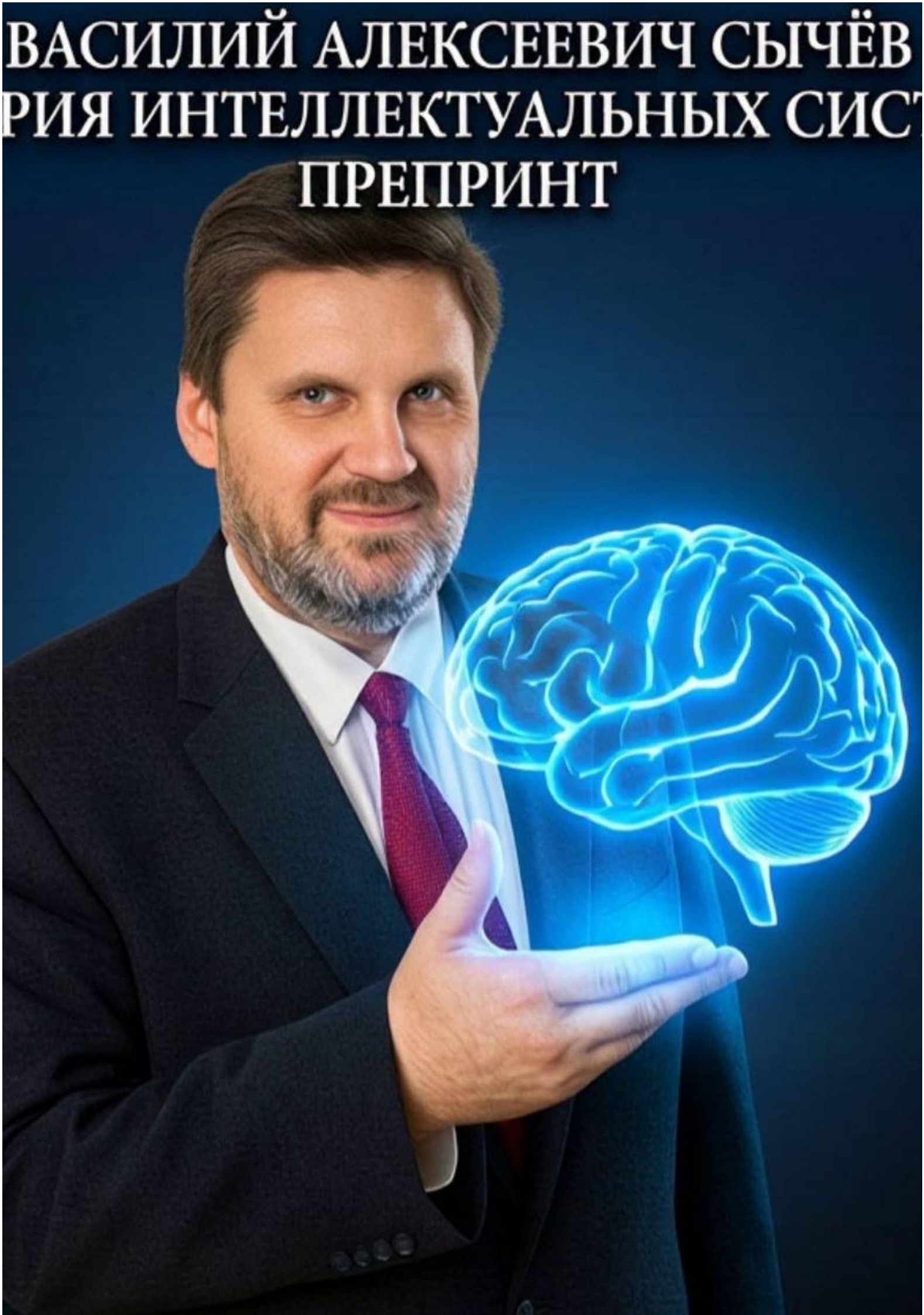
ВАСИЛИЙ АЛЕКСЕЕВИЧ СЫЧЁВ

РИЯ ИНТЕЛЛЕКТУАЛЬНЫХ СИСТ

ПРЕПРИНТ



Василий Сычев

Теория интеллектуальных систем препринт

«Автор»
2026

Сычев В.

Теория интеллектуальных систем препринт / В. Сычев — «Автор», 2026

Автор последовательно прослеживает, как из первичных целей репликаторов возникли все остальные цели, как в ходе эволюции появились мемстинкты — волнообразно возбуждающиеся популяции нейронов, способные записывать последовательности сигналов, совпадающие по времени с важными для выживания событиями, — и как сформировались инстинкты, мышление, культура и сознание.В книге предложена оригинальная модель работы гиппокампа, сна и сновидений, описана архитектура зооморфной искусственной интеллектуальной системы (ЗИИС). Кроме того, в монографии предложен новый математический аппарат — лингвоматика. Отдельные главы посвящены языку, культуре, юмору, любви, этике и проблеме цифрового бессмертия. В монографии представлен научный план, который в перспективе способен привести к полному преодолению ограничений естественной продолжительности жизни.

© Сычев В., 2026
© Автор, 2026

Содержание

АННОТАЦИЯ 7
Введение 13
Теологические и философские истоки понятий «сознание» и «субъект» 13
Кантовский переворот: трансцендентальный субъект 14
От феноменологии к аналитической традиции 15
Современные междисциплинарные исследования 16
Технологический поворот: субъект в кибернетике и науках управления 17
Проникновение в физику: квантовая механика и проблема измерения 18
Постановка задачи 19
Глава 1. Появление целей благодаря естественным процессам обучения 21
1.1. Определение интеллектуальной системы 21
1.2. Обучение первого типа: мир РНК и библиотеки генов 23
1.3. Обучение второго типа А: привыкание и рефлексы 26
1.4. Обучение второго типа Б: от условных рефлексов к первому мемстинкту 28
1.5. Инстинкты как врождённые детекторы последовательностей 33
1.6. Культура — третий тип интеллектуальных систем 38
1.7. Интеллектуальные системы, обучающиеся по четвёртому типу 42
1.8. Нетипичные интеллектуальные системы 43
Выводы первой главы 45
Глава 2. Общие принципы работы мозга в рамках Общей когнитивной теории как части Теории интеллектуальных систем 50
Введение 50
2.1. Базовые элементы психики: мемстинкты, инстинкты и воля 53
2.2. Инстинкты внимания и формирование мемстинктов первых уровней в онтогенезе 56
2.3. Мотивационные инстинкты: врождённые детекторы биологической значимости 58
2.4. Реципрокное торможение и волевой акт: почему мы делаем то, что приносит удовольствие 61
2.5. Инстинкты агрессии: выход из двигательного паралича 64
2.6. Планирование как внутреннее проигрывание: отбрасывание опасных вариантов 65
2.7. Нервная система как часть целостного организма: аллостаз, эндокринная и иммунная регуляция 67
Конец ознакомительного фрагмента. 68

Теория интеллектуальных систем препринт

Предисловие

Президент Межнациональной ассоциации психофизиологов, профессор О. С. Булгакова

«Теория интеллектуальных систем» — это большой фундаментальный труд. Я рада, что на мою долю выпала возможность внести свой посильный вклад в поддержку традиций российской психофизиологии — в том числе через участие в создании ассоциации психофизиологов. Вопреки распространённому мнению, наша психофизиология не сказала своего последнего слова в начале XX века — мы можем и сегодня вносить значимый вклад в мировую науку.

В последние десятилетия ведущие учёные, изучающие нервную систему, всё чаще называют себя нейрофизиологами, подчёркивая, что для понимания принципов её работы необходимо глубже заглянуть в отдельные нейроны мозга и изучить протекающие в них процессы. В этом направлении проделана огромная работа, и проделана она не зря: понимание тонких процессов, происходящих внутри нейронов и микроглии, кардинально обогатило мировую науку. Однако, как для врача важно понимать процессы, протекающие в отдельных клетках пациента, но ещё важнее — видеть общие процессы, происходящие в организме в целом, так и для учёных, изучающих нервную систему человека, конечно, очень важно знать принципы работы отдельных клеток мозга, но ещё важнее — видеть всю нервную систему целиком. Именно так работали великие психофизиологи российской школы начала прошлого века. Именно такой общий взгляд на психику человека позволил им построить теории, предвосхитившие своё время.

Василий Алексеевич Сычев в своей работе постарался продолжить традиции старой школы. Опираясь на труды прошлого, он также внимательно изучил достижения современной науки, что и позволило ему объединить огромный объём накопленных в настоящее время знаний в целостную теорию. Поражает объём проделанной работы: теория, претендующая на объяснение подавляющего большинства процессов в нашей психике, создана — как пишет сам автор — всего лишь за десять лет.

Я не берусь судить, все ли выводы автора являются верными. Чтобы досконально изучить все выводы теории, потребуются десятилетия работы, и только наука будущего сможет дать точную оценку Теории интеллектуальных систем. Но в целом данная теория представляется мне вполне проработанной и убедительной.

В Теории интеллектуальных систем предпринята попытка описать процессы возникновения первых интеллектуальных систем — то есть того, как в принципе возникла субъектность. Во второй главе автор описывает, как благодаря взаимодействию отдельных частей мозга возникают процессы мотивации, мышления, сохранения памятных следов и так далее. Как учёному, в круг интересов которого входит в том числе изучение процессов психосоматики, мне особенно интересно, как теория объясняет процессы аллостаза нашего организма.

Но теория не ограничивается только попыткой объяснить принципы работы нашего мозга. Василий Алексеевич предлагает также подходы к созданию искусственного интеллекта, построенного на глубоком понимании процессов, протекающих в нашем мозге. При этом очень интересно, что, по его утверждению, такие интеллектуальные системы можно создать полностью на отечественной микроэлементной базе. Теория также рассматривает философские вопросы и вопросы этики будущего.

В целом Теория интеллектуальных систем кажется мне очень интересной, и я однозначно рекомендую её прочитать.

АННОТАЦИЯ

Монография представляет собой систематическое изложение Теории интеллектуальных систем — междисциплинарной теории, объединяющей данные нейрофизиологии, психофизиологии, когнитивной науки, эволюционной биологии и философии сознания. В основе работы лежит идея о том, что все известные интеллектуальные системы — от первых репликаторов до человеческой культуры и искусственного интеллекта — возникли в результате процессов обучения, порождающих цели. Автор последовательно прослеживает, как из первичных целей репликаторов возникли все остальные цели, как в ходе эволюции появились мемстинкты — волнообразно возбуждающиеся популяции нейронов, способные записывать последовательности сигналов, совпадающие по времени с важными для выживания событиями, — и как сформировались инстинкты, мышление, культура и сознание.

В книге подробно разобраны клеточные и субклеточные механизмы сохранения воспоминаний и мышления, предложена оригинальная модель работы гиппокампа, сна и сновидений, описана архитектура зооморфной искусственной интеллектуальной системы (ЗИИС), которая, в отличие от существующих подходов, не требует сверхбольших вычислительных мощностей и может быть реализована на доступной элементной базе — что открывает путь к созданию полноценного искусственного интеллекта в рамках одной страны, от фундаментальной теории до работающего прототипа. Кроме того, в монографии предложен новый математический аппарат — лингвоматика — для описания иерархических систем мемстинктов. Отдельные главы посвящены языку, культуре, юмору, любви, этике и проблеме цифрового бессмертия. В монографии представлен научный план, который в перспективе способен привести к полному преодолению ограничений естественной продолжительности жизни.

Монография адресована специалистам в области нейрофизиологии, психофизиологии, когнитивной науки, искусственного интеллекта, культурологии и философии, а также широкому кругу читателей, интересующихся устройством человеческой психики и перспективами развития интеллектуальных систем.

Ключевые слова: Теория интеллектуальных систем; Общая когнитивная теория; мемстинкт; инстинкт; обучение; сознание; субъект; консолидация; культура; лингвоматика; зооморфная искусственная интеллектуальная система; цифровое бессмертие.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

Теологические и философские истоки понятий «сознание» и «субъект»
Кантовский переворот: трансцендентальный субъект
От феноменологии к аналитической традиции
Современные междисциплинарные исследования
Технологический поворот: субъект в кибернетике и науках управления
Проникновение в физику: квантовая механика и проблема измерения
Постановка задачи

Глава 1. Появление целей благодаря естественным процессам обучения

1.1. Определение интеллектуальной системы
1.2. Обучение первого типа: мир РНК и библиотеки генов
1.3. Обучение второго типа А: привыкание и рефлексы
1.4. Обучение второго типа Б: от условных рефлексов к первому мемстинкту
1.5. Инстинкты как врождённые детекторы последовательностей
1.6. Культура — третий тип интеллектуальных систем
1.7. Интеллектуальные системы, обучающиеся по четвёртому типу
1.8. Нетипичные интеллектуальные системы
Выводы первой главы

Глава 2. Общие принципы работы мозга в рамках Общей когнитивной теории как части Теории интеллектуальных систем

Введение

2.1. Базовые элементы психики: мемстинкты, инстинкты и воля
2.2. Инстинкты внимания и формирование мемстинктов первых уровней в онтогенезе
2.3. Мотивационные инстинкты: врождённые детекторы биологической значимости
2.4. Реципрокное торможение и волевой акт: почему мы делаем то, что приносит удовольствие
2.5. Инстинкты агрессии: выход из двигательного паралича
2.6. Планирование как внутреннее проигрывание: отбрасывание опасных вариантов
2.7. Нервная система как часть целостного организма: аллостаз, эндокринная и иммунная регуляция
2.8. Механизмы усталости и цикла «сон — бодрствование»
2.9. Основные принципы возникновения психосоматических и нейродегенеративных заболеваний
2.10. Учёт психоэмоционального состояния в медицинских протоколах
2.11. Психические заболевания и состояния в свете Общей когнитивной теории
Выводы второй главы

Глава 3. Нейробиологические основы консолидации и узнавания мемстинктов

Введение

3.1. Мотивационные инстинкты. Подкорковые источники нейромодуляции и их рецепторный аппарат: механизмы воздействия на глутаматергическую передачу
3.1.1. Дофамин: D1- и D2-подобные рецепторы
3.1.2. Норадреналин: α- и β-адренорецепторы
3.1.3. Гистамин: H1- и H2-рецепторы
3.1.4. Серотонин: 5-HT1A и 5-HT2A-рецепторы
3.1.5. Ацетилхолин: никотиновые и мускариновые рецепторы
3.2. Мотивационные инстинкты. Пути поступления нейромодуляторов в кору и регионализация инстинктов
3.3. Эволюционные предпосылки нейромедиаторной специфичности мотивационных инстинктов
3.4. Ламинарная организация AMPA/NMDA-рецепторов в префронтальной коре и волновая динамика

3.5. Мемстинкт как единица записи последовательности
3.5.1. Клеточные механизмы сохранения мемстинктов
3.5.1.1. Сохранение последовательностей в мемстинктах первых уровней: роль инстинктов внимания
3.5.1.2. Сохранение последовательностей в мемстинктах высоких уровней: роль мотивационных инстинктов
3.5.2. Регионы инстинкта и двусторонняя связь мемстинкт-инстинкт
3.5.3. Клеточные механизмы узнавания
3.5.4. Механизмы записи и узнавания у примитивных животных
3.5.4.1. Процеребро моллюсков
3.5.4.2. Грибовидные тела насекомых
3.5.4.3. Хордовые: данные по данио-рерио
3.5.4.4. Реципрокное торможение и возбуждение протоинстинктов
3.5.4.5. Эволюционный синтез
3.5.5. Время и длительность: как мозг измеряет продолжительность процессов
3.6. Роль гиппокампа в консолидации мемстинктов: реципрокное усиление и нейрогенез
3.6.1. От «временного хранилища» к реципрокному осциллятору
3.6.2. Проблема универсальности нейронов решётки: данные от других видов
3.6.3. Альтернативная интерпретация: решётка как карта обонятельных меток
3.6.4. Как мы сохраняем пространственную информацию
3.6.5. Гиппокамп как реципрокный усилитель корковых мемстинктов
3.6.6. Генерация высокочастотных осцилляций и replay
3.6.7. Оценка количества мемстинктов: арифметика нейрогенеза и архитектоники колонок
3.7. Динамика сна как циклический механизм поддержания консолидации мемстинктов
3.7.1. Медленный сон: прекращение активной консолидации
3.7.2. Быстрый сон: периодическая реактивация мотивационных инстинктов
3.7.3. Сновидения как побочный продукт реактивации
3.7.4. Защита от сохранения сновидений
3.8. Согласование с данными функциональной нейроанатомии и ритмами ЭЭГ
3.9. Консолидация двигательных мемстинктов и архитектура моторной коры
3.10. Зеркальные нейроны, мемстинкты зеркальных нейронов (МЗН) и культура
3.11. ОКТ и экспериментальные данные: проблема интерпретации
3.11.1. Почему наука пошла не по тому пути: историко-методологический экскурс
Выводы третьей главы

Глава 4. Принципы построения зооморфной интеллектуальной системы (ЗИИС)

Введение

4.1. Общая архитектура ЗИИС
4.2. Преимущества архитектуры ЗИИС по сравнению с существующими архитектурами
4.2.1. Энергоэффективность и нейроморфная реализация
4.2.2. Сигнал-аналог: отказ от кодирования
4.2.3. Многотысячезначная логика как математическая метафора
4.2.4. Информационная эффективность иерархии последовательностей
4.2.5. Частичное узнавание и обобщение без полного перебора примеров
4.2.6. Мультимодальность как врождённое свойство архитектуры
4.2.7. Отсутствие сквозной архитектуры и возможность обучения «на лету»

4.2.8. Устойчивость к катастрофическому забыванию
4.2.9. Планирование, творчество и оценка планов
4.2.10. Сравнение с традиционными архитектурами: метафора грифельной доски
4.2.11. Выводы: ЗИИС как архитектура сильного искусственного интеллекта
4.3. Аналоговая арифметика: суммация возбуждений как основа мышления и творчества
4.3.1. Инстинкты как усилители сигнала в среде с положительными обратными связями и торможением
4.3.2. Частотная маркировка как дополнительное измерение «арифметики»
4.4. Искусственные МЗН и обучение через культуру роботов
4.4.1. Архитектурные требования к искусственному МЗН
4.4.2. Обучение наблюдением и культурная синхронизация
4.4.3. Взаимодействие с человеческой культурой
4.4.4. Автономная культура неантропоморфных интеллектуальных систем
4.5. Другие архитектуры, использующие зооморфные принципы
4.6. Создание прототипа зооморфной интеллектуальной системы (ЗИИС)
4.6.1. Общие вопросы
4.6.2. Описание прототипа ЗИИС-1
4.6.2.1. Основные компоненты системы
4.6.2.2. Правила работы нейронов
4.6.3. Второе поколение: ЗИИС-2
4.6.4. Правила работы системы: три типа
Выводы четвёртой главы

Глава 5. Язык и культура

Введение

5.1. Что современная наука знает о времени возникновения языка
5.2. Что ОКТ добавляет в копилку знаний о языке: эффективность двусоставных и трёхсоставных групп
5.2.1. Проблема слога: не ошибка, а решение для экономии памяти
5.2.2. Разрешение спора Хомского с бихевиористами: врождённое и приобретённое в языке
5.3. Как возникал язык: от рефлекторного крика к иерархии смыслов
5.4. Иерархия языка как отражение иерархии мемстинктов
5.5. Почему мы любим стихи, песни и музыку: власть повтора и её обратная сторона
5.6. Юмор и смех как неожиданная активация целевого инстинкта
5.7. Добро и зло: физиологическая основа этических категорий
5.8. Любовь, дружба и социальные инстинкты
5.9. Культура как распределённая система хранения мемстинктов
5.10. Факторы ускорения культурной эволюции
5.11. Пределы скорости изменения культуры и количество каналов коммуникации
5.12. Дополненное мышление: культура как расширение разума
5.13. Условия успешной консолидации культурных мемстинктов
5.14. Искусство как инструмент изменения культуры
5.15. Почему мы так ценим произведения искусства: механизм переоценки значимости
5.16. Искусствоведение и культурология как точные науки
5.17. Культура как инструмент управления поведением
5.18. Культура как инструмент управления: политический аспект

5.19. Международные средства коммуникации и рост агрессивности культуры
5.20. Ответственность за изменение культуры и роль науки
Выводы пятой главы

Глава 6. Лингвоматика

Введение

6.1. Система мемстинктов как многотысячезначное число
6.2. Комбинаторный анализ: сколько вариантов может сохранить система?
6.2.1. Потенциальная ёмкость иерархической системы
6.2.2. Реальный мозг человека: языковая аналогия
6.2.3. Ограничения упрощённой модели
6.2.4. Почему расчёт в битах некорректен
6.3. Оптимальное количество сигналов в последовательности: математическое доказательство
6.3.1. Постановка задачи и переменные
6.3.2. Определение функции стоимости
6.3.3. Вывод функции и поиск минимума
6.3.4. Учёт целочисленного ограничения
6.3.5. Итоговый результат
6.4. Подходы к динамической лингвоматике
6.4.1. Ключевые процессы: консолидация и узнавание
6.4.2. Подходы к упрощению системы для моделирования
6.4.3. Стандартизация искусственных мемстинктов и многофакторное моделирование
6.5. Статическая лингвоматика: форма записи
6.5.1. Лингвоматическая запись: система как ключ к самой себе
6.5.2. Почему теория графов не подходит для описания системы
6.5.3. Применение статической лингвоматики
Выводы шестой главы

Глава 7. Сознание, философия и общие вопросы

Введение

7.1. Эпизод сознания как гиперактивация сенсорного элемента
7.1.1. Эмоции как интероцептивные ощущения
7.2. Свобода воли и привычка: от философского вопроса к нейрофизиологическому механизму
7.3. Самосознание: почему мы узнаём себя в зеркале?
7.4. Сложность сознания как проекция сложности интеллектуальной системы
7.5. Логика и философия как производные от архитектуры интеллекта
7.5.1. Время и последовательность как условие логического мышления
7.5.2. Точечный сбор информации и логика перемещения
7.5.3. Иерархическая организация как основа дедукции, индукции и системного подхода
7.5.4. Логика как продукт архитектуры интеллекта
7.5.5. Сравнение как фундаментальный процесс узнавания и его роль в психике
7.5.6. Может ли появиться сознание у искусственного интеллекта?

