/pqdd2/search/doi?query=au\(Goodman,%20Nelson\)%20and%20da\(1940\)](http://wwwlib.umi.com/pqdd2/search/doi?query=au(Goodman,%20Nelson)%20and%20da(1940)).
245. *Sun, R.* Prolegomena to Cognitive Social Sciences [Текст] / R. Sun // Grounding social sciences in cognitive sciences / ed. by R. Sun. — Cambridge, MA : MIT Press, 2012. — P. 3—32.

246. *Kim, J.* Supervenience and Mind [Текст] / J. Kim ; ed. by E. Sosa. — New York, NY, US : Cambridge University Press, 11/1993. — 377 p. — (Cambridge studies in philosophy.) — URL: <https://www.cambridge.org/core/product/identifier/9780511625220/type/book>.
247. *Houng, A. Y.* Levels of analysis: philosophical issues [Текст] / A. Y. Houng // Wiley Interdisciplinary Reviews: Cognitive Science. — 2012. — Т. 3, № 3. — С. 315—325. — URL: <http://dx.doi.org/10.1002/wcs.1179> %20<http://onlinelibrary.wiley.com/doi/10.1002/wcs.1179/abstract>.
248. *Ellis, G.* Top-down effects in the brain [Текст] / G. Ellis // Physics of Life Reviews. — 2019. — Vol. 31. — P. 11—27. — URL: <http://www.sciencedirect.com/science/article/pii/S1571064518300794>.
249. *Кастельс, М.* Информационная эпоха: экономика, общество и культура [Текст] / М. Кастельс. — Москва : Изд. дом ГУ-ВШЭ, 2000. — 608 с.
250. *Межуев, Б.* Вступление к выпуску "Мануэль Кастельс: Власть идентичности в сетевом обществе после конца тысячелетия [Текст] / Б. Межуев. — 2002. — URL: <http://www.archipelag.ru/geoeconomics/soobshestva/power-identity/introduction/> (дата обр. 13.10.2013).
251. *Dijk, J. A. G. M. van.* The one-dimensional network society of Manuel Castells [Текст] / J. A. G. M. van Dijk // New media & society. — 1999. — Vol. 1, no. 1. — P. 127—138.
252. *Castels, M.* The Rise of the Network Society, The Information Age: Economy, Society and Culture, Vol. I [Текст] / M. Castels. — Cambridge, MA; Oxford, UK : Blackwell, 1996. — 556 p.
253. *Castels, M.* The Power of Identity, The Information Age: Economy, Society and Culture, Vol. II [Текст] / M. Castels. — Cambridge, MA; Oxford, UK : Blackwell, 1997. — 461 p.
254. *Castels, M.* The End of the Millennium, The Information Age: Economy, Society and Culture, Vol. III [Текст] / M. Castels. — Cambridge, MA; Oxford, UK : Blackwell, 1997. — 418 p.

255. *Castels, M.* The Network Society: from Knowledge to Policy [Текст] / М. Castels // The Network Society: From Knowledge to Policy / ed. by M. Castels, G. Cardoso. — Washington, DC : Johns Hopkins Center for Transatlantic Relations, 2005. — P. 3—21.
256. *Castels, M.* Informationalism, Networks, and the Network Society: A Theoretical Blueprint [Текст] / М. Castels // The network society: a cross-cultural perspective. — Northampton, MA : Edward Elgar, 2004. — P. 3—48.
257. War, space, and the evolution of Old World complex societies [Текст] / P. Turchin [et al.] // Proceedings of the National Academy of Sciences. — 2013. — Sept. — P. 201308825. — URL: <http://www.pnas.org/content/early/2013/09/20/1308825110.abstract>.
258. *Михайлов, И. Ф.* Человек в сетевом обществе [Текст] / И. Ф. Михайлов // Новое в науках о человеке. К 85-летию со дня рождения академика И. Т. Фролова / под ред. Г. Л. Белкина, М. И. Фролова. — Москва : ЛЕЛАНД, 2015. — С. 364—387.
259. *Михайлов, И. Ф.* Коммуникация и онтология мышления [Текст] / И. Ф. Михайлов // Человек. — 2015. — № 6. — С. 23—31.
260. *Михайлов, И. Ф.* К гиперсетевой теории сознания [Текст] / И. Ф. Михайлов // Вопросы философии. — 2015. — № 11. — С. 87—98.
261. *Евтушенко, Г.* Метасетевой подход к идентификации ошибок искусственных нейронных сетей [Текст] / Г. Евтушенко, К. Отраднов // Теоретические и прикладные аспекты современной науки: сборник научных трудов по материалам VII Международной научно-практической конференции 31 января 2015 г.: в 10 ч. Часть III. — Белгород : ИП Петрова М.Г, 2015. — С. 32—35.
262. *Stich, S. P.* From folk psychology to cognitive science: The case against belief. [Текст] / S. P. Stich. — Cambridge, MA, US : The MIT Press, 1983. — 266 p.
263. Network science of biological systems at different scales: A review [Текст] / M. Gosak [et al.] // Physics of Life Reviews. — 2018. — Vol. 24. — P. 118—135.
264. Grounded Language Learning Fast and Slow [Текст] / F. Hill [и др.]. — 2020. — Сент. — arXiv: 2009.01719. — URL: <https://arxiv.org/abs/2009.01719>.