7.5.7. Проблема определения живого: как инстинкты внимания сформировали интуитивную категорию
7.6. Самоформирование: сознательное управление системой мемстинктов
7.6.1. Общие принципы самоформирования
7.6.2. Практический подход к ослаблению стрессовых мемстинктов
7.6.3. Пример: работа с травмирующим воспоминанием
7.6.4. Формирование полезных привычек
Выводы седьмой главы

Глава 8. Будущее: реальный научный план перехода к цифровому бессмертию

Введение

8.1. Первые шаги: протезирование нервной системы
8.2. Протезирование инстинктов и мемстинктов
8.3. Полная замена и цифровое бессмертие
8.3.1. Киберсимбиот
8.3.2. Постепенная замена
8.4. Новая этика: субъектность в эпоху сильного интеллекта и протезирования мозга
Заключение

Список литературы

Глоссарий

Введение

Теологические и философские истоки понятий «сознание» и «субъект»

Понятия «сознание» и «субъект» представляют собой одни из наиболее фундаментальных и одновременно проблематичных концептов в истории человеческой мысли. В отличие от происхождения жизни, где современная наука располагает наиболее разработанной на сегодняшний день гипотезой «мира РНК» (RNA World hypothesis), объясняющей переход от неживой материи к самовоспроизводящимся системам, возникновение сознания и субъективности остается глубочайшей загадкой. Эти понятия появились в человеческом дискурсе как бы «из воздуха» — из теологических размышлений о духах, природе души, божественного познания и человеческого места в творении.

Парадоксальность ситуации заключается в том, что, несмотря на отсутствие научного консенсуса относительно самих определений сознания и субъекта, эти концепты оказываются принципиально важными для развития целых отраслей современной науки. Без них кибернетика, науки управления, разработка искусственного интеллекта и когнитивные исследования не могут адекватно концептуализировать свои объекты исследования. Современные эмпирические исследования показывают, что ученые по-разному интерпретируют эти термины, что создает методологические трудности для междисциплинарного диалога.

Разумеется, сейчас мы не сможем точно определить, когда в человеческой культуре появились размышления о субъекте и сознании, но мы можем увидеть следы подобных размышлений в шаманских верованиях. Как показывают исследования, шаманизм представляет собой одну из древнейших форм культурной и когнитивной практики, связанной с измененными состояниями сознания [1].

Зарождение научных идей можно наблюдать и в трудах древнегреческих философов, в частности, в работах Платона и в психологии Аристотеля, роль деятельного ума в которой была глубоко проанализирована Францем Брентано [2].

В средние века генезис понятия субъекта неразрывно связан с теологическими размышлениями о природе божественного и человеческого познания. Фома Аквинский (1225–1274) в своей «Сумме теологии» (Summa Theologica) заложил фундаментальные основы для понимания субъекта через призму отношения между интеллектом и волей. В его концепции intellectus и voluntas образуют единство познающего субъекта, способного к рефлексии не только о внешнем мире, но и о собственных познавательных актах [3].

Аквинский вводит критически важное различение между esse naturale (естественным бытием) и esse intentionale (интенциональным бытием), которое предвосхищает современные дискуссии об интенциональности сознания. Его анализ cogitativa (оценивающей способности) как связующего звена между чувственным познанием и интеллектуальным пониманием содержит в зародыше идею сознания как специфически человеческой способности к самопознанию и саморефлексии.

Дальнейшее развитие этих идей происходило в рамках поздней схоластики, где Дунс Скот и его последователи разрабатывали концепцию haecceitas (этости), а номиналисты подготавливали почву для радикального пересмотра отношений между универсальным и индивидуальным в познании.

Кантовский переворот: трансцендентальный субъект

Иммануил Кант (1724–1804) осуществил подлинную революцию в понимании субъекта, переведя проблематику из метафизической плоскости в эпистемологическую. В «Критике чистого разума» (1781/1787) Кант формулирует концепцию трансцендентального субъекта как условия возможности всякого опыта и познания. Его знаменитое утверждение о том, что «я мыслю должно сопровождать все мои представления» (Das: Ich denke, muß alle meine Vorstellungen begleiten können), устанавливает единство апперцепции как фундаментальную структуру сознания [4].

Кантовское различение между эмпирическим и трансцендентальным субъектом открыло новые горизонты для понимания сознания как активной синтетической деятельности, а не пассивного восприятия данных. Концепция спонтанности рассудка и его категориальной структуры предвосхитила многие идеи современной когнитивной науки, хотя и в философско-трансцендентальной оболочке.

От феноменологии к аналитической традиции

XIX и XX века ознаменовались драматическим расширением проблемного поля исследований сознания и субъективности. Эдмунд Гуссерль (1859–1938) в «Логических исследованиях» (1900–1901) и «Идеях к чистой феноменологии» (1913) развил систематическую науку о сознании, введя понятие интенциональности как основной структуры сознательных актов [5]. Его концепция трансцендентального эго и методы феноменологической редукции оказали влияние на всю последующую философию сознания.

Параллельно в аналитической традиции Готлоб Фреге, Бертран Рассел и ранний Людвиг Витгенштейн разрабатывали логико-лингвистические подходы к проблеме субъекта и сознания. Витгенштейновский анализ «приватного языка» в «Философских исследованиях» (1953) поставил под сомнение возможность прямого доступа к содержаниям сознания и заложил основы для социально-лингвистических концепций субъективности [6].

Современные междисциплинарные исследования

Во второй половине XX века исследования сознания становятся интенсивно междисциплинарными. Дэвид Чалмерс в работе «Сознающий ум» (1996) сформулировал различение между «легкими» и «трудной» проблемами сознания, где последняя касается объяснения феноменального опыта [7]. Томас Нагель в классической статье «Каково быть летучей мышью?» (1974) поставил вопрос о принципиальной субъективности сознательного опыта [8].

Дэниел Деннет в «Объясненном сознании» (1991) предложил элиминативистский подход, рассматривая сознание как иллюзию, создаваемую сложными информационными процессами [9]. Джон Серл в работах по философии сознания развил концепцию биологического натурализма, утверждая, что сознание является естественным биологическим феноменом [10].

Технологический поворот: субъект в кибернетике и науках управления

Особого внимания заслуживает роль понятий субъекта и сознания в современных технологических науках. В кибернетике, основанной Норбертом Винером [11], концепция субъекта приобретает операциональное значение через понятия обратной связи, саморегуляции и целенаправленного поведения. Современные исследования в области искусственного интеллекта, особенно в рамках архитектур трансформеров — особого типа архитектуры искусственных нейронных сетей, на основе которого построены современные большие языковые модели (GPT, Claude и другие), — и больших языковых моделей, вынуждены обращаться к проблематике сознания и субъективности для объяснения эмерджентных свойств сложных систем.

В науках управления понятие субъекта становится центральным при анализе организационного поведения, принятия решений и стратегического планирования. Работы Питера Сенге по обучающимся организациям [12] и исследования в области organizational cognition демонстрируют необходимость концептуализации коллективного субъекта и распределенного сознания.

В области человеко-машинного взаимодействия и когнитивной эргономики понятие субъекта приобретает практическое значение при проектировании интерфейсов и систем поддержки принятия решений. Дональд Норман в работе «Дизайн привычных вещей» [13] показал, как понимание когнитивных процессов субъекта влияет на эффективность технологических систем.

Проникновение в физику: квантовая механика и проблема измерения

Парадоксальным образом понятие субъекта пытается найти место даже в физике, традиционно стремящейся к объективному описанию природы. Проблема измерения в квантовой механике, начиная с копенгагенской интерпретации Нильса Бора и Вернера Гейзенберга, вводит наблюдателя как конституирующий элемент физической реальности. Юджин Вигнер в работе «Замечания о проблеме тела-разума» (1961) предложил рассматривать сознание как фактор, вызывающий коллапс волновой функции [14].

Современные интерпретации квантовой механики, такие как многомировая интерпретация Хью Эверетта III или объективные теории коллапса Джанкарло Гирарди, Альберто Римини и Туллио Вебера, пытаются элиминировать субъективный элемент, но дискуссии о роли сознания в квантовых процессах продолжаются. Роджер Пенроуз и Стюарт Хамерофф в теории «orchestrated objective reduction» предполагают, что сознание возникает из квантовых процессов в микротрубочках нейронов [15].

Постановка задачи

Представленный исторический обзор выявляет три фундаментальные методологические проблемы. Во-первых, отсутствие научного консенсуса относительно базовых определений создаёт терминологическую неопределённость. Современные исследования в области consciousness studies демонстрируют поразительное разнообразие в понимании самого предмета исследования среди учёных [16]. Во-вторых, в современной науке полностью отсутствует представление о том, как возникли сознание и субъект. Наконец, в-третьих, понятие субъекта является основополагающим для целого ряда наук, и именно эта проблема — центральная для данной работы. В результате неоднократно предпринимались попытки заменить понятие субъекта другим понятием, к которому проще дать определение. Например, в кибернетике часто используется понятие управления как перевода системы из одного состояния в другое. Согласно классическому определению, управление — это целенаправленное воздействие на объект (систему) с целью перевода его из исходного состояния в требуемое [17]. Если теперь заменить управление понятием алгоритма (совокупности точно заданных правил) для перевода из одного состояния в другое, казалось бы, вообще можно обойтись без понятия субъекта: необходим просто определённый алгоритм перехода из одного состояния в другое, и мы получаем управление объектом. Однако, поскольку у системы может быть множество различных состояний, обязательно необходимо задать состояние, в которое нужно перевести систему, то есть должна быть определена цель, а это опять невозможно без субъекта. В своей работе «Введение в кибернетику» Уильям Росс Эшби сформулировал и обосновал закон необходимого разнообразия, из определения которого вытекает, что целью управления является снижение энтропии управляемого объекта или процесса [18]. И кажется, вот оно, решение: цель любого управления — снижение энтропии. Однако мы видим, что классическое определение энтропии в данном случае не подходит. Живые системы могут как увеличивать энтропию в результате своей деятельности, так и уменьшать её. Например, фикусы-душители в тропических лесах стремятся быстро достичь максимальной высоты, увеличивая при этом свою кинетическую энергию и метаболизм, а некоторые виды животных при неприятных внешних условиях, например засухе, могут впадать в состояние, подобное спячке, при этом они зарываются как можно глубже в землю, то есть минимизируют свою кинетическую энергию и снижают метаболизм. Понимая это, в своём определении Эшби использовал статистическую или информационную энтропию, однако, даже если использовать это дискуссионное понятие, мы можем говорить только о вероятности перевода системы из одного состояния x в другое состояние y, и в этом случае нам опять не обойтись без определения состояния y, то есть нам нужно знать цель системы.

В данной работе мы попробуем разобраться, как появились первые интеллектуальные системы и как в результате обучения появилась первая цель, из которой в последующем возникли все остальные цели всех интеллектуальных систем, которые мы знаем. Отделить интеллектуальные системы от других и провести их классификацию нам поможет процесс обучения. Как оказалось, именно этот процесс, в результате которого и возникают цели, позволяет навести порядок в данном вопросе.

Теория интеллектуальных систем появилась как естественное развитие Общей когнитивной теории [19], которая входит в Теорию интеллектуальных систем как её составная часть. Общая когнитивная теория описывает, как устроена психика животных и, прежде всего, человека, а Теория интеллектуальных систем, кроме того, что ещё полнее и подробнее описывает устройство психики животных, объясняет, как в целом возникли интеллектуальные системы, как они развивались, а также затрагивает ряд других общих вопросов, в том числе рассматривает необходимость введения нового математического аппарата для описания особенно слож-

ных интеллектуальных систем, таких как человеческий мозг или наша культура. Разработка подобного математического аппарата в будущем позволит практически стереть грань между гуманитарными и точными науками. На всех этапах разработки Теории интеллектуальных систем автор строго придерживался системно-процессного подхода, всегда рассматривались и описывались системы и процессы, которые можно наблюдать благодаря научным экспериментам, при этом указываются также подсистемы, которые участвуют в данных процессах или из которых состоят описываемые системы и которые, в свою очередь, также можно наблюдать в ходе научных экспериментов.

В результате наших размышлений мы также сможем дать определение понятий «сознание» и «субъект», но мы увидим, что данные понятия не являются ключевыми для понимания того, как устроен доступный для изучения мир.

Глава 1. Появление целей благодаря естественным процессам обучения

1.1. Определение интеллектуальной системы

Как мы уже упоминали, основным процессом, который позволит нам отделить интеллектуальные системы от всех прочих и также позволит их классифицировать, будет процесс обучения. Для начала дадим определение интеллектуальной системы — это система, которая после обучения (обязательно включающего внешнее воздействие) так изменяет свои подсистемы и процессы, что статистически значимо — то есть не случайно, а закономерно — увеличивается вероятность достижения определённого состояния (цели обучения) при разнообразных, даже диаметрально противоположных и противоречащих достижению цели состояниях надсистемы. В результате обучения создаётся модель надсистемы, которая позволяет отслеживать её изменения и реагировать в случае необходимости. Ключевое в этом определении слово — «разнообразных». Если в результате внешнего воздействия какая-либо система, например камень треснет, то есть изменится его внутренняя среда (подсистемы), и от него отколется часть, которая покатится с горы, в соответствии с нашим определением мы не сможем сказать, что наш камень прошёл обучение и приобрёл цель — двигаться вниз. Вот если какая-либо система научится не только скатываться с горы, но и забираться на неё, то есть достигать цели при разных, диаметрально противоположных углах наклона, нам придётся признать, что данная система стала интеллектуальной.

Кроме того, поскольку интеллектуальная система создаёт модель надсистемы, у неё должна быть подсистема моделирования, которая и хранит модель надсистемы (сама является моделью надсистемы). Её можно было бы назвать подсистемой памяти, но в этом случае мы начали бы её путать с памятью компьютерной, в отличие от которой подсистема моделирования способна изменяться под внешним воздействием (обучением), соответствующим определённому состоянию надсистемы, и в дальнейшем похожим образом реагировать на похожие воздействия внешней среды, в свою очередь воздействуя на подсистемы интеллектуальной системы таким образом, чтобы увеличивать вероятность достижения целей обучения. Другими словами, подсистема моделирования должна напрямую или через другие подсистемы интеллектуальной системы воздействовать на внешнюю среду, чтобы достичь цели. При разных типах обучения подсистемой моделирования могут быть, например, несколько поколений ДНК или РНК (библиотеки генов), популяции нейронов, сохраняющие последовательность активации рецепторов, искусственная нейронная сеть, отдельные химические каскады, органоиды или процессы внутри клетки и т. д. Возможно ли обучение системы, не имеющей подсистемы моделирования? Мы можем представить себе систему, например, ёмкость с химически чистой водой, которая при достижении нуля градусов Цельсия не будет замерзать. Если мы добавим в эту ёмкость центры образования кристаллов, то при достижении нуля градусов Цельсия она начнёт кристаллизоваться и превращаться в лёд. Причём если после этого воду растопить, она опять сможет замёрзнуть при нуле градусов. Наверное, добавив большую долю воображения, мы могли бы сказать, что обучили данную систему кристаллизоваться при нуле градусов, однако в данном конкретном случае процесс «обучения» (добавление центров кристаллизации) никак не связан с охлаждением воды, повторяющимся внешним воздействием (которые должна «запомнить» подсистема моделирования), кроме того, отсутствует сама подсистема моделирования, поэтому назвать такой процесс обучением, а данную систему интеллектуальной мы не можем. Отдельно следует заметить, что просто наличие в какой-то системе

подсистемы, способной сохранять внешнее воздействие, не делает такую систему интеллектуальной. Например, флеш-память не является интеллектуальной системой, так же как и наш спутник Луна, хранящая информацию о поздней тяжёлой бомбардировке, также не является интеллектуальной системой.

Итак, интеллектуальная система обязательно должна иметь подсистему моделирования, которую также можно назвать подсистемой прогнозирования (ПП). Такая подсистема может принимать сигналы из внешней среды (изменяться под воздействием внешних процессов) и при повторении подобного воздействия, в свою очередь, воздействовать на свои эффекторные системы/манипуляторы таким образом, чтобы увеличить вероятность достижения цели обучения. При поступлении похожих сигналов из внешней среды ПП всегда реагирует похожим образом, то есть по определённому алгоритму. Иначе говоря, ПП — это такая система, которая как бы говорит: «Такое уже было, я помню, надо действовать вот так». Благодаря описанным процессам, интеллектуальная система может узнавать различные состояния надсистемы и реагировать на них определённым образом. Например, иммунные клетки могут запомнить молекулярную метку «агрессора» и в дальнейшем атаковать её; рыбак, выйдя утром на берег и увидев определённую конфигурацию волн, предвещающую шторм, может отказаться от рыбалки и т. д. Прогнозирование в самой простой форме, в принципе, сводится к узнаванию. Процессы внешнего мира, благодаря законам природы, могут повторяться, при этом мы можем понимать процесс как последовательность определённых состояний системы, и если эти состояния повторяются, то, узнав одно из них, мы автоматически узнаём о вероятном наступлении следующего. Подробнее о прогнозировании последовательностей событий мы будем говорить ниже, а сейчас рассмотрим, каким образом мы можем классифицировать интеллектуальные системы.

1.2. Обучение первого типа: мир РНК и библиотеки генов

Теоретически возможны разнообразные интеллектуальные системы, которые могли возникать в разных частях Вселенной независимо друг от друга, но все известные нам интеллектуальные системы возникли один раз, по всей вероятности на нашей планете, в результате процесса, описываемого гипотезой мира РНК. В рамках этой гипотезы молекула РНК смогла стать первой ПП [20; 21], которая обладает следующими характеристиками. Во-первых, она содержит геном, а значит, может манипулировать внешней средой, создавая молекулы белка и другие молекулы, наращивая себя из частей внешней среды. Во-вторых, она может копировать себя, создавая библиотеки копий. В-третьих, она может незначительно изменяться, мутировать, а значит, копии в таких библиотеках будут не совсем идентичные.

Процесс обучения в данном случае, назовём его обучением по первому типу, состоит из трёх этапов. Первый репликатор — молекула или группа молекул РНК, способных создавать свою копию и воздействовать на окружающую среду, изменяя её молекулы по определённому алгоритму, — создал несколько своих копий. На первом этапе обучения под воздействием мутагенных факторов в большое количество этих копий вносятся изменения, мутации, которые, в том числе, меняют алгоритм воздействия на внешнюю среду (молекула начинает делать другие белки и т. д.). Подавляющее большинство таких изменений вредны для процесса передачи генетической информации следующему поколению, поэтому обязательно требуется второй этап — отбор репликаторов, способных передавать генетическую информацию. На данном этапе самые вредные мутации отбраковываются вместе с несущими их репликаторами, и остаются только те немногие репликаторы, которые смогли достичь определённого состояния — передать генетическую информацию следующему поколению. В результате повторения данных процессов создаётся большая библиотека, содержащая наборы генов — разные РНК, способные по-разному воздействовать на внешнюю среду, создавая разные белки и другие молекулы. Только такая большая система, содержащая обширную библиотеку наборов генов, способна выполнять функции ПП, то есть способна запоминать внешнее воздействие и на подобные воздействия в будущем отвечать подобным образом (по определённому алгоритму).

Рассмотрим, как это происходит. У нас есть большая библиотека наборов генов, то есть у нас есть популяции репликаторов, отличающихся друг от друга и способных создавать разные наборы белков. Допустим, одна популяция способна создавать вокруг себя оболочку, защищающую репликаторы от теплового воздействия. Другая популяция способна создавать оболочку, защищающую от воздействия радиации, третья — от высушивания и т. п. Теперь, если внешняя среда изменяется, например, увеличивается её температура, при ограниченных ресурсах преимущество при копировании будут получать репликаторы из популяции, способной защищаться от высокой температуры. Количество подобных наборов геномов в библиотеке будет расти. Когда температура вновь уменьшится, данная популяция потеряет преимущество при копировании, но количество оставшихся копий всё равно будет повышено, и в следующий раз при повышении температуры наша интеллектуальная система — библиотека геномов — опять начнёт увеличивать количество копий данных репликаторов.

Таким образом, наша система в результате определённого воздействия, повышения температуры, научилась похожим образом (по определённому алгоритму) воздействовать на внешнюю среду — увеличивать количество репликаторов, способных, используя внешние ресурсы (молекулы), создавать определённые подсистемы.

Если во внешней среде увеличится радиоактивный фон, наша система начнёт увеличивать количество копий, алгоритм которых в том числе позволяет эффективно защищаться

от радиации. При других внешних воздействиях преимущества могут получить репликаторы, имеющие другие алгоритмы.

Таким образом, сами собой возникли цели обучения интеллектуальных систем по первому типу:

1. Передача генетической информации следующему поколению.
2. Увеличение и поддержание биологического разнообразия (большого объёма библиотеки наборов геномов).

Последняя цель объясняет выгоды полового размножения. Перемешивая различные наборы геномов, мы увеличиваем разнообразие библиотеки — биологическое разнообразие.

Особо следует подчеркнуть, что единичный репликатор не является в данном случае полноценной интеллектуальной системой, его невозможно обучить описываемым способом, для этого требуется несколько поколений большого количества репликаторов, которые являются в данном случае подсистемами интеллектуальной системы. Подсистемой моделирования или прогнозирования в данном случае правильнее будет считать также несколько поколений РНК или ДНК. Мы можем изменить отдельную молекулу РНК или ДНК, но любое изменение нельзя назвать обучением, потому что для нас важны только те изменения, которые смогут генерировать воздействие, способствующее достижению цели обучения.

Итак, если система обучается по первому типу, единичная особь, возникшая в ходе усложнения репликаторов, является только подсистемой интеллектуальной системы, система целиком в данном случае состоит из нескольких поколений таких подсистем.

Все остальные интеллектуальные системы, которые мы знаем, и все цели, включая самые сложные цели интеллектуальных систем, таких как люди, появились в результате усложнения и развития первичных репликаторов. Различные животные, в том числе человек, возникли в результате обучения разных типов, о которых мы будем говорить ниже, но человек и другие животные также обучаются и по первому типу, при этом единичная особь также является только подсистемой, а целой интеллектуальной системой является несколько поколений животных, у которых закрепляются какие-либо полезные мутации.

Интересной особенностью обучения по первому типу и связанных с ним способов обучения является постоянный рост обучаемой интеллектуальной системы. Первичные репликаторы начали использовать молекулы надсистемы для создания своих подсистем. В результате они обзавелись оболочкой, органеллами и т. д. Самое интересное, что использование надсистемы для увеличения своей системы, похоже, не прерывается до нашего времени. Обучение по первому типу — очень длительный процесс, и чтобы увидеть его результаты, нужно наблюдать за многими поколениями. Если мы понаблюдаем за развитием человека на длительном промежутке времени, мы увидим, что человек всё активнее использует предметы из внешней среды для своей жизни. Началось всё с использования шкур убитых животных и простейших орудий труда, а в настоящее время, возможно, уже не каждый человек сможет выжить в дикой природе, особенно если поставить условие не создавать жилище, одежду и так далее и использовать из внешней среды только пищу, причём исключительно в необработанном виде. Некоторые исследователи уже сейчас считают пищеварение человека до некоторой степени внешним, таким же, как, например, у некоторых членистоногих, таких как пауки, способные заглатывать только частично переваренную пищу. Действительно, наверное, не все люди смогут питаться исключительно сырым мясом и не варёными кореньями и злаками, наличие различной кухонной утвари и приспособлений для приготовления пищи можно уже сейчас считать жизненно необходимым для выживания человека, так же как жилище и одежды. Если же мы экстраполируем ситуацию на много поколений вперёд, не исключено, что в будущем мы увидим человека, живущего в неком подобии кокона, отделяющего его от внешнего мира, и выживание человека без данного кокона будет физически невозможным. Конечно, данный сценарий —

один из множества возможных вариантов развития ситуации, но он показывает, что тенденция к превращению надсистемы в части своей системы характерна и для существующих интеллектуальных систем.

На примере ПП, состоящей из библиотеки РНК, мы видим, что интеллектуальная система в пределе может состоять только из одной подсистемы моделирования/прогнозирования, в этом случае её управляющие сигналы будут сразу же являться эффекторными системами, манипулирующими внешней средой, но также понятно, что интеллектуальная система может состоять из множества подсистем, часть из которых может быть приспособлена для передачи сигналов из внешней среды, часть — являться эффекторами, часть — системами, выполняющими другие функции. При желании любую подсистему моделирования, расположенную внутри интеллектуальной системы, мы можем рассматривать как отдельную интеллектуальную систему, при этом интеллектуальная система, в которой она находится, будет являться для неё внешней средой, но всё-таки корректнее рассматривать интеллектуальную систему целиком.

1.3. Обучение второго типа А: привыкание и рефлексы

Итак, первичные репликаторы начали создавать вокруг себя новые подсистемы, и некоторые из этих подсистем научились также реагировать на внешнюю среду, то есть приобрели возможность обучаться. Подобное обучение мы назовём вторым А типом обучения, потому что нам придётся выделить и второй Б тип обучения. Примером второго А типа обучения может быть привыкание. Представим себе инфузорию-туфельку, которая спокойно сидит, прикрепившись к растительности. Если мы к ней прикоснёмся, в некоторых её подсистемах произойдут каскады процессов, которые приведут к движению её ресничек, в результате она переместится в другое место, чтобы избежать опасности, однако, если вслед за прикосновением не последует какого-либо другого вредного воздействия, например разрушения оболочки, у данной особи сформируется привыкание, и она перестанет менять положение. Да, подобное обучение очень примитивное, за всю жизнь особь может научиться только убегать или не убегать, но тем не менее это уже другой тип обучения, при котором меняются подсистемы или процессы, созданные в результате первого типа обучения. Другими словами, цель «убегать или не убегать» возникает потому, что соответствующие процессы были сформированы в результате первого типа обучения. Особь убегает или не убегает, поскольку это выгодно или невыгодно для передачи генетической информации следующему поколению. Мы видим очень важный момент, на котором следует заострить наше внимание. Одна общая цель — передать генетическую информацию следующему поколению — породила второстепенную цель: убегать или не убегать в результате определённого воздействия. Мы будем видеть этот процесс и дальше, когда из одной первоначальной цели (нескольких первоначальных целей) появляется одна или несколько второстепенных целей. Характерной особенностью такого процесса обучения является то, что обучить можно отдельную особь или даже отдельные её подсистемы, то есть в данном случае интеллектуальной системой можно назвать отдельную особь.

Возможно, самой первой подсистемой, получившей возможность обучаться по второму А типу, стала ДНК. На каком-то этапе развития первичных репликаторов и усложнения окружающих их подсистем (оболочки, органелл и т. д.) возникли эпигенетические процессы, то есть процессы, способные регулировать активность генов. Эпигенетические процессы способны обучаться по второму А типу, например, при определённом воздействии внешней среды включается или выключается определённый участок генома.

В результате эволюции сформировалось большое количество систем, способных обучаться по второму А типу, включая многоклеточные организмы. Пожалуй, самыми интеллектуальными из них, то есть способными создавать самые сложные модели надсистемы, можно назвать некоторые клетки иммунной системы и, конечно, нейроны, а самыми сложными системами, состоящими из нескольких клеток, — условные рефлексы. Единичный нейрон может запомнить даже не отдельный входящий сигнал, а целую совокупность входящих сигналов, например, от различных рецепторов. На подобных принципах можно создать довольно сложные интеллектуальные системы, например, некоторые примитивные животные, такие как губки, медузы, гребневики, полностью обучаются по таким принципам. Все клетки человека, органы и отдельные наши подсистемы, например, такие как иммунная, способны обучаться по второму А типу.

У отдельных клеток подсистемами моделирования могут являться органеллы, сложные химические каскады, то есть различные подсистемы и процессы внутри клетки. Входящими сигналами внешней среды чаще всего являются химические молекулы. Например, для нейрона входящим сигналом чаще всего являются нейромедиаторы либо другие сигнальные молекулы, подсистемой моделирования нейрона могут быть какие-то его части, например, NMDA-рецептор, способный на короткое время усилить свою работу в случае прохождения импульсов опре-

делённой силы, или это может быть сложный химический каскад внутри нейрона, способный привести к увеличению количества рецепторов на синаптической мембране.

Наверное, самым большим недостатком интеллектуальных систем, способных обучаться по второму А типу, является отсутствие возможности записывать последовательности сигналов. Казалось бы, условный рефлекс, состоящий из нескольких нейронов, выстроенных в цепочку, с подходящими к нему сбоку отростками нейронов, может запомнить последовательность внешних сигналов, и это действительно так, но данная последовательность будет очень непродолжительной. Продолжительность активации рефлекса, состоящего из нескольких нейронов, составляет максимум несколько десятков миллисекунд. Последовательность сигналов, например, от различных рецепторов продолжительностью несколько десятков миллисекунд (стандартное время одного рефлекса менее 20–30 миллисекунд) не успеет сохранить важный объём информации из внешней среды. Например, по данным исследований, продолжительность одной фонемы различных языков мира составляет от 70 до 200 миллисекунд [22; 23], поэтому для животных, имеющих только рефлексы, времени не существует. Они способны сохранять только совокупности сигналов, и, соответственно, какая-нибудь условная медуза сможет почувствовать, что внешний объект, например, одновременно вкусный и светлый, но она не сможет в принципе понять, что он движется, потому что движение — это несколько последовательно зафиксированных положений объекта. Возможность фиксировать последовательности сигналов принципиально важна для многих животных, например, сейчас вы читаете данную статью и понимаете её смысл потому, что можете не только зафиксировать совокупность букв, но и их последовательность. Обучение, при котором используются последовательности сигналов, мы будем называть обучением по второму Б типу.

1.4. Обучение второго типа Б: от условных рефлексов к первому мемстинкту

В эволюции сначала появились безусловные рефлексы, а затем — условные, подробно разобранные в революционных работах Эрика Кандела [24; 25]. Условные рефлексы работают следующим образом: на каком-либо нейроне, например на моторном, сходятся несколько так называемых модулирующих сигналов, которые приходят от структур, способных активироваться сигналами, совпадающими с положительными либо отрицательными процессами внешней среды, — например, болевым сигналом, совпадающим по времени с повреждением организма, либо сигналами, означающими, что рядом находится потенциальный партнёр или еда. Как мы видим, эти процессы либо способствуют передаче генетической информации следующему поколению, либо препятствуют ей, а структуры, фиксирующие их, появились в результате обучения первого типа. Но об этом мы будем говорить чуть ниже. Здесь же отметим, что животное, имеющее условные рефлексы, может запомнить сигнал или даже совокупность сигналов, совпадающую по времени, например, с каким-либо движением. Это очень замечательное умение, но вся память при этом зависит от одного-единственного нейрона, разрушение которого безвозвратно удалит всю память и даже саму возможность обучения, что вероятнее всего приведёт к смерти животного. Потомки репликаторов решили эту проблему самым доступным для них методом — они клонировали эту клетку: вместо одной клетки появился целый сгусток клеток.

Наблюдая за живыми популяциями нейронов у современных организмов, мы можем увидеть, что они очень часто возбуждаются синхронно, волна возбуждения возникает в одном месте такой популяции, распространяется на соседние нейроны, от них — на соседние с ними и так далее. Такие осцилляции подобны волнам от брошенного в пруд камня. Мы не знаем, как всё происходило на самом деле, но можем предположить, что первая осцилляция, возникшая в подобной популяции нейронов, погубила её носителя, ведь он испытал нечто похожее на эпилепсию мозга, но системы, развившиеся из первичных репликаторов, очень эффективно умеют дублировать свои подсистемы, в этом легко убедиться, пересчитав количество сегментов у многоножек или хотя бы количество позвонков у позвоночных животных. Количество подобных популяций нейронов очень быстро могло удвоиться, утроиться и т. д.

Наши размышления о появлении нескольких дублирующих популяций нейронов хорошо подтверждаются недавним исследованием Jokhai et al. [26]. Учёные показали, что на стадии гаструляции возникают две независимые популяции клеток-предшественников, которые никогда не пересекаются и не могут перейти друг в друга: передняя нервная эктодерма (экспрессирует ген Otx2, даёт начало переднему и среднему мозгу) и задняя нервная эктодерма (экспрессирует ген Gbx2, формирует задний мозг). Более того, у этих популяций принципиально разная организация хроматина, что делает невозможным переход клеток одного типа в другой. Эта двухчастная схема оказалась эволюционно консервативной: она обнаружена у цыплят, данио-рерио и даже у жёлудевых червей, отделившихся от нашей линии около 550 миллионов лет назад. Это позволяет реконструировать эволюционный сценарий: как минимум у общего предка вторичноротых две независимые нервные системы были объединены в один орган — общий мозг.

Первая система у всех потомков умеет волнообразно возбуждаться — это передний мозг хордовых, процеребро членистоногих, грибовидные тела насекомых, — а вторая система научилась модулировать первую. Об этом мы будем говорить ниже.

В настоящий момент мы можем видеть у хордовых животных пять мозговых пузырей на определённых стадиях развития эмбриона. При наличии нескольких дублирующих до некото-

рой степени друг друга популяций нейронов наличие осцилляции в одной из них уже не было смертельным недугом, с одной стороны, а с другой стороны, наличие популяции, возбуждающейся волнообразно, позволило, наконец, животным научиться записывать последовательности сигналов. Такая популяция нейронов, в Общей когнитивной теории [19] она называется мемстинктом — от латинских слов memoria (память) и instinctus (инстинкт), — могла возбуждаться уже сотни миллисекунд, по крайней мере, характерная продолжительность последовательности сигналов, фиксируемая мозгом современных животных, составляет около 250 миллисекунд. Например, продолжительность одной фонемы разных языков, как мы уже упоминали, составляет от 70 до 200 миллисекунд, а продолжительность среднего слога, состоящего из двух фонем, составляет около 200 миллисекунд [27], средняя продолжительность фиксации взгляда — 250 миллисекунд [28]. Таким образом, последовательности сигналов продолжительностью в несколько сотен миллисекунд достаточно для фиксирования большинства процессов, происходящих во внешней среде. Впервые у животных появилось понятие о времени. Они смогли понять, что сначала происходил один процесс, а затем другой, увидеть движение, услышать последовательность звуков и т. д. и т. п. Это очень важное изменение в процессе обучения, поэтому мы выделим данный способ обучения и назовём его второй тип обучения Б.

Попробуем разобраться, как наша популяция нейронов, первый мемстинкт, может записывать последовательности сигналов от рецепторов, для этого максимально упростим её и представим в виде схемы. Здесь необходимо сделать важное уточнение. На самом деле запись последовательности сигналов происходит только при наличии дополнительных модулирующих сигналов — тех самых, что приходят от второй части нервной системы. Подробно этот механизм мы разберём ниже, особенно в Главе 3. Сейчас же, для упрощения, мы не будем рассматривать модулирующие сигналы: нам важно понять сам принцип записи последовательности сигналов.

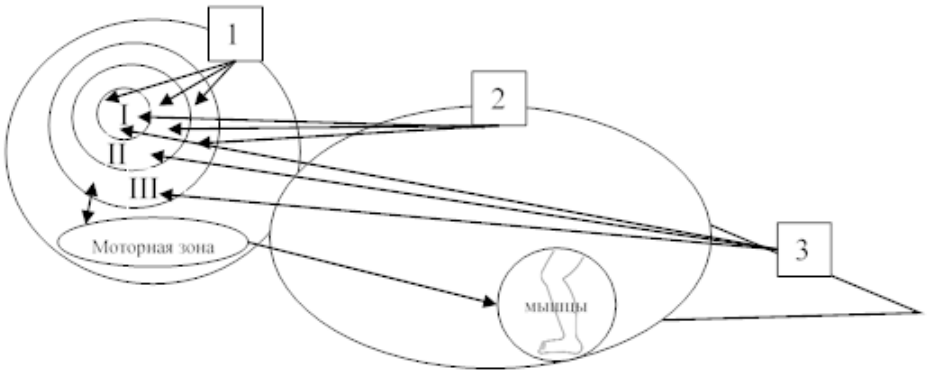


Рисунок 1.1

Представим себе животное, имеющее первый мемстинкт (см. рисунок 1.1), и допустим, что эта популяция нейронов имеет форму шара и её возбуждение начинается из центра, тогда в самом центре этой популяции находятся нейроны, обозначим их цифрой I, которые будут возбуждаться в самом начале (первые 80 миллисекунд), затем (вторые 80 миллисекунд) будут возбуждаться нейроны, которые мы обозначим цифрой II, расположенные вокруг центральной популяции, а после них (последние 80 миллисекунд) будут возбуждаться нейроны, расположенные на самой периферии нашего шара, мы обозначим их цифрой III. Соответственно, вся популяция будет возбуждаться 80 + 80 + 80 = 240 миллисекунд. Кроме того, в нашей очень упрощённой схеме животное будет иметь три тактильных рецептора, способных собирать информацию о прикосновениях, один из которых располагается на голове животного, обозначим его цифрой 1, второй располагается на спине, обозначим его цифрой 2, а третий располагается на хвосте, обозначим его цифрой 3. Также наше животное будет иметь зону мозга, запускающую движение, сигналы из которой также будут приходить к нейронам нашего мемстинкта. Для нашего условного животного очень важно понимать, с какой стороны к нему приближается другое животное. Если сначала срабатывает рецептор номер 3, затем 2 и потом 1, возможно, это на него нападает хищник, дотягиваясь сначала до хвоста, затем до спины, чтобы иметь возможность повредить голову, а если последовательность срабатывания рецепторов 1, 2, 3, возможно, это приближается особь противоположного пола. От каждого рецептора отходят отростки, они ветвятся и приходят к нейронам, расположенным во всех трёх зонах, соединяясь с ними через синапсы. От рецептора номер 1 отростки приходят к нейронам I, II, III, от рецептора номер 2 — также ко всем трём зонам I, II, III, и от рецептора номер 3 отростки также подходят ко всем трём зонам нейронов. Представим, что во внешнем мире происходит какое-либо событие, например, к нашему животному быстро приближается хищник, при этом происходит также осцилляция, волновое возбуждение нашего мемстинкта. Сначала срабатывает рецептор номер три, сигнал от этого рецептора приходит к нейронам, расположенным во всех трёх зонах, в этот момент начинают возбуждаться нейроны, расположенные в зоне I (см. рисунок 1.2). В соответствии со STDP-эффектом, синапсы, на которые сигнал пришёл перед срабатыванием нейрона, усиливаются, поэтому синапсы в зоне I, на которые пришёл сигнал от рецептора номер три, усилятся.