265. Language for action: Motor resonance during the processing of human and robotic voices [Текст] / G. Di Cesare [et al.] // Brain and Cognition. — 2017. — Nov. — Vol. 118. — P. 118–127.
266. *Lupyan, G.* Linguistically Modulated Perception and Cognition: The Label-Feedback Hypothesis [Текст] / G. Lupyan // Frontiers in Psychology. — 2012. — Vol. 3. — P. 1–54. — URL: <https://www.frontiersin.org/article/10.3389/fpsyg.2012.00054><http://journal.frontiersin.org/article/10.3389/fpsyg.2012.00054/abstract>.
267. The function of words: Distinct neural correlates for words denoting differently manipulable objects [Текст] / S. A. Rueschemeyer [et al.] // Journal of Cognitive Neuroscience. — 2010. — Aug. — Vol. 22, no. 8. — P. 1844–1851.
268. *Schonbein, W.* The Linguistic Subversion of Mental Representation [Текст] / W. Schonbein // Minds and Machines. — 2012. — Vol. 22, no. 3. — P. 235–262. — URL: <http://dx.doi.org/10.1007/s11023-012-9275-1><http://link.springer.com/article/10.1007%7B%5C%7D2Fs11023-012-9275-1>.
269. *Lazebnik, Y.* Can a biologist fix a radio?—Or, what I learned while studying apoptosis [Текст] / Y. Lazebnik // Cancer Cell. — 2002. — Sept. — Vol. 2, no. 3. — P. 179–182. — URL: <https://linkinghub.elsevier.com/retrieve/pii/S1535610802001332>.
270. *May, R. M.* Uses and Abuses of Mathematics in Biology [Текст] / R. M. May // Science. — 2004. — Vol. 303, no. 5659. — P. 790–793.
271. *Turing, A. M.* The Chemical Basis of Morphogenesis [Текст] / A. M. Turing // The Essential Turing: Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life: Plus The Secrets of Enigma / ed. by B. J. Copeland. — Oxford : Oxford University Press, 2004. — Chap. 15. P. 519–561.
272. *Bascompte, J.* Biology and mathematics [Текст] / J. Bascompte // Arbor. — 2007. — Vol. 182, no. 725. — P. 347–351.
273. *Hofmeyr, J.-H. S.* Mathematics and biology [Текст] / J.-H. S. Hofmeyr // South African Journal of Science. — 2017. — Mar. — Vol. Volume 113, Number 3/4. — URL: <http://sajs.co.za/article/view/3628>.