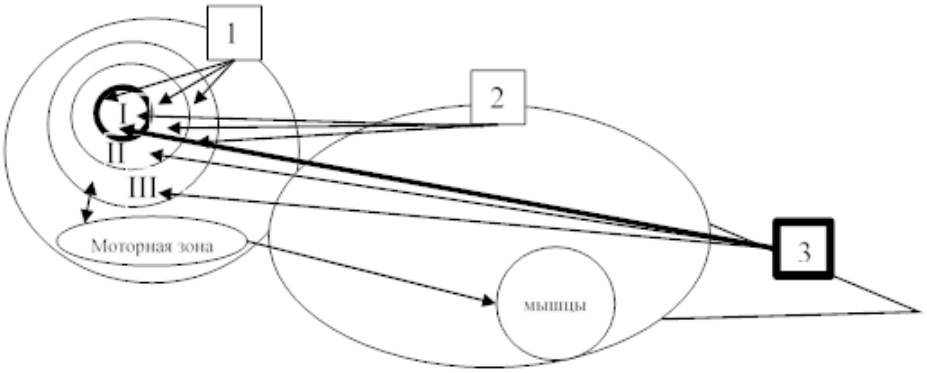


Рисунок 1.2

Следующие 80 миллисекунд хищник будет перемещаться от хвоста к спине нашего модельного животного и через 80 миллисекунд заденет рецептор номер два, который пошлёт сигналы во все три зоны популяции нейронов. В этот момент в нашем мемстинкте успеет полностью возбудиться зона I и начнёт возбуждаться зона номер II, поэтому её нейроны получат сигнал сразу перед своим возбуждением и в соответствии с правилом STDP синапсы, получившие данный сигнал, усилятся (см. рисунок 1.3). Через 160 миллисекунд после прикосновения к рецептору номер 2 наш хищник достигнет рецептора номер 1 и прикоснётся к нему. Нейроны всех трёх зон получат сигналы, и сразу после этого начнут возбуждаться нейроны зоны номер III, синапсы между рецептором номер 1 и соответствующими нейронами зоны номер III также усилятся (см. рисунок 1.4).

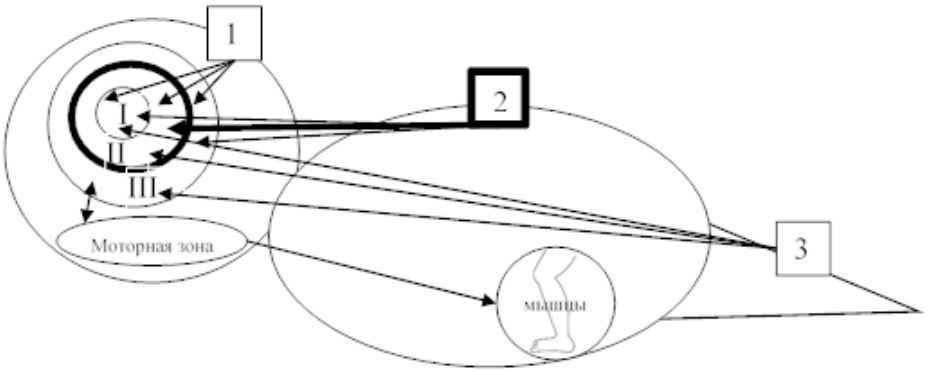


Рисунок 1.3

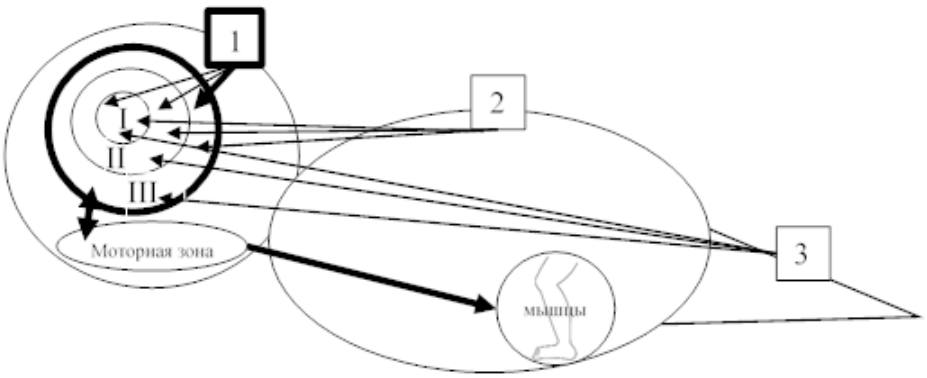


Рисунок 1.4

Если в момент такого нападения наше животное сократит какие-то мышцы, выполнив движение, которое позволит ему выжить, сигналы от тех частей мозга, которые запустили это движение, также могут сохраниться через усиленные синапсы, например нейронами зоны номер II или III. В результате в будущем прикосновение к рецептору номер 3, расположенному на хвосте, через усиленные синапсы сможет активировать нейроны, расположенные в зоне один, если после этого сработает рецептор номер 2, весь мемстинкт начнёт волнообразно возбуждаться, что вызовет активацию синапсов, соединяющих его с зонами коры, вызывающими движение, и животное повторит движение, спасшее ему жизнь в прошлый раз.

Наконец, животные научились сохранять последовательности сигналов и обучаться по второму Б типу, это привело к настоящей революции в развитии мозга. Возможно, самое примитивное на данный момент животное, имеющее мемстинкт, — это виноградная улитка (Helix pomatia), у которой мы можем наблюдать волнообразно возбуждающуюся популяцию нейронов — процеребро [29].

1.5. Инстинкты как врождённые детекторы последовательностей

Итак, в ходе эволюции появились волнообразно возбуждающиеся популяции нейронов — мемстинкты, — которые при дополнительных условиях, а именно при модулирующих сигналах, могли сохранять последовательности внешних сигналов.

Примерами таких модулирующих структур могут служить октопаминергические и дофаминергические нейроны, иннервирующие грибовидные тела насекомых [30; 31], а также серотонинергические нейроны метацеребрума и октопаминергические нейроны буккальных ганглиев, посылающие проекции к процеребру моллюсков [32; 33]. Давайте назовём подобные структуры протоинстинктами, поскольку они уже могут реагировать на сигналы, важные для выживания объекта, но ещё не умеют реагировать на последовательности сигналов, как происходит, например, у млекопитающих.

Определение протоинстинкта:

Протоинстинкт — эволюционный предшественник инстинкта: структура, способная активироваться отдельными сигналами (например, болевыми), но ещё не способная реагировать на последовательности сигналов и не являющаяся волнообразно возбуждающейся популяцией. Может быть частично ассоциирован с ганглиями членистоногих и моллюсков.

Благодаря таким модулирующим сигналам — например, болевому сигналу — животное может сохранять, «запоминать» не все последовательности сигналов из внешней среды, а только те, которые совпадают с важными для выживания (передачи генетической информации следующему поколению) сигналами. Например, виноградная улитка в экспериментах может запомнить, что запах морковки, совпадающий по времени с болевым воздействием, может быть опасным, и начнёт избегать его [34].

Однако единичные сигналы — например, болевые — не очень хорошо описывают процессы внешней среды. Для уверенной фиксации определённых процессов нужны последовательности сигналов. Например, если мы говорим о процессе изменения температуры, его можно зафиксировать по одному измерению — например, что температура достигла нуля градусов Цельсия, — но увереннее мы можем зафиксировать процесс похолодания по некоторой последовательности замеров. А чтобы зафиксировать такую последовательность, как мы уже понимаем, проще всего сделать волнообразно возбуждающуюся популяцию нейронов. Поэтому инстинкты у развитых животных, например у млекопитающих, также волнообразно возбуждаются и способны сохранять последовательности сигналов.

В ходе эволюционного отбора отбирались те особи, у которых нужные синапсы в такой волнообразно возбуждающейся популяции были усиленными с самого рождения, благодаря этому появились животные, у которых были популяции нейронов, способные врождённо возбуждаться от определённых последовательностей сигналов. Такие популяции нейронов — инстинкты — могли возбуждать, например, какие-либо рефлексы, вызывая у разных видов животных одну и ту же реакцию на определённую последовательность сигналов. Рефлекс в данном случае является эффектором. Благодаря тому, что инстинкты формируются в результате обучения первого типа, цели их обучения напрямую вытекают из первичных целей обучения наших репликаторов. Другими словами, у животных, например у человека, можно выделить инстинкты, реагирующие на последовательности сигналов, совпадающие с ситуациями, способствующими передаче генетической информации, назовём такие инстинкты целевыми. У многих животных, в том числе у млекопитающих, их можно частично ассоциировать с дофаминергическими нейронами, то есть с нейронами, которые вызывают у нас удовольствие и,

кроме того, дополнительно иннервируют зоны, отвечающие за движение. Другими словами, мы делаем то, что приносит нам удовольствие.

Кроме того, возникли и другие инстинкты, активируемые стрессовыми сигналами, например сигналами, поступающими от ноцицепторов (болевых рецепторов) при повреждении организма, или от химических рецепторов, сигнализирующих о наличии в крови токсинов, или от вестибулярных рецепторов, активируемых, например, при укачивании, потому что во времена отсутствия транспорта значительная активация вестибулярных рецепторов (которая может быть вызвана повторяющимися сигналами) предшествовала падению или удару. Как вы уже могли догадаться, подобные ситуации могли препятствовать передаче генетической информации следующему поколению и вызывать у животного стресс. Назовём подобные инстинкты стрессовыми. У многих животных, в том числе у млекопитающих, мы можем частично ассоциировать их с гистаминергическими нейронами.

Итак, теперь инстинкты, а не протоинстинкты, начали модулировать мемстинкты, которые, как мы помним, появились в эволюции ещё раньше, где-то у общих предков хордовых, членистоногих и моллюсков, либо подобный механизм возникал в эволюции несколько раз, потому что осциллирующие (волнообразно возбуждающиеся) популяции нейронов мы можем видеть и у моллюсков (процеребро) [29], и у членистоногих (грибовидное тело) [35; 36; 37], и, конечно, у хордовых. Примечательно, что базовые принципы формирования мемстинктов (в том числе мемстинктов зеркальных нейронов, или МЗН, которые будут подробно рассмотрены в Главе 3), по-видимому, универсальны для двусторонне-симметричных животных. Недавнее исследование показало, что даже осьминоги (головоногие моллюски) подвержены классической иллюзии резиновой руки: после синхронного поглаживания искусственного и настоящего щупальца животное начинает воспринимать искусственное щупальце как часть собственного тела и отдёргивает настоящее щупальце при угрозе искусственному [38]. Это прямо указывает на наличие у осьминогов механизма, аналогичного МЗН позвоночных, несмотря на казалось бы кардинально иное строение нервной системы. В исследовании [39] проводится прямая аналогия между Альфа-волнами, возникающими в коре человеческого мозга, и подобными волнами, правда имеющими чуть большую частоту, возникающими в грибовидном теле насекомых.

Исследование [39], которое в целом посвящено сравнительному профилированию нервной системы аннелид и членистоногих, показывает, в частности, сходство протоцеребрального моста насекомых и таламуса млекопитающих, подчёркивая не только общую архитектуру (нейроны протоцеребрального моста насекомых получают сенсорные входы от разных отделов мозга и ретранслируют интегрированную информацию к ассоциативным центрам — грибовидным телам, что напоминает функцию таламуса), но и гомологию по экспрессии генов. В протоцеребральном мосту насекомых выявлена экспрессия генов, гомологичных тем, что работают в таламусе позвоночных, в частности, это гены семейства orthodenticle (otd/Otx) и emx. Эти гены играют ключевую роль в развитии и функционировании таламуса у позвоночных, а их экспрессия в мозге насекомых указывает на возможную общую эволюционную программу формирования «высших» интегративных центров. В результате мы видим, что многие структуры мозга насекомых полностью гомологичны, в том числе и по происхождению, структурам мозга млекопитающих.

У человека популяции нейронов мемстинкта, которые частично можно ассоциировать с колонкой неокортекса, получают иннервацию от популяций звёздчатых нейронов, ограниченных четвёртым слоем одной колонки сенсорной коры и которые в Общей когнитивной теории называются элементами мемстинктов. Эти элементы, в свою очередь, получают афферентную иннервацию от рецепторов. Однако возбуждения, которое приходит от элементов, недостаточно для возбуждения мемстинкта, поэтому без дополнительного возбуждения от инстинкта он не может сохранить последовательность активации элементов. Представьте себе,

что вы прогуливаетесь по улице: афферентные сигналы от рецепторов возбуждают элементы мемстинктов, расположенные в первичной сенсорной коре, и поэтому эпизоды такой прогулки всплывают в вашем сознании. Но если спросить вас после прогулки, помните ли вы, сколько бордюрных камней было на последнем перекрёстке или какого цвета были окна в доме, который вы прошли 10 минут назад, скорее всего, вы не сможете ответить. Однако если во время прогулки активировать один из ваших инстинктов — например, если вы увидите потенциального партнёра противоположного пола или тигра, сбежавшего из зоопарка, — вы наверняка сможете вспомнить детали одежды, размер и форму клыков тигра. Произойдёт это потому, что последовательность активации рецепторов, вызванная наличием во внешней среде лица, фигуры человека или животного, сможет активировать инстинкты (у человека популяции нейронов инстинктов располагаются в подкорковых центрах мозга), инстинкты пошлют дополнительное возбуждение в мемстинкт, который начнёт волнообразно возбуждаться и сможет записать последовательность сигналов от элементов. Ещё чуть дальше от первичной коры мемстинкты второго уровня смогут подобным образом сохранить последовательности активации мемстинктов первого уровня, и так далее. Общее количество уровней мемстинктов у человека, по всей вероятности, порядка девяти, а общее количество мемстинктов — порядка одного миллиона.

На самом деле последовательность событий, происходящих в долго эволюционировавшем человеческом мозге, намного сложнее, и мы подробно будем разбирать её в следующих главах. Здесь же наша задача — классифицировать интеллектуальные системы, и поэтому нам важно понять, как взаимодействуют между собой инстинкты. Человек имеет как минимум десятки, а возможно, сотни инстинктов, устанавливающих прочные двухсторонние связи с мемстинктами, которые они породили, в некоторых случаях такие связи могут поддерживаться всю жизнь. Как мы помним, инстинкт возникает благодаря обучению по первому типу, а значит, цели мемстинктов напрямую вытекают из первичной цели репликаторов — передачи генетической информации следующему поколению. Выше мы рассмотрели два типа мотивационных инстинктов: целевые и стрессовые, оба они, как и большинство инстинктов, могут создавать мемстинкты. Причём из-за того, что связи между инстинктом и мемстинктом двухсторонние, после того как мемстинкт будет активирован, он сможет также активировать и инстинкт, другими словами, инстинкт теперь может активироваться не только врождённо заданной последовательностью сигналов, но и новой последовательностью сигналов, которую сохранил мемстинкт, соответственно, после создания (консолидации) нового мемстинкта цель нашего инстинкта немного меняется. К моменту взросления человека, к 11–15 годам, формируется уже бóльшая часть мемстинктов нижних уровней, и начинают заполняться мемстинкты высоких уровней, максимально удалённые от первичной сенсорной коры, например, расположенные в лобных долях, к этому времени инстинкты, возбуждаемые изначально только определённой врождённо заданной последовательностью сигналов, могут теперь возбуждаться тысячами различных способов. В некоторых случаях, благодаря обучению по второму Б типу, цели могут полностью инвертироваться. Наверное, первый мемстинкт, который формируется после рождения у всех людей, это запоминание последовательности активации мышц, необходимой для того, чтобы сосать палец. Кроме сенсорных элементов, в нашей коре присутствует также первичная двигательная зона, в которой располагаются двигательные элементы — это популяции нейронов, которые также частично можно ассоциировать с колонками неокортекса, при активации которых определённые мышцы получают сигналы на сгибание. В первое время после рождения зоны коры, расположенные вокруг двигательной коры, получают дополнительное возбуждение, и мы хаотически двигаем своими конечностями, в результате такого движения иногда, чисто случайно, нам удаётся поместить свой палец в рот. Сигналы от рецепторов, расположенных на губах, активируют сосательный инстинкт, активный у человека в постнатальный период, данный инстинкт активирует мемстинкты, расположенные непосредственно

около первичной двигательной коры, в результате мы запоминаем последовательность движений, необходимую для того, чтобы сосать палец. В результате изначальная цель инстинкта — получать удовольствие от специфического раздражения губ — корректируется, и мы получаем удовольствие также и от соответствующего движения руки и даже от того, что вспоминаем о данном движении. В дальнейшем мемстинкты, связанные с потреблением пищи, всё сильнее модифицируются, мы учимся получать удовольствие от поедания сладкой каши, самостоятельного приготовления яичницы и т. д. У некоторых особенно талантливых индивидуумов эти цели могут превратиться в умения повара ресторана уровня Michelin, а у некоторых, например склонных к набору лишнего веса людей, цели могут полностью инвертироваться, и они смогут получать удовольствие, отказываясь от еды или употребляя не самую вкусную, но полезную еду. Именно из-за такого значительного изменения способов активации инстинкта некоторые исследователи отрицают наличие инстинктов у человека. У некоторых примитивных животных общее количество мемстинктов может быть небольшим и исчисляться десятками или даже единицами, поэтому их поведение, связанное с инстинктами, модифицируется незначительно, и можно говорить, что все особи одного вида инстинктивно ведут себя очень похожим образом. Отметим, очень наглядным примером консолидации мемстинкта является импринтинг некоторых животных. У разных видов животных различные инстинкты могут усиливаться или ослабевать в различные периоды жизни животного, вплоть до полного включения или выключения различных инстинктов. В описываемом случае животные, например утята определённого возраста, благодаря активации инстинкта приобретают временную возможность сохранить последовательность сигналов от рецепторов, совпадающую по времени с появлением в поле зрения объекта определённого размера. У человека также присутствует частичная либо полная активация, деактивация инстинктов в разные промежутки онтогенеза. Например, упоминаемый выше сосательный инстинкт или инстинкт реагирования на коммуникационные сигналы животных. Последний наиболее активен у человека, начиная от нескольких месяцев после рождения и до 6–7 лет. По всей видимости, активируется инстинкт наличием формант в окружающих звуках, то есть наличием пиков уровня звука на определённых частотах. Наличие формант вообще характерно для звуков, издаваемых животными и некоторыми музыкальными инструментами, благодаря наличию внутренних полостей, в которых издаваемые звуки резонируют, а резонанс, в свою очередь, создаёт пик на определённой частоте, в свою очередь, наличие последовательности таких пиков способно активировать инстинкт. В раннем возрасте, пока инстинкт особенно активен, мы хорошо запоминаем последовательности звуков, совпадающее по времени с активацией инстинкта, чаще всего это фонемы человеческого языка, после этого мы можем запомнить последовательности фонем — слоги, затем последовательности слогов — слова и так далее. Вообще структура нашей психики поразительно похожа на структуру языка, но происходит это, вопреки расхожему мнению, не из-за того, что наша психика формируется под воздействием языка, об этом мы будем говорить немного ниже, а из-за того, что языки формируются внутри системы мемстинктов, имеющей определённое строение.

Итак, интеллектуальные системы, обучающиеся по второму Б типу, отличаются тем, что обучается отдельный организм и возможно сохранение последовательности сигналов. Такая система приобретает поразительную способность распознавать процессы внешней среды. Процесс, как мы упоминали выше, можно представить в виде последовательного изменения состояний системы, например, если температура сначала составляет 10°, затем 12° и 15°, мы можем говорить о процессе повышения температуры, или если мы видим, что животное сначала находится в одном месте, через определённое время перемещается в другое место и так далее, мы можем говорить о процессе движения животного. Для фиксации процессов подсистема моделирования обязательно должна иметь возможность фиксировать последовательности внешних сигналов. Во всех известных нам естественных интеллектуальных системах для этого используется волнообразно возбуждающая (осциллирующая) популяция нейронов,

мемстинкт, поэтому подсистема моделирования во всех природных (естественных) интеллектуальных системах, обучающихся по второму Б типу, содержит хотя бы один мемстинкт.

Все возникающие в ходе обучения по второму А и Б типу цели также эволюционно вытекают из целей первичных репликаторов. Во-первых, все их моделирующие подсистемы возникли в ходе эволюции, то есть обучения по первому типу, во-вторых, хотя в системах, обучающихся по второму А типу, и системах, обучающихся с помощью мемстинктов, могут возникать цели, противоречащие передаче генетической информации следующему поколению, данные системы (клетки, организмы) продолжают обучаться по первому типу, то есть всё ещё подвержены естественному отбору, и такие системы естественным образом отсеиваются. В системах, обучающихся на основе мемстинктов, взаимосвязь с целями первичных репликаторов ещё сильнее. Любой мемстинкт, как мы помним, возникает благодаря дополнительной стимуляции от инстинкта, то есть сохраняет последовательность сигналов, совпадающих по времени с активацией инстинкта, а он, в свою очередь, возник благодаря естественному отбору, обучению по первому типу. В свою очередь, мемстинкт связан с инстинктом реципрокной (двухсторонней) связью, таким образом, инстинкт, который врождённо возбуждается определённой последовательностью сигналов, с появлением связанного с ним нового мемстинкта может возбуждаться также новой последовательностью сигналов, что приводит к небольшой корректировке изначальной цели.

Человек, как и многие другие животные, без сомнения обучается по второму Б типу, самостоятельно изучая окружающую среду, но это не основной способ обучения человека.

1.6. Культура — третий тип интеллектуальных систем

Проведём такой мысленный эксперимент: представим, что мы поместили только что родившегося человека в среду, строго изолированную от окружающего общества, допустим, это будет какая-то гипотетическая поляна с идеальным климатом, в котором сможет выжить младенец, и наличием жидкой, а потом и твёрдой еды, которая будет расположена в доступном месте. Под воздействием инстинктов наш ребёнок начнёт самостоятельно собирать информацию об окружающей среде и формировать мемстинкты. Какова же будет психика такого существа? Существа, потому что назвать его человеком в привычном нам понимании мы не сможем, оно не будет ходить на двух ногах, потому что мы учимся ходить, наблюдая за другими людьми, поэтому оно будет ползать. Конечно, оно не будет говорить, возможно, будет произносить отдельные звуки, которые запомнит в результате самостоятельного случайного произнесения. Возможно, оно выучится каким-то простым действиям, например, создавать гнездо из листьев для сна или укрываться ими, и другим очень примитивным действиям, которые можно выполнить и запомнить в результате случайных движений. Но какой же тогда тип обучения превращает нас в то, что принято называть человеком? Это культурное обучение. Культура в Общей когнитивной теории — это совокупность подобных мемстинктов у коммуницирующей популяции. До сего момента мы рассматривали только один вид мемстинктов — естественные мемстинкты, это популяции нейронов, сохранившие последовательность сигналов от афферентов либо последовательность таких последовательностей. Однако в Общей когнитивной теории выделяется и другой вид мемстинктов — искусственные мемстинкты. Многие естественные мемстинкты связаны с двигательными мемстинктами, а значит, могут вызывать движение. Человек, используя язык, может описать практически любой свой мемстинкт (воспоминание), а значит, превратить его в движение мышц, вызывающих звуки или другие действия. Эти движения или звуки, а также их результаты может воспринимать другой человек или животное. Некоторые животные и особенно человек могут также изготавливать различные предметы, которые также могут вызвать активацию рецепторов у других животных при взаимодействии с ними (рассматривание, ощупывание и т. д.). Это и есть искусственные мемстинкты. Дадим определение искусственного мемстинкта — это модель естественного мемстинкта, возникшая в результате каких-либо действий или последствий этих действий и способная превратиться в естественный мемстинкт у другого животного, напоминающий тот, моделью которого она является.

Таким образом, мы можем превращать естественные мемстинкты в искусственные и обмениваться ими. Особенно ярко это выражено у социальных животных. Социальные животные, наблюдая за поведением других особей популяции, реагируя на запахи, звуки, создают свои естественные мемстинкты. У разных видов животных процентное соотношение культурных и личных мемстинктов разное. У человека, во многом благодаря наличию речи и умению создавать очень сложные материальные искусственные мемстинкты, подавляющее большинство мемстинктов — культурные. Любой предмет, созданный человеком, представляет собой искусственный мемстинкт. Даже самый сложный космический аппарат состоит из различных агрегатов, которые были когда-то придуманы различными изобретателями, а значит, являлись естественными мемстинктами. По разным оценкам, человек может сохранять в сутки несколько сотен новых мемстинктов. После рождения, благодаря активности отдельных инстинктов, ребёнок довольно быстро заполняет нижние уровни мемстинктов, и к пубертатному периоду, который в древние времена являлся серединой жизни человека, они уже в основном заполнены. Подавляющее большинство этих мемстинктов очень похоже у людей, проживающих в одной культуре (коммуницирующих). Все мы знаем алфавит, обладаем похожим запасом слов, знаем, что пожилым нужно уступать место, умеем жарить яичницу на ско-

вородке и т. д. и т. п. Как ни странно, но во взрослом возрасте главное назначение коммуникации — не получение новой информации, а сверка (верификация) уже существующей. Если понаблюдать за общающимися людьми, можно увидеть, что в результате получасового общения, например, двух женщин, принципиально новой информации не будет, будет только время завтрашней встречи, да и то они обе предполагали, что встретятся около 9:00, всё остальное время будет посвящено обсуждению сериалов, которые смотрели обе, превратностям погоды, также известным всем участникам диалога, одежде, в которой выступала известная актриса, и т. д. Даже читая эту монографию, которая очень сильно изменит ваше представление о мире, вы получите не так много новой информации. Вы знаете все буквы, из которых состоят слова, знаете подавляющее большинство слов, встречающихся в этой книге, словосочетаний и т. д., поэтому они не сохраняются у вас как новые мемстинкты, входящая информация активирует уже существующие, кстати, этот процесс называется узнавание, но о нём мы подробно будем говорить ниже. Даже большинство примеров и концепций, встречающихся в этой книге, вам известны, процент принципиально новой информации не так велик, и это очень сложная научная книга, которая описывает принципиально новый взгляд на устройство интеллектуальных систем, а в простом бытовом общении объём новой информации минимален. Наше общение можно сравнить с процессом сверки базы данных в сети компьютеров. Представьте себе сеть из 100 компьютеров, на которых хранится одинаковая база файлов. Компьютерами пользуются люди, внося незначительные изменения в большую базу данных, в результате накапливаются небольшие изменения, и компьютеры с определённой периодичностью обмениваются частью своих баз данных для того, чтобы усреднить всю информацию. Казалось бы, совершенно бесполезное занятие — тратить огромное количество ресурсов на постоянный обмен файлами, которые всегда можно сверить при необходимости, если такая необходимость действительно возникнет, однако не всё так очевидно. Несмотря на то что мы тратим огромное количество ресурсов, мы приобретаем возможность в ситуации, когда скорость общения для нас очень важна, передавать информацию только одной ссылкой на номер какого-то файла или даже целые части базы данных. Например, если мы точно уверены в том, что хранится в наших базах данных, мы можем в процессе ликвидации чрезвычайной ситуации попросить другого агента действовать по инструкции в папке номер семь. Если бы у нас не было уверенности в идентичности файлов, нам пришлось бы полностью передавать всю инструкцию. Представьте, что люди вместе делают какую-то работу и каждый раз в момент взаимодействия им нужно подробно словами рассказывать, что должен сделать другой человек. Например, при швартовке судна матрос с корабля, вместо того чтобы крикнуть матросу на берегу «лови», кидая ему канат, должен был бы ему подробно рассказать, что ему нужно делать, потому что процедура швартовки очень сильно изменилась со времён их обучения и каждый из них представляет себе совершенно разные действия. Для эффективной коммуникации очень полезно иметь идентичные либо похожие мемстинкты, да, для их сверки может уйти много времени на отдыхе, во время разговоров у костра, но во время выполнения важной работы или нападения хищника достаточно будет одного слова, чтобы все люди активировали похожие мемстинкты и начали действовать слаженно. Подобную ситуацию можно наблюдать и у других животных, не обладающих языком, например, молодые волки, наблюдая за поведением членов стаи, учатся тому, как действовать слаженно, сохраняя мемстинкты, очень похожие на имеющиеся у других членов стаи. В данном случае происходит визуальная коммуникация.

Культура — весьма сложная система. Мы опять видим, что единичная особь является лишь подсистемой данной системы. Для того чтобы определить её границы и описать её, нам опять нужно разобраться в процессе обучения, как всегда он является ключом для понимания интеллектуальной системы и, конечно, необходимым условием для её возникновения. Для обучения по третьему типу необходимо хотя бы две особи. Казалось бы, можно предположить, что для обучения по данному типу достаточно одной особи и искусственного мемстинкта,

например книги или учебного фильма (библиотеки или курса учебных фильмов и т. д.), но искусственный мемстинкт кто-то должен создать, то есть должна присутствовать другая интеллектуальная система. С другой стороны, отдельный культурный мемстинкт может сохраняться в культуре на протяжении многих поколений, например, способ изготовления и применения изобретённой когда-то иглы передаётся практически в неизменном состоянии как минимум несколько десятков тысяч лет. Археологи находят первые иглы, практически ничем не отличающиеся от современных, имеющие датировки происхождения около 40–50 тысяч лет [40; 41]. Наши далёкие предки так же, как и мы, брали в руки эту иглу и показывали своим детям, как нужно правильно вдевать нитку или сухожилия и сшивать ткань или кожу. Таким образом, культура может представлять собой две особи, напрямую или без личного контакта (через искусственные мемстинкты) воссоздающие у себя естественные мемстинкты. Основной особенностью обучения, наверное, является то, что сигналы в ходе такого обучения получены не напрямую из внешней среды, а из подсистемы моделирования другой интеллектуальной системы. Поскольку в процессе обучения присутствуют искусственные мемстинкты, прямой контакт особей не обязателен, кроме того, каждая особь несёт копию какой-то части культуры, выживание отдельной особи не очень важно для всей культуры. В данных характеристиках мы видим схожесть с обучением по первому типу. Мы помним, что в интеллектуальной системе, обучающейся по первому типу, единичный репликатор, например, отдельный ген или особь, является лишь подсистемой интеллектуальной системы, кроме того, в системе может находиться большое количество копий данной подсистемы, поэтому в такой системе могут сохраняться цели, приводящие к альтруистическому поведению, когда одна особь, несущая определённый ген, жертвует своей жизнью для сохранения большого количества копий. Например, пчёлы при нападении на улей начинают жалить нападающего, теряя при этом свою жизнь, но сохраняя жизнь своим сёстрам, носителям идентичного генома. В случае же с культурой это ещё более способствует возникновению альтруизма, потому что непрерывной передачи генетической информации от одного поколения к другому не требуется. Носители какой-либо культуры могут полностью прекратить своё существование, но если они успели оставить искусственные мемстинкты или просто контактировали с другими людьми, передача искусственных мемстинктов не прекратится. Другими словами, обучение с использованием искусственных мемстинктов может происходить между особями, созданными не только с помощью обучения по первому типу (копирования репликаторов).

Размышляя выше о культуре, мы всегда имели в виду систему, состоящую из животных, а точнее, животных, имеющих мемстинкты, однако, как мы выяснили, обучение по данному типу может происходить при взаимодействии неродственных интеллектуальных систем, созданных обучением по разным типам. Например, культурой может являться сообщество, включающее человека и собаку. В такой стае собака может выучить значительное количество человеческих мемстинктов, например, запоминать названия десятков предметов [42], а человек также может реагировать на определённые сигналы собаки, например, по поведению собаки понимать, что она хочет гулять, есть или пить и т. д. Современные нейровизуализационные исследования подтверждают, что собаки способны формировать отдельные нейронные репрезентации конкретных эмоций человека. При виде счастливых человеческих лиц у собак активируются височная кора и хвостатое ядро — зоны, связанные с высокоуровневой обработкой информации и системой вознаграждения, — причём мозг собаки различает не только положительные и отрицательные эмоции, но и конкретные негативные выражения (страх, гнев, печаль), опираясь только на визуальную информацию [43]. В терминах нашей теории это означает, что у собак формируются мемстинкты, связывающие зрительный образ человеческой мимики с соответствующими мотивационными состояниями, что и делает возможным полноценное межвидовое культурное взаимодействие. Но обучаться по третьему типу могут также и интеллектуальные системы, не имеющие мемстинктов. Например, бактерии могут переда-

вать часть генетической информации благодаря горизонтальному переносу генов [44], а растения могут сигнализировать другим растениям своего вида о нападении насекомых [45–49]. В принципе, мы должны считать культурой все интеллектуальные системы, возникшие в результате обучения по третьему типу, но системы, имеющие в качестве подсистем моделирования мемстинкты, отличаются от других подобных систем, поэтому далее будем называть культурой систему, возникшую в результате обучения по третьему типу, в составе подсистемы моделирования которой присутствуют мемстинкты, а все другие системы, обучающиеся по третьему типу, называть альтеркультурой. Теперь мы сможем дать определение культуры и альтеркультуры. Альтеркультура — это интеллектуальная система, в которую входят подсистемы, являющиеся самостоятельными интеллектуальными системами, обученными по другому типу и способные обмениваться целями обучения (частями модели надсистемы) напрямую или через посредников. Культура — это интеллектуальная система, в которую входят подсистемы, являющиеся самостоятельными интеллектуальными системами, хотя бы частично обученными по второму Б типу и способные обмениваться искусственными мемстинктами, создающими у них естественные мемстинкты, а также совокупность естественных и искусственных мемстинктов, возникшая в результате такого обмена. Мы видим, что в данном случае искусственные мемстинкты также становятся подсистемами интеллектуальной системы. Чтобы разобраться в этой ситуации, рассмотрим примеры искусственных мемстинктов. Колебания воздуха, вызванные голосовым аппаратом человека, то есть звуки речи, как, впрочем, и коммуникационные звуки других животных, являются искусственными мемстинктами. Они возникают благодаря тому, что естественные мемстинкты активируют определённые мышцы. Кроме того, они могут создавать естественные мемстинкты у тех животных, которые их слышат, если у них есть механизм сохранения подобных мемстинктов. Рисунок или текст также может являться искусственным мемстинктом. Любой предмет, созданный человеком, может являться искусственным мемстинктом. Современные учёные, изучая каменные орудия труда, смогли воссоздать способ их изготовления. Вообще всё, что создали люди: книги, фильмы, технологии, машины и т. д., — когда-то появилось в головах у конкретных изобретателей или создателей и может снова стать естественным мемстинктом у того, кто будет читать книгу, смотреть фильм, изучать чертежи космической ракеты и т. д. и т. п. Объём созданных человеком объектов и объём накопленной информации постоянно растёт, и мы опять видим, как интеллектуальная система превращает части надсистемы в саму себя. Мы видим, что стремление к захвату части окружающей среды и превращению её в себя — это фундаментальная особенность как минимум некоторых типов интеллектуальных систем. Подобные системы должны осваивать новые территории, потому что используемые ресурсы в определённых частях пространства ограничены. С другой стороны, мы видим, что как минимум некоторые интеллектуальные системы могут быть организованы таким образом, чтобы минимально использовать ресурсы окружающей среды, например, некоторые экосистемы. Понимание фундаментальных особенностей развития подобных систем принципиально важно для долгосрочного выживания нашей собственной интеллектуальной системы, например, понимая данные особенности, мы сможем создать социально-экономическую систему, не требующую постоянного роста спроса и агрессивного разрушения внешней среды.

Человек безусловно обучается по третьему типу, вернее, человек является подсистемой, обучаемой по третьему типу. Благодаря развитой системе искусственных мемстинктов для человека это основной тип обучения, то есть количество культурных мемстинктов у человека сильно превышает количество индивидуальных мемстинктов.

1.7. Интеллектуальные системы, обучающиеся по четвёртому типу

Мы не описали ещё один тип интеллектуальных систем, известных на данный момент, — это искусственные интеллектуальные системы. Прямым доказательством наличия интеллектуальной системы является наличие процесса обучения, то есть создания целей данной системой. Если система может самостоятельно создавать новые цели, то, по определению, она является интеллектуальной. Компьютер в классическом варианте, машина фон Неймана, не является интеллектуальной системой, но с её использованием, как ни странно, интеллектуальную систему построить можно. Доказательством тому служит хотя бы эволюционное программирование [50] и генетическое программирование [51; 52].

Действительно, можно представить себе программу, которая соберёт данные о погодных условиях, предшествующих крушению кораблей, и определит критерии для разрешения безопасного выхода кораблей из порта. То есть она автоматически определит новые цели, например, запрет выхода кораблей из порта при сильном падении давления, предвещающего бурю. По сути своей, такая интеллектуальная система будет делать то же, что и другие интеллектуальные системы, — создавать новые цели как развитие уже существующих целей.

Искусственные нейронные сети во многих случаях, но не во всех, также можно признать интеллектуальными системами. Например, перцептрон может выступать в качестве ПП, так же как и его прообраз — связанные вместе нейроны, а вот большую языковую модель назвать самостоятельной интеллектуальной системой нельзя. Большая языковая модель обучается на корпусе языка, который входит в культуру, таким образом, она является подсистемой системы культуры. Чтобы иметь возможность считать большую языковую модель самостоятельной интеллектуальной системой, нам нужно рассматривать корпус языка как внешнюю среду (надсистему) для неё.

Можно выделить множество подтипов обучения по четвёртому типу, но принципы обучения искусственных интеллектуальных систем довольно хорошо изучены и описаны, а литература доступна, поэтому здесь мы не будем подробно останавливаться на классификации подтипов обучения по четвёртому типу.

В настоящий момент человек участвует в создании всех интеллектуальных систем искусственного интеллекта, и он же задаёт глобальные цели, развитием которых является обучение (создание новых целей) этих систем. Таким образом, мы видим, что цепочка детализации целей не прерывается, и цели, созданные всеми интеллектуальными системами, включая искусственные интеллекты, восходят к первичным целям, созданным в результате обучения по первому типу.