274. *Bhalla and Ravi Iyengar, U. S.* Emergent Properties of Networks of Biological Signaling Pathways [Текст] / U. S. Bhalla and Ravi Iyengar // Science. — 1999. — Jan. — Vol. 283, no. 5400. — P. 381—387. — URL: <https://www.sciencemag.org/lookup/doi/10.1126/science.283.5400.381>.
275. *Csete, M. E.* Reverse Engineering of Biological Complexity [Текст] / M. E. Csete // Science. — 2002. — Mar. — Vol. 295, no. 5560. — P. 1664—1669. — URL: <https://www.sciencemag.org/lookup/doi/10.1126/science.1069981>.
276. *Bray, D.* Protein molecules as computational elements in living cells [Текст] / D. Bray // Nature. — 1995. — July. — Vol. 376, no. 6538. — P. 307—312. — URL: <http://www.nature.com/articles/376307a0>.
277. A Multi-scale View of the Emergent Complexity of Life: A Free-Energy Proposal [Текст] / C. Hesp [et al.] // Springer Proceedings in Complexity. — Springer, 2019. — P. 195—227. — URL: [http://link.springer.com/10.1007/978-3-030-00075-2%7B%5C\\_%7D7](http://link.springer.com/10.1007/978-3-030-00075-2%7B%5C_%7D7).
278. *Es, T. van.* Living models or life modelled? On the use of models in the free energy principle [Текст] / T. van Es // Adaptive Behavior. — 2020. — May. — P. 1059712320918678. — URL: <https://doi.org/10.1177/1059712320918678>.
279. Morphogenesis as Bayesian inference: A variational approach to pattern formation and control in complex biological systems [Текст] / F. Kuchling [et al.] // Physics of Life Reviews. — 2019. — June. — URL: <https://linkinghub.elsevier.com/retrieve/pii/S1571064519300909>.
280. *Hulme, O. J.* Neurocomputational theories of homeostatic control [Текст] / O. J. Hulme, T. Morville, B. Gutkin // Physics of Life Reviews. — 2019. — Vol. 31. — P. 214—232. — URL: <http://www.sciencedirect.com/science/article/pii/S1571064519301009>.
281. Variational ecology and the physics of sentient systems [Текст] / M. J. D. Ramstead [et al.] // Physics of Life Reviews. — 2019. — Vol. 31. — P. 188—205. — URL: <http://www.sciencedirect.com/science/article/pii/S157106451930003X>.

282. *Ramstead, M. J. D.* Answering Schrödinger's question: A free-energy formulation [Текст] / M. J. D. Ramstead, P. B. Badcock, K. J. Friston // *Physics of Life Reviews*. — 2018. — Mar. — Vol. 24. — P. 1–16. — URL: <http://www.sciencedirect.com/science/article/pii/S1571064517301409%20https://linkinghub.elsevier.com/retrieve/pii/S1571064517301409>.
283. *Kirmayer, L. J.* Ontologies of life: From thermodynamics to teleonomics: Comment on “Answering Schrödinger's question: A free-energy formulation” by Maxwell James Désormeau Ramstead et al. [Текст] / L. J. Kirmayer // *Physics of Life Reviews*. — 2018. — Vol. 24. — P. 29–31. — URL: <http://www.sciencedirect.com/science/article/pii/S1571064517301860>.
284. *Friston, K.* Life as we know it [Текст] / K. Friston // *Journal of The Royal Society Interface*. — 2013. — Sept. — Vol. 10, no. 86. — P. 20130475. — URL: <https://royalsocietypublishing.org/doi/10.1098/rsif.2013.0475>.
285. Knowing one's place: a free-energy approach to pattern regulation [Текст] / K. Friston [et al.] // *Journal of The Royal Society Interface*. — 2015. — Apr. — Vol. 12, no. 105. — P. 20141383. — URL: <https://royalsocietypublishing.org/doi/10.1098/rsif.2014.1383>.
286. *Martyushev, L. M.* Living systems do not minimize free energy [Текст] / L. M. Martyushev // *Physics of Life Reviews*. — 2018. — Mar. — Vol. 24. — P. 40–41. — URL: <http://www.sciencedirect.com/science/article/pii/S1571064517301574%20https://linkinghub.elsevier.com/retrieve/pii/S1571064517301574>.
287. *Auletta, G.* Information and Metabolism in Bacterial Chemotaxis [Текст] / G. Auletta // *Entropy*. — 2013. — Jan. — Vol. 15, no. 1. — P. 311–326. — URL: <http://www.mdpi.com/1099-4300/15/1/311>.
288. *Tremlin, T.* Minds and gods: The cognitive foundations of religion [Текст] / T. Tremlin. — Oxford : Oxford University Press, 2006. — 222 p.
289. *Singh, M.* Multiagent Systems. A Theoretical Framework for Intentions, Know-How, and Communications [Текст] / M. Singh. — Berlin, Heidelberg : Springer-Verlag, 1994. — 168 p.
290. *Knoll, J.* The Brain and Its Self. A Neurochemical Concept of the Innate and Acquired Drives [Текст] / J. Knoll. — Berlin, Heidelberg : Springer-Verlag, 2005. — 176 p.