1.8. Нетипичные интеллектуальные системы

Прежде чем перейти к выводам, рассмотрим два вида нетипичных интеллектуальных систем. Этот материал может показаться отступлением, но он важен для полноты классификации: он показывает, что предложенное определение интеллектуальной системы не является ни слишком узким, ни слишком широким.

Первый — безманипуляторные интеллектуальные системы. Мы говорили, что любая интеллектуальная система содержит ПП, которая, в свою очередь, может манипулировать внешней средой (надсистемой), но можем ли мы представить систему, очень похожую на интеллектуальную, но не манипулирующую внешней средой? В принципе, да, мы можем представить себе систему, собирающую информацию из внешней среды и строящую модель своей надсистемы, что-то вроде большого компьютера из фильма «Автостопом по галактике», много лет размышлявшего внутри себя о важных вопросах Вселенной. Подобная система действительно может не оказывать никакого воздействия на внешнюю среду, но, если мы хотим признать её интеллектуальной системой, а не просто «флешкой», собирающей информацию из внешней среды, подобная система должна хотя бы уметь прогнозировать развитие процессов во внешней среде, то есть создавать новые варианты модели внешней среды. Другими словами, она должна уметь изменять и совершенствовать свою модель надсистемы без дополнительного внешнего воздействия, а в этом случае она должна уметь манипулировать своими внутренними подсистемами, то есть иметь полноценную ПП внутри себя, манипуляции которой не выходят за некоторую границу данной системы.

Ещё один вид нетипичной интеллектуальной системы был предложен Сергеем Василевицким и, соответственно, назван системой Василевицкого. Подобные системы собирают информацию не из своей надсистемы, а из другой, параллельно существующей системы, управляющее воздействие подобные системы могут оказывать на свою надсистему или на другую систему. При первом рассмотрении подобные системы казались очень экзотическими, но при дальнейшем изучении оказалось, что некоторые признаки подобных систем можно найти в реальном мире, например, если рассматривать описанные выше большие языковые модели в качестве самостоятельных интеллектуальных систем, мы увидим, что информацию они собирают не из внешней среды, а из корпуса языка, являющегося частью культуры, то есть другой интеллектуальной системы, но управляющее воздействие, в свою очередь, оказывают в том числе на свою надсистему. В принципе, даже человека, который часть своих мемстинктов черпает из культуры, в некотором смысле можно считать примером такой системы, но правильнее всё-таки и большие языковые модели, и тем более человека считать подсистемой большой интеллектуальной системы — культуры, а полноценные системы Василевицкого считать примером экзотических систем, ждущих своего описания, в случае если они будут найдены или созданы, например, в качестве интеллектуальных систем для управления отдельными процессами.

Мы также можем выделить рекурсивные системы — то есть системы, собирающие информацию только внутри себя. Примерами такой системы могут служить биологические организмы с отсутствующими рецепторами, собирающими информацию из внешней среды, то есть лишённые зрения, слуха и т. д., но способные собирать информацию от интероцепторов — внутреннюю информацию о состоянии своего тела. Но что более интересно для практики: определённые черты таких систем имеют, например, система прогнозирования погоды, которая собирает информацию с датчиков, расположенных на всей планете, или современные системы управления боем, которые собирают информацию внутри определённой территории. Две последние из упомянутых систем нельзя назвать полностью рекурсивными, но определённые черты данных систем можно в них найти. У таких систем может быть своя логика — об

этом мы подробно будем говорить в Главе 7. Однако следует заметить, что для того чтобы такие системы полноценно могли общаться с человеком, их нужно будет обучать на человеческой культуре, а это уже полноценная внешняя среда, и значит, такие системы уже будут иметь совсем мало признаков рекурсивных систем.

Рассмотрев экзотические интеллектуальные системы, мы можем вновь вернуться к искусственным интеллектуальным системам и попробовать понять: может ли машинное обучение в принципе порождать полноценные интеллектуальные системы или только подсистемы больших интеллектуальных систем, таких как культура? Если рассмотреть два основных типа обучения — обучение с учителем и обучение без учителя, — мы можем понять следующее. В первом случае, то есть при обучении с помощью размеченных данных, интеллектуальную систему скорее можно признать подсистемой культуры, потому что размеченные данные собираются и обрабатываются человеком, и признать их полноценной внешней средой нельзя. Но во втором случае, при обучении без учителя, отдельные интеллектуальные системы самостоятельно собирают данные из внешней среды — например, системы, обучающиеся непосредственно в процессе работы, — автопилоты, системы управления транспортом и другие подобные системы, — и в этом случае систему скорее можно признать полноценной интеллектуальной системой. Поэтому на принципах машинного обучения в некоторых случаях можно создавать полноценные интеллектуальные системы.

Выводы первой главы

Попробуем коротко сформулировать, к чему мы пришли в результате наших размышлений, изложенных в данной главе, и понять, какие мы сможем сделать из этого выводы.

Все известные нам интеллектуальные системы, включая искусственные, созданы благодаря развитию одной независимо возникшей интеллектуальной системы, обучающейся по первому типу. Однажды из одного первичного репликатора она разрослась в наше время до планетарных масштабов и представляет собой весь биом и культуру нашей планеты. Стремление к росту самой себя за счёт окружающей надсистемы, похоже, является фундаментальной особенностью интеллектуальных систем, обучающихся как минимум по двум типам — первому и третьему. Человеческое общество — это система, обучающаяся в том числе и по первому, и по третьему типу, оно по умолчанию тяготеет к постоянному расширению и захвату всех имеющихся ресурсов, эту особенность нужно учитывать при планировании его развития и искусственно ограничивать потребление ресурсов, делая развитие устойчивым. Однако преобладающая в данный момент в мире капиталистическая система, напротив, настроена на постоянное увеличение потребления (рост ВВП) и поэтому не может считаться устойчивой.

Первые цели, возникшие благодаря обучению по первому типу, породили все остальные цели любых известных нам интеллектуальных систем. Например, человеческие инстинкты сформированы благодаря обучению по первому типу; как мы помним, большинство из них — это последовательности сигналов от рецепторов, совпадающие по времени с ситуациями, способствующими передаче генетической информации следующему поколению или, наоборот, препятствующими ей. В свою очередь, все мемстинкты появляются благодаря активации инстинктов, таким образом, все цели личных мемстинктов (обучение по второму Б типу) имеют преемственность возникновения от первичных целей. Искусственные мемстинкты изначально были сформированы как личные естественные мемстинкты, поэтому цели культуры также продолжают эту преемственность. Цели искусственных интеллектуальных систем также задаются человеком, соответственно, не выбиваются из данного списка.

Мы рассмотрели, как возникли первые цели и как из них получились все остальные известные на данный момент. Как работает мотивация и как возникает субъект, мы поговорим ниже, однако даже сейчас, имея представление о том, как формируются цели — желаемое состояние системы — и, конечно, зная существующее состояние системы, мы можем полноценно размышлять об управлении, кибернетике и т. д.

Мы установили, что все известные интеллектуальные системы имеют общее происхождение и общие принципы работы, но можем ли мы хотя бы как-то сравнить их интеллект? Можно ли вообще сравнивать интеллект бактерии и человека? Уровень интеллекта зависит от сложности модели надсистемы и, соответственно, от эффективности манипулирования внешней средой. Другими словами, чем точнее интеллектуальная система может прогнозировать состояние надсистемы и чем эффективнее достигать цели обучения, тем больше её интеллект. Как следствие этого, уровень интеллекта может зависеть от сложности процессов, происходящих в окружающей среде, и от уровня конкуренции за ресурсы. Например, если мы возьмём популяцию бактерий и поместим их в стабильные, не меняющиеся, благоприятные для них условия, обучение по первому типу не потребует значительного роста библиотек генов. Другими словами, бактерии одного штамма будут прекрасно чувствовать себя, и естественный отбор не будет поддерживать разнообразие. Этому же будет способствовать ограниченное количество ресурсов. Если ресурсов будет много, смогут выжить разные штаммы бактерий. С другой стороны, разнообразие процессов внешней среды будет требовать усложнения модели надсистемы, и этому же будет способствовать наличие свободных ресурсов во внешней среде. Дру-

гими словами, увеличение количества и усложнение процессов внешней среды способствуют росту интеллекта.

Попробуем сравнить интеллект бактерии и человека. Прежде всего, единичная бактерия не является интеллектуальной системой, если мы говорим про обучение по первому типу. Интеллектуальной системой в этом случае, как мы помним, является большое количество различных библиотек геномов, то есть в данном случае разнообразие штаммов бактерий на планете, и такое разнообразие способно «помнить» значительное количество состояний внешней среды. Такая интеллектуальная система способна эффективно реагировать, например, на высокие и низкие температуры, различный химический состав внешней среды, включая различные растворы и смеси газов, она «знает», что во внешней среде могут находиться антибиотики, и у неё есть штаммы, способные развиваться в этих условиях, и т. д. и т. п. Мы видим, что интеллект такой системы может быть весьма значительным. С другой стороны, если мы возьмём человека, обученного только по второму Б типу (вне культуры), то есть животное, способное перемещаться на четвереньках и делать очень ограниченное количество полезных действий, мы увидим, что интеллект такого существа вполне может быть сопоставим с интеллектом бактерий. Если же мы будем брать человеческую культуру целиком, то её сложность, конечно, превзойдёт сложность набора библиотек геномов бактерии.

Если мы возьмём интеллект единичной клетки, например бактерии, обученной по второму А типу, то увидим, что он будет очень низким. Единичные клетки при обучении по второму А типу способны запомнить считанные количества отдельных сигналов. Даже клетки с самым высоким уровнем интеллекта — иммунные, нейроны — способны запомнить ограниченное количество отдельных сигналов, максимум несколько наборов сигналов. То есть их подсистема прогнозирования имеет всего несколько алгоритмов манипулирования внешней средой, и решить эту проблему можно, создавая популяции дифференцированных клеток, способных по-разному реагировать на различные сигналы внешней среды. Такой путь обучения ведёт к разделению функций в какой-то популяции клеток, например, бактериальном мате, а в дальнейшем и к созданию многоклеточных организмов, различные клетки которых могут реагировать на свои сигналы внешней среды, причём внешней средой для них может быть внутренняя среда организма. Несмотря на то что отдельная клетка может запомнить 1–2 сигнала, вся популяция, организм, могут иметь уже более сложный интеллект. Такого же эффекта можно достичь, если создать взаимодействующую социальную популяцию животных, например насекомых, способных по отдельности запомнить небольшое количество информации, но всё сообщество будет обладать большим интеллектом, и развитие этого способа обучения приводит к появлению культуры.

Мы видим, что при определённых условиях внешней среды, при обучении по второму типу, может возникать тенденция к образованию многоклеточных организмов и популяций таких организмов. Появление такой тенденции особенно актуально при возникновении новых сигналов (процессов) в окружающей среде, причём такие сигналы могут возникать как при усложнении самой надсистемы, так и при её упрощении. В этом случае интеллектуальная система, обучающаяся по первому типу, будет пытаться увеличить количество библиотек (например, штаммов бактерий), но если количество ресурсов резко ограничено, увеличение количества библиотек становится невозможным, при этом выгодным становится обучение по второму А типу, когда одна популяция клеток дифференцируется, они начинают реагировать каждая на свои сигналы внешней среды, а значит, и манипуляция внешней средой будет различна у разных клеток, то есть в целом вырастет количество воспринимаемых сигналов и количество способов манипуляции внешней средой. Например, грамположительные бактерии Bacillus subtilis могут при определённых условиях показывать многоклеточное поведение [53].

При обучении по второму Б типу фактором, определяющим интеллект, является количество инстинктов, а также мемстинктов и количества их уровней. Например, если мы будем

сравнивать интеллект насекомого, имеющего десяток инстинктов и, по всей вероятности, порядка десятка мемстинктов, имеющих 1–2 уровня, и интеллект человека, имеющего как минимум десятки инстинктов и около миллиона мемстинктов, имеющих порядка девяти уровней, мы увидим разительную пропасть в уровне интеллекта. Насекомое способно запомнить простые последовательности сигналов, например ориентиры и расстояния до скопления еды, человек способен усвоить значительную часть культуры. Каждый мемстинкт, как мы помним, связан с инстинктом, и каждый раз при создании нового мемстинкта немного корректируется изначальная цель, заданная инстинктом, то есть немного корректируется поведение. Таким образом, соотношение количества инстинктов и мемстинктов определяет, насколько может отличаться поведение особей, имеющих одинаковые инстинкты. Если у насекомого количество инстинктов и мемстинктов примерно одинаковое, последние не могут сильно скорректировать изначальные цели инстинктов, поэтому поведение различных особей одного вида будет практически идентичным при одинаковых сигналах внешней среды. У человека же на один инстинкт приходятся сотни тысяч мемстинктов, и, соответственно, поведение разных взрослых особей человека при одинаковых сигналах внешней среды будет очень сильно отличаться, поэтому многие исследователи до сих пор считают, что у человека нет инстинктов.

Большое количество мемстинктов и также большое количество их уровней позволяют нам создавать очень сложные мемстинкты, объединяющие большое количество мемстинктов нижнего уровня. Именно это, а также наличие способа передавать искусственные мемстинкты — речи, позволило нам создать такую сложную культуру.

Компилируя существующие мемстинкты, высокоорганизованные обезьяны могут создать новые, например, они могут научиться колоть орехи камнем или научиться пить воду, используя мох в качестве губки, и т. д. Такому приёму также могут научиться другие группы, но из-за того, что общее количество мемстинктов ограничено, будет также ограничено количество таких инноваций, и они с большой вероятностью будут несочетаемы друг с другом. Практически невозможно соединить камень для колки орехов и кусок мха. Один человек может создать десятки подобных инноваций, а коллектив людей — многие десятки, и в таком коллективе могут появиться сочетаемые инновации. Например, если вы умеете делать каменный наконечник и умеете создавать клей, вы можете сделать комплексное орудие. Таким образом, сочетание большого количества мемстинктов и их уровней, а также наличие речи, способной создавать качественные искусственные мемстинкты, ну и, конечно, кисти руки позволило нам создать сложную культуру, и только этими, в основном количественными, признаками мы отличаемся от других животных, обучаемых по второму Б типу и третьему типу.

У животных, имеющих большое количество мемстинктов, особенно у человека, в моменты бодрствования или фазы быстрого сна одновременно бывают активны множество мемстинктов. Таким образом, наше поведение всегда определяется комплексом активных мемстинктов, и поэтому, говоря о целях человека, корректнее будет говорить о комплексных целях, состоящих из всех активных в данный момент мемстинктов.

Выше мы определили, что самое примитивное прогнозирование возможно с помощью ПП, способной сохранять один сигнал, так как этот один сигнал может показывать, что во внешнем мире начался определённый процесс, однако, если мы говорим о планировании, возможности сохранять один сигнал недостаточно. Для того чтобы сформировать полноценный план, интеллектуальная система должна уметь генерировать определённую последовательность манипуляций, то есть как минимум сохранять какие-то последовательности сигналов, а лучше — и последовательности таких последовательностей. Даже самые примитивные интеллектуальные системы, обучающиеся по второму Б типу, могут сформировать примитивный план, например, пчела может иметь план полёта до определённого поля. Однако для создания полноценного и сложного плана также требуется иметь большое количество мемстинктов и их уровней, поэтому на сложное планирование способен только человек. Сложная система

мемстинктов позволяет не только создать план своих действий, но и оценить последствия его реализации. Например, если мы запланируем прикоснуться к оголённому проводу, имея опыт подобного взаимодействия, мы ассоциативно активируем мемстинкт, хранящий подобное воспоминание, а он, в свою очередь, активирует стрессовый мемстинкт, который вызовет реакцию избегания, и это значит, что даже на этапе создания плана мы имеем возможность оценивать план и, значит, корректировать своё поведение. Выше мы говорили, что все цели имеют преемственность и формируются благодаря процессу обучения, таким образом, все цели формируются внешней средой, и только у животных, обучающихся по второму Б типу, с очень развитой системой мемстинктов есть возможность как бы немного заглядывать в будущее, создавая модель развития процессов надсистемы, свой план действия при этом и оценку успешности его реализации. Всё вышеописанное позволяет нам оценивать свои цели и изменять их ещё до реализации плана. Эта очень важная эволюционная особенность в наибольшей мере развита именно у нашего вида, опять же из-за большого количества мемстинктов и их уровней.

Понимая, что человек может строить планы своего поведения, основанные на моделировании процессов внешней среды, мы можем предложить метод оценки сложности модели надсистемы конкретного человека и эффективности его манипулирования внешней средой. Именно на таком принципе может быть построена корректная оценка интеллекта человека. Подобный тест должен содержать описание состояния внешней среды, то есть какой-то ситуации, а человек должен разработать план реагирования на данную ситуацию. По эффективности данного плана можно оценивать сложность модели надсистемы, а значит, интеллект человека. В некоторых случаях данный тест может быть построен в виде игры. При этом следует понимать, что нельзя оценивать общий интеллект человека, записанный в виде одного числа, мы можем оценивать интеллект человека в определённых областях деятельности. Например, один и тот же человек может иметь высокий уровень интеллекта для игры в футбол, но небольшой уровень интеллекта в биологии. Общий уровень интеллекта зависит от количества не ложных мемстинктов, уложенных в корректную и не противоречащую друг другу сеть, от особенностей архитектоники нервной системы конкретного человека, а также от эффективности процессов в ней, то есть как от наследственных факторов, так и от обучения. Например, ген FoxP2, определяющий в том числе длину отростков нейронов в части конечного мозга, отвечающей за слух у млекопитающих и птиц, может влиять на музыкальный интеллект этих животных [54; 55]. Но если мозг не имеет серьёзных отклонений от нормы, уровень интеллекта в большей степени определяется именно обучением.

Когда мы сравниваем интеллект человека с интеллектом другого животного, наше сравнение некорректно, так как мы сравниваем интеллектуальную систему животного, обученную в основном по второму Б типу, и интеллектуальную систему человека, обученную по третьему типу. Если бы человеческий ребёнок самостоятельно собирал информацию и обучался на культуре других животных, то его интеллект, возможно, и превосходил бы в некоторых аспектах интеллект других животных, но тотального, подавляющего превосходства не было бы. Другими словами, если взять человека, обученного в стае обезьян, и сравнить его интеллект с интеллектом других обезьян, а особенно с интеллектом обезьян, обученных в культуре людей, разница была бы не очень большой.

Итак: во внешней среде часто происходят повторяющиеся процессы. Такие процессы могут быть идентифицированы даже по одному измерению — то есть по одному сигналу от определённого рецептора или датчика. По одному сигналу мы можем с определённой точностью прогнозировать последующие события. В природе существует множество систем, способных копировать себя, например с помощью автокатализа, но просто копирование недостаточно для начала процесса обучения. В определённый момент времени появилась РНК, которая самостоятельно или с помощью другой РНК могла не только копироваться, но и воздействовать на окружающую среду по строго заданному алгоритму — создавать другие молекулы, которые,

что очень важно, могут быть использованы для создания более сложной системы. При этом РНК устроена так, что при копировании может измениться алгоритм воздействия на внешнюю среду — будут производиться новые белки. После этого запустился процесс обучения: РНК наделала много своих копий, при этом благодаря мутациям — ошибкам при копировании — разные копии могли иметь разные алгоритмы воздействия на внешнюю среду, а дальше преимущество при копировании получали те репликаторы, которые были приспособлены к происходящим в данный момент во внешней среде процессам. Стали собираться целые библиотеки геномов, приспособленных к определённым процессам внешней среды. Сохраняя копии, приспособленные для определённых условий (имеющие определённые алгоритмы), система при повторении процессов могла отвечать успешными алгоритмами. При этом, поскольку большинство алгоритмов позволяли строить систему из внешних молекул, система постоянно усложнялась, и в итоге интеллектуальная система, которая начала обучаться буквально с нескольких репликаторов, превратилась в настоящий момент в биоту Земли, во многом определяющую её внешний вид и даже строение. Интеллектуальной системой можно считать всю биоту Земли, но также и отдельный биоценоз можно считать интеллектуальной системой, обучающейся по первому типу.

Обучение по второму А типу является очень неэффективным, потому что единичный организм может запомнить очень небольшое количество сигналов. Выход из этого тупика — создать многоклеточный организм, отдельные клетки которого будут сохранять разные сигналы. Так появилась многоклеточность.

Обучение по второму Б типу благодаря уникальной архитектуре позволило сохранять огромное количество последовательностей сигналов из внешней среды, но знания умирали вместе с носителем.

Культура позволяет накапливать эти знания, аккумулируя достижения отдельных индивидуумов за сотни тысяч лет.

И наконец, искусственные интеллектуальные системы помогут в будущем человеку стать бессмертным и преодолеть ограничения естественной интеллектуальной системы.

Глава 2. Общие принципы работы мозга в рамках Общей когнитивной теории как части Теории интеллектуальных систем

Введение

Изучение взаимодействия структур мозга прошло путь от представления о жёстко локализованных «центрах» к пониманию нервной системы как динамической, самоорганизующейся сети. Настоящая монография опирается на теоретическое наследие российских и зарубежных учёных, чьи гипотезы сформировали методологический фундамент современной психофизиологии. Ниже подробно рассмотрены ключевые концепции, объясняющие механизмы межрегиональной координации и интеграции мозговой активности.

И.М. Сеченов заложил рефлекторную парадигму, отвергнув дуалистическое разделение «мозг–душа» и доказав, что психическая деятельность есть результат материальных процессов. Согласно его взглядам, взаимодействие структур мозга носит иерархический характер: кора больших полушарий и средние отделы головного мозга не просто запускают рефлексы, а модулируют их через центральное торможение, подавляя или усиливая сигналы, поступающие от спинного мозга и продолговатого отдела [56]. Таким образом, высшие центры выступают как интегративные «узлы», преобразующие разрозненные афферентные потоки в целенаправленные поведенческие акты.

И.П. Павлов развил учение о высшей нервной деятельности, обосновав материалистический подход к психике через призму физиологии коры больших полушарий [57; 58]. В его модели мозг функционирует как динамическая мозаика очагов возбуждения и торможения, где взаимодействие корковых и подкорковых структур регулируется законами иррадиации, концентрации и взаимной индукции. Ключевым механизмом адаптации выступает условный рефлекс — временная нервная связь, формирующаяся между нейронными популяциями, представляющими условный и безусловный раздражители. Согласно Павлову, образование условного рефлекса подчиняется строгим физиологическим принципам:

1. Временна́я сопряжённость — условный раздражитель должен предшествовать безусловному на оптимальный интервал (обычно 0,5–5 с), что обеспечивает пространственно-временное совпадение нейронных активаций [57, с. 87–92].
2. Многократное подкрепление — повторные совпадения активируют пластические механизмы замыкания временных связей, укрепляя синаптическую передачу между соответствующими корковыми полями [58, с. 134–141].
3. Отсутствие интерферирующих стимулов — сильные посторонние раздражители вызывают ориентировочный рефлекс или запредельное торможение, разрушая формирующуюся связь [57, с. 156–160].
4. Оптимальный корковый тонус — чрезмерное возбуждение или утомление коры блокируют нейропластичность, тогда как умеренная активация ретикулярной формации и подкорковых центров создаёт благоприятный фон для обучения [58, с. 201–208].
5. Генерализация и дифференцировка — на ранних этапах возбуждение иррадиирует, вызывая реакцию на сходные стимулы; по мере тренировки вокруг очага возбуждения формируется концентрическое торможение (отрицательная индукция), обеспечивающее точную избирательность [57, с. 178–185].

В процессе выработки условного рефлекса возбуждение распространяется по коре, затем локализуется в зоне условного стимула, а взаимная индукция формирует очаги торможения вокруг него, что демонстрирует динамическую перестройку межзональных взаимодействий. Подкорковые образования (гипоталамус, лимбические структуры) обеспечивают мотивационно-эмоциональный компонент подкрепления, а ретикулярная формация модулирует общий уровень корковой возбудимости. Таким образом, по Павлову, мозг представляет собой самонастраивающуюся систему, где функциональные связи не фиксированы анатомически, а непрерывно перестраиваются в ответ на опыт, обеспечивая адаптивное поведение через пластичную корково-подкорковую интеграцию [57; 58].

А.А. Ухтомский сформулировал принцип доминанты, согласно которому в каждый момент времени в мозге возникает временный функциональный очаг повышенной возбудимости, подчиняющий себе работу остальных структур [59]. Доминантный центр не привязан к строго определённой анатомической области; он формируется в результате конвергенции афферентных сигналов и поддерживается за счёт отрицательной индукции, подавляющей конкурирующие очаги. Взаимодействие частей мозга здесь носит кооперативно-конкурентный характер: доминанта перенаправляет афферентные потоки, синхронизирует активность ретикулярной формации и лимбических структур, обеспечивая устойчивость целенаправленного поведения даже при изменении внешних условий.

П.К. Анохин создал теорию функциональных систем, описывающую мозг как сеть временных альянсов структур, объединяющихся для достижения конкретного приспособительного результата [60; 61]. Взаимодействие разворачивается в циклическом контуре: афферентный синтез → принятие решения → акцептор результата действия → эфферентный синтез → обратная афферентация. Кора, гиппокамп, миндалевидное тело, ретикулярная формация и мозжечок не действуют изолированно, а образуют ситуативные функциональные блоки, где каждая зона вносит специфический вклад (память, эмоция, тонус, прогноз, моторная коррекция). Акцептор результата служит внутренним сравнительным устройством, сопоставляющим ожидаемый и реальный сенсорный поток, что обеспечивает непрерывную коррекцию межрегиональных связей.

В зарубежной психофизиологии и нейронауке параллельно развивались модели, акцентирующие распределённую организацию и вероятностную природу мозговой активности. Д. Хебб (1949) ввёл понятие клеточных ансамблей и фазовых последовательностей, показав, что нейроны, активируемые одновременно, усиливают синаптические связи по принципу «нейроны, которые разряжаются вместе, связываются вместе» [62]. Взаимодействие зон мозга здесь объясняется через формирование распределённых сетей, охватывающих корковые и подкорковые области. Повторяющаяся коактивация порождает реверберирующие контуры, обеспечивающие кратковременное удержание информации, её консолидацию и последующее автоматизированное воспроизведение. Хеббовский механизм стал нейрофизиологическим базисом для современных моделей связности и пластичности.

М. Газзанига на основе исследований пациентов с комиссуротомией продемонстрировал модульную организацию полушарий и роль межполушарного взаимодействия в формировании целостного сознания [63]. При интактном мозолистом теле полушария обмениваются информацией, компенсируя функциональную асимметрию; при его рассечении каждая гемисфера обрабатывает сигналы автономно, но левое полушарие активно «интерпретирует» разрозненные данные, создавая связный нарратив. Взаимодействие структур, по Газзаниге, носит не только электрофизиологический, но и когнитивно-смысловой характер: мозг постоянно достраивает причинно-следственные связи между разрозненными нейронными процессами, обеспечивая субъективную непрерывность опыта.

М. Райчл с соавторами открыл сеть пассивного режима работы мозга (Default Mode Network, DMN), показав, что в отсутствие внешних задач медиальная префронтальная кора,

задняя поясная извилина, precuneus, угловая извилина и гиппокамп образуют устойчивый функциональный контур [64]. Взаимодействие этих структур антикоррелирует с активностью сетей внешней направленности (dorsal/ventral attention networks), что отражает динамический баланс между экстроспекцией и интроспекцией. DMN служит нейроанатомическим субстратом для консолидации памяти, мысленного моделирования будущего, самореференции и социального познания, демонстрируя, что мозг непрерывно интегрирует прошлый опыт с текущим контекстом даже в состоянии «покоя».

К. Фристон развил теорию предсказательного кодирования и принцип свободной энергии, описывая мозг как иерархическую байесовскую систему [65; 66]. Высшие корковые уровни генерируют прогнозы о сенсорных входах, а нижележащие структуры вычисляют ошибки предсказания и передают их вверх по иерархии. Взаимодействие частей мозга сводится к двустороннему потоку сигналов: нисходящие прогнозы модулируют восприятие, восходящие ошибки обновляют внутренние модели. Нейромодуляторные системы (дофаминергическая, холинергическая, норадренергическая) регулируют «точность» (precision) сигналов, усиливая или ослабляя синаптический вес определённых контуров. Мозг минимизирует неожиданность через активное исследование среды и структурную перестройку связей, что делает взаимодействие регионов не статичным, а непрерывно оптимизируемым.

Среди работ последнего времени, развивающих интегративный подход к пониманию мозга и сознания, необходимо особо отметить концепцию когнитома, разрабатываемую академиком Константином Владимировичем Анохиным. Термин «когнитом» был предложен им для обозначения полной совокупности познавательных способностей мозга — по аналогии с понятием генома [67]. В основе этой концепции лежит гиперсетевая теория мозга, которая пытается преодолеть разрыв между физиологией и субъективным разумом, отходя от традиционных картезианских представлений [68]. Согласно данной теории, сознание и высшие функции не локализованы в отдельных зонах, а возникают благодаря сложной нейронной архитектуре — тому, что сам Анохин называет «теорией реализованных степеней свободы мозга» [68]. Таким образом, концепция когнитома описывает, как именно многоуровневые нейронные ансамбли материализуют мысли, память и субъективный опыт [67].

Эти концепции, от рефлекторных дуг Сеченова и динамической мозаики Павлова до предсказательных иерархий Фристона, сетевой динамики Райчла и когнитома Анохина, отражают эволюцию представлений: мозг перестал рассматриваться как механический конгломерат центров и понимается как многоуровневая, пластичная система, где взаимодействие структур определяется задачами, контекстом и непрерывным потоком обратной связи. Настоящая монография систематизирует эти подходы, сопоставляет их эмпирическую базу и предлагает интегративную модель межрегионального взаимодействия, актуальную для современных психофизиологических и нейрокогнитивных исследований.

Мы видим, что принципы работы мозга, описанные в разных теориях, незначительно отличаются, особенно в деталях. Великие учёные прошлого как бы смотрят на процессы, которые мы подробно будем рассматривать ниже, с разных сторон и видят каждый своё. Всё дело в том, что большинство великих психофизиологов творили в прошлом веке, когда не было полноценных способов экспериментировать с живым работающим мозгом, наблюдая при этом процессы, происходящие внутри нервной системы. Говоря образно, можно представить, что психофизиологи прошлого пытались понять, как устроен слон, ощупывая его в тёмной комнате, поэтому все они описывали различные аспекты работы нервной системы. Без сомнения, поражает научная интуиция отцов психофизиологии, которая позволила им предсказать многие особенности описываемой ниже системы. Именно Общая когнитивная теория как часть Теории интеллектуальных систем позволит нам собрать воедино все эти немного различающиеся описания принципов работы мозга.

2.1. Базовые элементы психики: мемстинкты, инстинкты и воля

Основной принцип работы нашего мозга — одновременное возбуждение популяций нейронов из разных источников. Рассмотрим упрощённую архитектуру нашей психики, начнём с принципов взаимодействия подсистем центральной нервной системы, а далее рассмотрим, как центральная нервная система взаимодействует с другими клетками организма, в частности с эндокринной и иммунной системами. Конечно, на одной схеме нам не удастся показать все контуры взаимодействия отдельных подсистем нашего организма. Чтобы описать все процессы, протекающие в организме, недостаточно будет всего текста монографии и даже целой библиотеки монографий, поэтому мы ограничимся только основными контурами взаимодействия, тем более что задача данной главы — описать общие принципы взаимодействия, и большое количество второстепенных контуров не даст нам сосредоточиться на основных. Мы оставим за рамками нашего рассмотрения гомеостаз — он довольно неплохо описан и не оказывает решающего воздействия на центральную нервную систему, ограничиваясь в основном рефлекторными механизмами в ней. По той же причине мы вообще не будем рассматривать рефлексы, оставим за рамками нашего рассмотрения мозжечок, многие неосновные инстинкты и т. д. и т. п. Только так мы сможем осознать общую логику работы нашей психики и благодаря этому в следующей главе сможем глубже погрузиться в некоторые детали процессов межклеточного взаимодействия.

На рисунке (Рис. 2.1) схематично изображена кора головного мозга, обозначенная цифрой (1). Внутри неё расположены элементы мемстинктов (квадратики), сенсорные справа и двигательные вверху. Кружочки обозначают мемстинкты всех уровней, начиная от первого уровня, расположенного ближе всего к элементам, и заканчивая девятым уровнем, расположенным дальше всего от элементов. Подробно об их клеточном составе и процессах мы будем говорить в следующей главе. Элементы можно частично ассоциировать с колонками неокортекса первичных сенсорных и двигательных зон коры, мемстинкты первого и второго уровня — с колонками вторичных и третичных зон коры соответственно и так далее, заканчивая мемстинктами последнего, девятого уровня, которые частично можно ассоциировать с колонками неокортекса лобных долей коры.

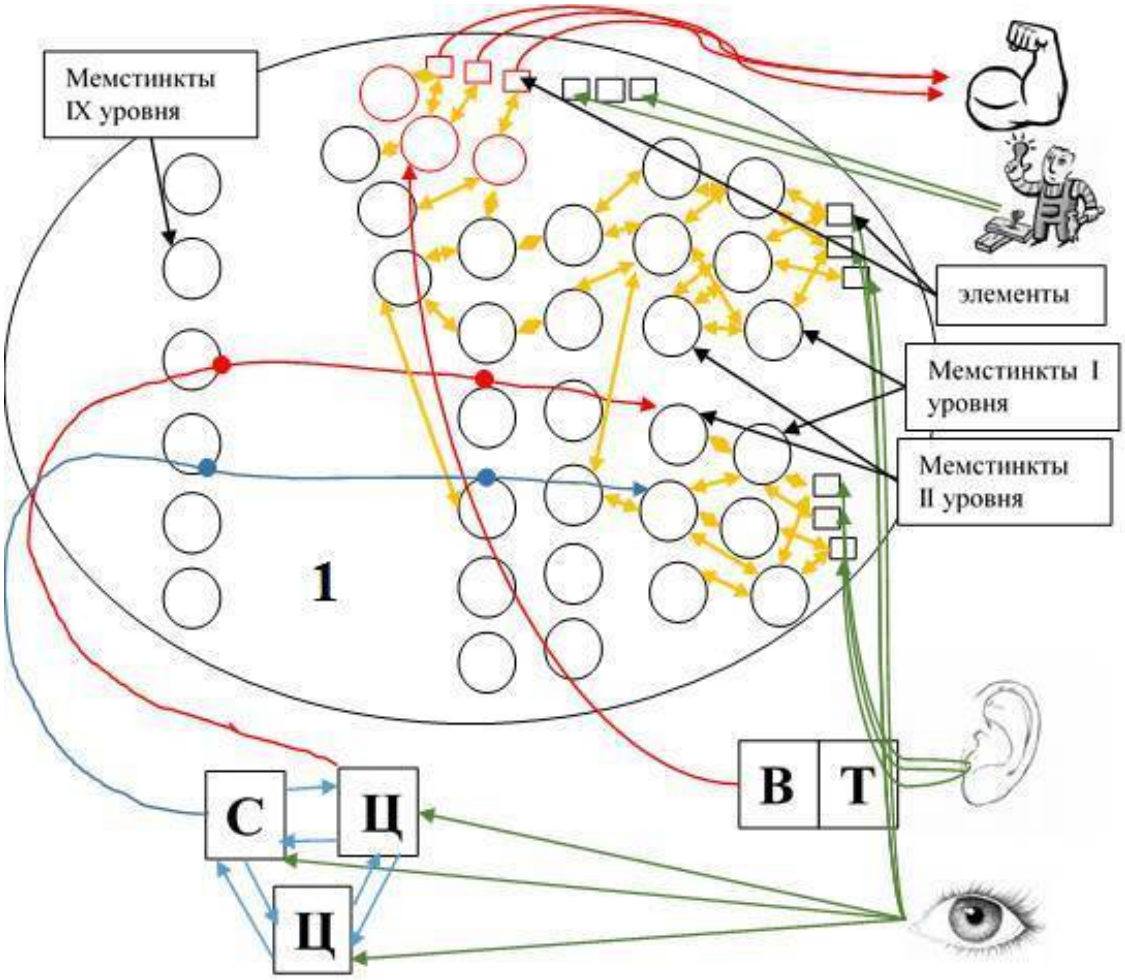


Рисунок 2.1

Кроме того, на рисунке схематично изображены разные типы рецепторов (пиктограмма глаза и уха), а также эффекторы, мышцы и т. д. (пиктограмма руки). Афференты от рецепторов приходят в элементы. Здесь мы, для упрощения схемы, не отображаем преобразование сигналов от рецепторов на пути до коры. Очень важно, что при активации определённых рецепторов всегда активируются одни и те же элементы, эти связи задаются врождённо и не меняются, таким образом, одинаковые сигналы из окружающей среды всегда попадают на одни и те же элементы. Гиперактивация элемента приводит к появлению эпизода сознания. Двигательные элементы, в свою очередь, имеют эфферентные связи с эффекторами, прежде всего с мышцами. При гипервозбуждении двигательного элемента активируется какая-либо мышца, группа мышц либо группа нейронов, запускающая последовательность движений или реакций организма.

Сигналы от рецепторов также попадают на инстинкты, расположенные в подкорковых ядрах. Элементы, в свою очередь, имеют связи с мемстинктами первого уровня и способны их возбуждать, но только при определённых условиях. Поскольку каждый конкретный элемент имеет входы от конкретных рецепторов, повторяющиеся процессы во внешней среде последовательно возбуждают одни и те же элементы или очень похожие последовательности элементов. Для того чтобы мемстинкт сохранил последовательность активации элементов, он должен получить дополнительное возбуждение от инстинкта.

Иерархическая организация системы мемстинктов. Как уже упоминалось в первой главе, система мемстинктов устроена иерархически. Проще всего понять принцип её организации можно, представив себе систему естественного языка. На первых уровнях иерархии находятся фонемы — минимальные смыслоразличительные единицы звучащей речи. На втором уровне последовательности фонем образуют слоги, на третьем последовательности слогов — слова. Далее следуют словосочетания, предложения, абзацы, тексты и так далее вплоть до самых верхних уровней, на которых могут находиться целые библиотеки текстов. Аналогичным образом, мемстинкты первого уровня кодируют последовательности активации отдельных сенсорных элементов (подобно фонемам), мемстинкты второго уровня — последовательности мемстинктов первого уровня (подобно слогам), и так далее до мемстинктов высших уровней префронтальной коры, кодирующих сложные последовательности событий, сопоставимые по сложности с текстами и библиотеками.

Консолидация и узнавание: базовые определения. Для дальнейшего изложения необходимо ввести формальные определения двух ключевых процессов. Консолидация — это усиление связей между нейронами внутри одного мемстинкта, а также его связей с инстинктом, его породившим, и с нижележащими мемстинктами, входящими в его состав. Усиление происходит благодаря одновременному возбуждению указанных структур и поддержке этого возбуждения другими системами, в частности гиппокампом (подробнее — в Главе 3). Консолидация переводит мемстинкт из временного, легко разрушаемого состояния в устойчивую долговременную форму. Узнавание — это активация мемстинкта в результате активации хотя бы части мемстинктов нижних уровней, которые в него входят, либо в результате активации мемстинкта более высокого уровня, в который он сам входит. Иными словами, узнавание возможно как при восходящем движении сигнала (от элементов к высшим уровням), так и при нисходящем (от высших уровней к элементам); в обоих случаях активируется уже существующая, ранее сохранённая структура.