291. *Городецкий, В. И.* Многоагентная самоорганизация в В2В сетях [Текст] / В. И. Городецкий // XII Всероссийское совещание по проблемам управления ВСПУ-214. Москва, 16–19 июня 2014 г.: Труды. — Москва : Институт проблем управления им. В.А. Трапезникова РАН, 2014. — С. 8954–8966.
292. *Редько, В.* Многоагентная модель прозрачной рыночной экономической системы [Текст] / В. Редько, З. Сохова // Труды НИИСИ РАН. — 2013. — Т. 3, № 2. — С. 61–65.
293. *Афанасьев, М.* Разработка и исследование многоагентной системы для решения задач технологической подготовки производства. Специальность 05.11.14 – Технология приборостроения. АВТОРЕФЕРАТ диссертации на соискание учёной степени кандидата технических наук. На правах р [Текст] / М. Афанасьев. — Санкт-Петербург, 2012. — 22 с.
294. *Аронов, И. З.* Исследование времени достижения консенсуса в работе технических комитетов по стандартизации на основе регулярных марковских цепей [Текст] / И. З. Аронов, О. В. Максимова, А. В. Зажигалкин // Компьютерные исследования и моделирование. — 2015. — Т. 7, № 4. — 941950.
295. *Thomas, R. W.* Cognitive Networks. Dissertation in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Computer Engineering [Текст] : PhD thesis / Thomas R. W. — Virginia Polytechnic Institute, State University. Blacksburg, Virginia. — P. 184.
296. Программа фундаментальных научных исследований государственных академий наук на 2013-2020 годы [Текст]. — 2012. — URL: <http://www.ras.ru/scientificactivity/2013-2020plan.aspx> (дата обр. 28.07.2020).
297. *Макаров, В. Л.* Искусственные общества и будущее общественных наук [Текст] / В. Л. Макаров // Лекции и доклады членов Российской Академии наук в СПбГУП (1993-2013): в 3т. — Санкт-Петербург : СПбГУП, 2013. — С. 536–551.
298. *Новиков, Д.* Рефлексия и управление (математические модели [Текст] / Д. Новиков, А. Г. Чхартишвили. — Москва : Физматлит, 2013. — 412 с.
299. *Niazi, M. A.* Cognitive Agent-based Computing-I. A Unified Framework for Modeling Complex Adaptive Systems Using Agent-based & Complex Network-based Methods [Текст] / M. A. Niazi, A. Hussain. — New York ; London : Springer, 2013. — 55 p.

300. *Маркс, К.* Капитал, Т. 1 [Текст] / К. Маркс // Маркс К., Энгельс Ф. Сочинения. 2-е изд. Т. 23. — Москва : Госполитиздат, 1960. — С. 907.
301. *Kim, P. S.* Increased longevity evolves from grandmothing [Текст] / P. S. Kim, J. E. Coxworth, K. Hawkes // Proceedings of the Royal Society B: Biological Sciences. — 2012. — Dec. — Vol. 279, no. 1749. — P. 4880—4884. — URL: <https://doi.org/10.1098/rspb.2012.1751>.
302. *Eiben, A. E.* On the Dynamics of Communication and Cooperation in Artificial Societies [Текст] / A. E. Eiben, M. C. Schut, N. Vink // Complexus. — 2004. — Vol. 2, no. 3/4. — P. 152—162. — URL: <https://www.karger.com/DOI/10.1159/000093687>.
303. *Wooldridge, M.* An Introduction to MultiAgent Systems [Текст] / M. Wooldridge. — 2nd. — Glasgow : Wiley Publishing, 2009. — 484 p.
304. *Vlassis, N.* A Concise Introduction to Multiagent Systems and Distributed Artificial Intelligence [Текст] / N. Vlassis // Synthesis Lectures on Artificial Intelligence and Machine Learning. — 2007. — Jan. — Vol. 1, no. 1. — P. 1—71. — URL: <https://doi.org/10.2200/S00091ED1V01Y200705AIM002>.
305. Reasoning About Knowledge [Текст] / R. Fagin [et al.]. — 01/2003. — 544 p.
306. *Austin, J. L.* How to do things with words [Текст] / J. L. Austin. — Cambridge : Harvard University Press, 1962. — 192 p.
307. *Searle, J. R.* Speech acts : an essay in the philosophy of language [Текст] / J. R. Searle. — London : Cambridge University Press, 1969. — 203 p.
308. *Grice, H. P.* Studies in the Way of Words [Текст] / H. P. Grice. — Cambridge, Mass. : Harvard University Press, 1989. — 402 p.
309. *O'Regan, J. K.* A sensorimotor account of vision and visual consciousness [Текст] / J. K. O'Regan, A. Noë // Behavioral and Brain Sciences. — 2001. — Vol. 24, no. 5. — P. 883—917.
310. *Felin, T.* Rationality, perception, and the all-seeing eye [Текст] / T. Felin, J. Koenderink, J. I. Krueger // Psychonomic Bulletin & Review. — 2017. — Aug. — Vol. 24, no. 4. — P. 1040—1059. — URL: <http://link.springer.com/10.3758/s13423-016-1198-z>.



311. *Simons, D. J.* Gorillas in our midst: Sustained inattention blindness for dynamic events [Текст] / D. J. Simons, C. Chabris // Perception. — 1999. — Vol. 28, no. 9. — P. 1059–1074.
312. *Shirayev, E. B.* Cross-Cultural Psychology: Critical Thinking and Contemporary Applications [Текст] / E. B. Shirayev, D. Levy. — New York : Routledge, 2016. — 704 p.
313. *Schooler, L.* Schooler, L. Rational Theory of Cognition in Psychology [Текст] / L. Schooler // International Encyclopedia of Social & Behavioral Sciences. Pergamon. — Pergamon, 2001. — P. 12771–12775.
314. *Bender, A.* The cultural constitution of cognition: Taking the anthropological perspective [Текст] / A. Bender, S. Beller // Frontiers in Psychology. — 2011. — Vol. 2, no. 67. — P. 1–6. — URL: <http://journal.frontiersin.org/article/10.3389/fpsyg.2011.00067/abstract>.
315. *Драгалина-Черная, Е. Г.* Выправление имен: от грамматики к логике социального софтвера [Текст] / Е. Г. Драгалина-Черная // Философский журнал | Philosophy Journal. — 2016. — Т. 9, № 4. — С. 147–157. — URL: <https://pj.iph.ras.ru/article/view/161>.
316. *Turner, M.* Cognitive Dimensions of Social Science [Текст] / M. Turner. — Oxford New York : Oxford University Press, 2001. — 192 p. — URL: <https://books.google.co.il/books?id=8trnCwAAQBAJ>.
317. *Thagard, P.* Mapping Minds across Cultures [Текст] / P. Thagard // Grounding social sciences in cognitive sciences / ed. by R. Sun. — Cambridge, MA, London : The MIT Press, 2012. — P. 35–62.
318. On the Contributions of Cognitive Sociology to the Sociological Study of Race [Текст] / W. H. Brekhus [et al.] // Sociology Compass. — 2010. — Jan. — Vol. 4, no. 1. — P. 61–76. — URL: <http://doi.wiley.com/10.1111/j.1751-9020.2009.00259.x>.
319. Joint Action: Mental Representations, Shared Information and General Mechanisms for Coordinating with Others [Текст] / C. Vesper [et al.] // Frontiers in Psychology. — 2017. — Vol. 7. — P. 2039. — URL: <https://www.frontiersin.org/article/10.3389/fpsyg.2016.02039>.

320. *Emery, N. J.* The evolution of social cognition [Текст] / N. J. Emery // The Cognitive Neuroscience of Social Behaviour / ed. by A. Easton, N. J. Emery. — New York : Psychology Press, 2005. — P. 115—156.
321. *Heberlein, A. S.* Functional anatomy of human social cognition [Текст] / A. S. Heberlein, R. Adolphs // The Cognitive Neuroscience of Social Behaviour / ed. by A. Easton, N. J. Emery. — New York : Psychology Press, 2005. — P. 157—194.
322. *Lieberman, M. D.* The self and social perception: Three kinds of questions in social cognitive neuroscience [Текст] / M. D. Lieberman, J. H. Pfeifer // The Cognitive Neuroscience of Social Behaviour / ed. by A. Easton, N. J. Emery. — New York, 2005. — P. 195—235.
323. *Huebner, B.* Macrocognition: A Theory of Distributed Minds and Collective Intentionality [Текст] / B. Huebner. — OUP USA, 2014. — 278 p. — URL: <https://books.google.co.il/books?id=ObGcAQAAQBAJ>.
324. *Bogdan, R. J.* Predicative Minds: The Social Ontogeny of Propositional Thinking [Текст] / R. J. Bogdan. — MIT Press, 2009. — 259 p. — (A Bradford Book). — URL: <https://books.google.ru/books?id=Ki8BGogK2YUC>.

**Список рисунков**

3.1	Рациональное, когнитивное и интеллектуальное . . . . .	256
3.2	Иллюзия Понцо . . . . .	272
3.3	Иллюзия Эббингауза . . . . .	272
3.4	Куб Неккера . . . . .	273

## Список таблиц

1	Виды вычислений . . . . .	161
---	---------------------------	-----