2.2. Инстинкты внимания и формирование мемстинктов первых уровней в онтогенезе

Как мы уже упоминали в первой главе, проще всего понять принцип консолидации мемстинкта на примере импринтинга. У большинства людей импринтинг ассоциируется с наиболее изученной формой — «реакцией следования» зрелорождающихся птенцов серых гусей за родителями и друг за другом, открытой Конрадом Лоренцем [69]. У данных птиц есть инстинкт, который может быть активен только в первые часы после вылупления (сензитивный период). В это время зрительные сигналы, соответствующие предмету определённого размера, активируют данную группу клеток, которая, в свою очередь, дополнительно активирует популяции нейронов мемстинкта, расположенные у птиц в гомологе коры млекопитающих — плаще. В результате птенцы запоминают последовательности сигналов от зрительных рецепторов, совпадающие по времени с появлением в поле зрения предмета определённых габаритов — не очень большого и не очень маленького — и в дальнейшем следуют за ним, воспринимая его как свою мать. Мы видим, что инстинкты не только активируются врождённо заданной последовательностью сигналов от рецепторов, но и характеризуются ещё одной особенностью: они могут быть особенно активными в определённые периоды жизни особи, вплоть до полного включения и выключения возможности активироваться последовательностью сигналов от рецепторов.

Инстинкты человека также могут иметь определённые периоды наибольшей активности. Например, у нас есть инстинкты внимания, активируемые звуковыми сигналами животных. Звуковые сигналы животных часто отличаются от других сигналов наличием формант — пиков мощности на определённых частотах. Форманты возникают благодаря наличию в телах животных пустот; иногда животные даже выращивают специальные полости, предназначенные исключительно для резонирования звуков. Отметим, звуки многих музыкальных инструментов также характеризуются наличием формант. Подобные звуки активируют у человека и некоторых животных инстинкты внимания, обозначенные на нашей схеме буквой «В». Инстинкты внимания — это популяции нейронов, расположенные в подкорковых ядрах головного мозга, в значительной степени в неспецифических и ассоциативных ядрах таламуса. Они иннервируют вторичные и третичные сенсорные и двигательные зоны коры в составе таламической лучистости и активируются сигналами, важными для выживания особи.

По всей вероятности, все инстинкты внимания — или, как минимум, подавляющая их часть — объединены общей эволюционной функцией: они реагируют на последовательности сигналов от рецепторов, совпадающие по времени с появлением поблизости живых существ. В этот класс сигналов входят: форма предмета, напоминающая животное или человека; предмет, похожий на лицо или морду животного (глаза, рот); движение поблизости либо движение, выпадающее из общего паттерна движений (например, предмет, плывущий против волн, или объект, движущийся не так, как все листья на дереве); коммуникационные звуки животных, характеризующиеся наличием формант, и т. д. Такая специализация не случайна: наличие поблизости другого живого существа всегда несёт биологически значимую информацию — оно может представлять либо угрозу (хищник, конкурент), либо потенциальную добычу, либо партнёра. Соответственно, чем быстрее и надёжнее детектируются такие стимулы, тем выше шансы особи на выживание и размножение.

Активация инстинктов внимания, реагирующих на коммуникационные звуки животных, позволяет ребёнку прежде всего сохранять звуки речи. Данный инстинкт особенно активен в первые годы жизни человека, поэтому именно этот период является важным периодом для обучения языку [70]. Продолжительность последовательности сигналов, которую может запи-

сать один мемстинкт, примерно соответствует 250 миллисекундам. По всей вероятности, это универсальная продолжительность одного мемстинкта как минимум для всех млекопитающих [70]. Итак, ребёнок слышит звуки, содержащие форманты, активируются инстинкты внимания, которые, в свою очередь, посылают дополнительное возбуждение во вторичные и третичные зоны слуховой коры, в которых находятся мемстинкты первого и второго уровня, способные записать последовательность активации элементов, активируемых слуховыми рецепторами. Первичная слуховая кора имеет тонотопическую организацию, то есть определённые элементы активируются, когда в поле нашего слуха появляются определённые частоты. Таким образом, ребёнок может запомнить последовательность звуков разной частоты продолжительностью около 250 миллисекунд, совпадающую по времени с появлением во внешней среде коммуникационных сигналов животных. Для современного человека это прежде всего голос его родителей. Ребёнок не будет запоминать шум дождя, скрип двери и т. д., но запомнит отдельные фонемы языка, на котором говорят родители. Если посмотреть на огибающую частот одной фонемы, мы увидим, что каждая фонема состоит из конкретного рисунка мощности звука на определённых частотах, которые как раз и соответствуют последовательности активации отдельных звуковых элементов.

К 9–13 годам формируется большая часть мемстинктов первых уровней, и во взрослом возрасте они могут лишь незначительно дополняться. Это мы можем, например, увидеть при изучении нового языка взрослым человеком. Влияют два фактора: первое — время активности инстинкта внимания (0–5 лет), и второе — заполнение значительной части первых уровней мемстинктов, а также образование связей с высокими уровнями, которые в уже готовой структуре сложно переписывать.

2.3. Мотивационные инстинкты: врождённые детекторы биологической значимости

Итак, мы имеем в значительной мере сформированную сеть мемстинктов первых уровней. Кроме того, у человека есть мотивационные инстинкты — те самые целевые и стрессовые инстинкты, о которых мы упоминали в первой главе, а также инстинкты агрессии, о которых мы будем говорить далее. Напомним: целевые инстинкты — это подкорковые ядра и группы клеток, которые частично можно ассоциировать с дофамином; стрессовые инстинкты — с гистамином; инстинкты агрессии — с норадреналином. Следует подчеркнуть, что на нашей упрощённой схеме система мотивации показана как трёхконтурная, однако в действительности в ней участвуют и другие инстинкты, которые можно ассоциировать с другими нейромедиаторами (серотонин, ацетилхолин, различные нейропептиды). Они оказывают меньшее влияние друг на друга и на основную систему мотивации, но также вносят свой вклад в модуляцию корковой активности и формирование мемстинктов, как мы увидим ниже.

Из первой главы мы помним, что целевые, стрессовые инстинкты и инстинкты агрессии возбуждаются напрямую определёнными последовательностями сигналов от рецепторов. Целевые последовательности сигналов, как мы помним, вызывают у нас чувства удовольствия и связаны с внешними ситуациями и процессами, способствующими передаче генетической информации следующему поколению, например: наличие партнёра противоположного пола, детские черты внешности, наличие еды и т. д. Стрессовые последовательности сигналов связаны с болезнями, разрушениями организма, укачиванием (потому что до появления транспорта изменение ускорения прежде всего говорит о полёте и последующем падении), громкими звуками, которые могут быть связаны с падением предметов и криком опасных животных, и т. д.

Принципиально важным является вопрос о том, как именно эти подкорковые ядра узнают, какую валентность имеет текущая ситуация. Ответ заключается в их прямых или практически прямых входах от сенсорных систем — свойстве, отсутствующем у серотонинергических ядер шва и холинергического базального ядра Мейнерта. Мотивационные инстинкты в ходе эволюционного отбора (обучение прежде всего по первому типу) научились врождённо реагировать на определённые последовательности сигналов от рецепторов: стрессовые реагируют на последовательности сигналов, которые совпадают с процессами во внешнем мире, препятствующими передаче генетической информации, целевые — с процессами, способствующими ей.

Принципиальным является и тот факт, что у удовольствия нет своего отдельного типа рецепторов. Если для стрессовых инстинктов мы ещё можем выделить специализированные рецепторные системы — например, ноцицепторы (болевые рецепторы), хеморецепторы, регистрирующие изменение химического состава крови при воспалении или гипоксии, вестибулярные рецепторы, реагирующие на ускорение, — то удовольствие всегда представляет собой совокупность и определённую последовательность сигналов от разных типов рецепторов. Не существует специальных «рецепторов удовольствия»: один и тот же зрительный, слуховой или тактильный сигнал может вызывать удовольствие в одном контексте и отвращение в другом. Из этого следует фундаментальный вывод: любая теория, пытающаяся объяснить работу мозга без наличия врождённых интегрирующих центров — инстинктов или иных структур, способных оценивать биологическую валентность сложных межмодальных последовательностей сигналов, — обречена на провал, поскольку мозг в такой модели просто не имел бы критерия для различения полезного и вредного.

Наиболее изученным примером прямой активации стрессового инстинкта является передача болевых сигналов. Ноцицептивные афференты от спинного мозга и тройничного нерва через спино-парабрахиальный тракт поступают в латеральную парабрахиальную область (LPB). Нейроны LPB напрямую, моносинаптически возбуждают нейроны туберомамиллярного ядра (TMN) заднего гипоталамуса — главного источника гистамина в мозге [71]. Таким образом, болевой стимул непосредственно, через два синаптических переключения, активирует стрессовый инстинкт. Кроме того, LPB проецируется в центральное ядро миндалины, которое также является компонентом стрессовой системы и связано с TMN реципрокными связями. Резкие громкие звуки, вестибулярные нарушения (укачивание) и другие сигналы физической угрозы используют сходные быстрые пути к стрессовым ядрам через стволовые релейные станции [72].

Целевые инстинкты также имеют прямой сенсорный вход. VTA получает афференты от ядер солитарного тракта (вкусовая и висцеральная информация), а также от обонятельных структур через латеральный гипоталамус. Появление во внешней среде сигналов, коррелирующих с пищей или потенциальным партнёром, напрямую активирует дофаминергические нейроны VTA [73].

Контакты от этих и других инстинктов поступают в кору спереди, по всей вероятности, в составе медиального пучка переднего мозга. При активации мотивационных инстинктов (важная для выживания ситуация) регионы отдельных инстинктов — места, получающие наибольшую иннервацию в коре, — получают дополнительное возбуждение, благодаря чему могут сохраняться последовательности активации мемстинктов низких уровней. Происходит это следующим образом. Представим, что вы прогуливаетесь по городу. У взрослого человека, как мы уже говорили, практически все нижние уровни мемстинктов уже заполнены, то есть практически любая ситуация во внешнем мире активирует у него уже сохранённые мемстинкты нижних уровней. Даже если вы окажетесь совершенно в незнакомом месте, там всё равно будут деревья, дома, возможно, машины либо животные, будет небо, облака и т. д. Поэтому взрослый человек практически всегда находится в состоянии узнавания. Если активируется мемстинкт более высоких уровней, особенно очень хорошо консолидированный, то мы можем испытывать яркое узнавание, например, если вернёмся домой после длительного отсутствия. Но вернёмся к нашей ситуации.

Если вы прогуливаетесь по городу, определённые последовательности сигналов будут активировать мемстинкты нижних уровней, сигналы от них будут поступать выше, прежде всего в префронтальную кору, причём сигналы от определённых групп мемстинктов проходят по определённым трактам в преимущественно определённые зоны коры. Зоны коры, преимущественно связанные с определёнными сигналами снизу, называются архипелагами мемстинктов. В них в будущем будут появляться мемстинкты, содержащие похожие воспоминания, например, группу воспоминаний о ежедневном пробуждении с помощью будильника или о множественных закатах и т. д. Итак, вы прогуливаетесь, определённые архипелаги мемстинктов при этом возбуждаются и благодаря определённым механизмам, например механизму, описанному в исследованиях нейронов задержки, расположенных преимущественно во втором-третьем слое префронтальной коры, удерживаются определённое время (минуты) в активном состоянии [74]. Кроме того, по всей вероятности, создаются временные мемстинкты, способные запомнить пары последовательностей сигналов: сначала вы прошли мимо кондитерского магазина, а затем — рядом с почтой. Это как раз можно сравнить с нахождением в потоке информации, то есть вы помните, по какому маршруту вы шли, но эта информация не сохраняется надолго. Если вас отвлечь, например, случится очень важный звонок, а вы находитесь в незнакомой местности, особенно если вы находитесь не в городе, а в лесу, где все деревья похожи друг на друга, да ещё, например, в сумерках, то после звонка вам потребуется

значительное время, чтобы вспомнить, а где вообще вы находитесь и куда нужно дальше идти. Таким образом работает краткосрочная память.

Но если после этого произойдёт какое-либо важное событие, например, вы увидите возможного будущего партнёра или сбежавшего из зоопарка тигра, входящая последовательность сигналов от рецепторов активирует инстинкт, который в ходе естественного отбора врождённо может узнавать эту последовательность. Регион этого инстинкта в коре получит дополнительное возбуждение, и последовательности сигналов от мемстинктов нижних уровней будут сохранены. Например, вы на всю жизнь запомните, что встретили девушку своей мечты после того, как прошли кондитерскую и почту. Таким образом сохраняется постоянная память (подробнее мы об этом будем говорить в Главе 3).

То, что инстинкты, в том числе мотивационные, врождённо реагируют на определённые последовательности сигналов, находит прямое подтверждение в клинических наблюдениях за детьми с гидранэнцефалией — редким пороком развития, при котором ребёнок рождается практически без коры головного мозга, но с сохранным стволом и подкорковыми структурами. Если бы способность реагировать на сигналы внешней среды была исключительно функцией коры, такие дети должны были бы находиться в вегетативном состоянии. Однако масштабный опрос 108 основных опекунов показал обратное: дети с гидранэнцефалией демонстрируют отзывчивость, эмоциональную реактивность и реагируют на голос [75]. Отдельные сообщения о реакции на собственное имя могут указывать на то, что в некоторых случаях сохраняются остатки коры, позволяющие записывать отдельные сигналы, — однако и без этого дети проявляют эмоциональную реактивность и отзывчивость.

Для нашей теории это означает следующее. Эти данные указывают на то, что подкорковые структуры — в первую очередь верхний отдел ствола мозга — способны врождённо реагировать на биологически значимые последовательности сигналов [76]. При этом важно уточнить: сознание в терминах Теории интеллектуальных систем — это активность сенсорных элементов, и у детей с гидранэнцефалией, у которых кора практически отсутствует, этих элементов может и не быть. Речь идёт не о сохранении сознания в нашем понимании, а о сохранении врождённой способности реагировать на последовательности сигналов — то есть о работе инстинктов.

Человек с полноценной корой отличается от человека без коры не наличием сознания как такового, а прежде всего наличием большого количества мемстинктов, которые также могут реагировать на сигналы внешней среды и обеспечивать сложное, выученное поведение. Инстинкты же — врождённые детекторы последовательностей — работают и без коры. Это наглядно демонстрирует, что базовый механизм реакции на биологически значимые сигналы локализован именно в подкорковых структурах.

2.4. Реципрокное торможение и волевой акт: почему мы делаем то, что приносит удовольствие

На схеме мы видим, что целевые и стрессовые инстинкты соединены между собой стрелочками. Связано это с тем, что мотивационные инстинкты могут напрямую или через посредников тормозить либо возбуждать друг друга. Целевые инстинкты тормозят стрессовые, а стрессовые тормозят целевые. На бытовом уровне это можно проиллюстрировать тем, что тяжело быть счастливым, испытывая боль, и наоборот, счастливый человек испытывает меньше страданий. В этом смысле счастье можно описать как активацию целевых мемстинктов, которой не мешает активация стрессовых.

Данный принцип реципрокного торможения подтверждается конкретными нейроанатомическими данными. Боль подавляет дофамин: нейроны латеральной парабрахиальной области (стрессовый контур) напрямую активируют ГАМК-ергические интернейроны в VTA, что приводит к торможению дофаминергических нейронов и снижению выброса дофамина [77]. Стресс через таламус подавляет дофамин: активация среднего ядра шва (nucleus reuniens, RE) — структуры, рекрутируемой стрессовыми стимулами, — мощно снижает активность дофаминовых нейронов VTA, замыкая ещё один контур стресс-индуцированного торможения целевого инстинкта [78]. Дофамин, в свою очередь, подавляет стресс: введение дофамина и его предшественника L-DOPA непосредственно в префронтальную кору грызунов ускоряло угашение (extinction) страшных воспоминаний, то есть разрушало связь между стимулом и стрессовым мемстинктом [79].

Как же активация целевых и стрессовых мемстинктов влияет на наше поведение? Для этого нам нужно рассмотреть механизмы воли, работа которых во многом связана с базальными ганглиями и близлежащими структурами. Как это работает? Базальные ганглии и близлежащие структуры получают прямые входы от структур коры, причём входы эти выходят в основном из шестого слоя отдельной колонки кортекса, то есть от отдельных мемстинктов. Мемстинкты эти, по всей вероятности, прежде всего дофаминовые, но, вероятно, могут быть и стрессовые, и агрессии, и серотониновые, и т. д. В свою очередь, они по прямому пути (хвостатое ядро → внутренний бледный шар → таламус) через «торможение торможения» воздействуют на таламус, который в результате добавляет возбуждение первичным и вторичным моторным зонам коры. Непрямой путь (хвостатое ядро → внешний бледный шар → субталамическое ядро → внутренний бледный шар → таламус), наоборот, может тормозить эти зоны. Кроме моторных зон, в некоторой степени также могут быть заторможенными или возбуждёнными другие части коры.

Недавнее исследование Komiyama et al. напрямую подтверждает эту модель на уровне отдельных нейронов моторной коры [80]. Учёные обучали мышей нажимать на рычаг, регистрируя активность сотен нейронов с помощью двухфотонной кальциевой визуализации и стимулируя отдельные нейроны с помощью голографической оптогенетики. Оказалось, что кратковременная синхронная активация всего около двадцати нейронов, чья активность начиналась непосредственно перед движением, запускала движение, похожее на выученное. При этом в других, нестимулированных нейронах моторной коры возникал паттерн активности, сходный с тем, который наблюдался при произвольном выполнении выученного движения.

Принципиально важно, что эта стимуляция не была единственной и достаточной причиной движения: её успех зависел от исходного состояния моторной коры и от того, вовлечена ли мышь в задачу. Если животное не было сфокусировано на задаче, стимуляция не вызывала ни движения, ни «выученного» паттерна активности. Это полностью соответствует описанному

выше механизму: волевой акт не «запускает» движение с нуля, а снимает торможение с тех таламокортикальных цепей, которые уже были подготовлены обучением.

Более того, в отдельной работе той же группы было показано, что ключевым источником этого «подготовительного» возбуждения является моторный таламус [81]. В процессе обучения таламус начинает избирательно возбуждать именно те нейроны поверхностного слоя моторной коры, которые связаны с выученным движением, и перестаёт возбуждать «лишние». Инактивация моторного таламуса у обученных животных значительно нарушала выполнение выученного движения.

Таким образом, мы видим, как механизм воли работает непосредственно в моторной коре: активация отдельных мемстинктов (прежде всего целевых, связанных с базальными ганглиями — наше «желание») растормаживает моторный таламус, который, в свою очередь, дополнительно возбуждает моторную кору, но не всю, а только отдельные зоны, синапсы которых были усилены благодаря обучению. Целенаправленная стимуляция этих ансамблей и запускает заученные движения.

Что это даёт? Сигнал в системе мемстинктов может как подниматься вверх, так и спускаться вниз, при этом мемстинкт более высокого уровня может активировать мемстинкты более низкого уровня, которые в него входят. Таким образом, возбуждение от верхних уровней позволяет произойти узнаванию мемстинктов более низких уровней вплоть до самих элементов — точно так же, как при восходящем движении сигнала происходит узнавание мемстинктов высоких уровней. При этом активация одного сенсорного элемента — это и есть эпизод сознания. Однако элементы практически никогда не возбуждаются по отдельности: они возбуждаются либо рецепторами, и тогда возбуждается какая-то последовательность элементов, соответствующая процессу, происходящему во внешней среде, либо активируются благодаря мемстинкту высокого уровня, который хранит воспоминания о последовательности событий. Поэтому элементы практически всегда возбуждаются последовательно в виде историй сознания; истории практически всегда логически правильно уложены.

Однако возбуждение сверху может также дотягиваться до двигательных элементов; в этом случае должно происходить узнавание двигательных мемстинктов низких уровней и, потенциально, активация последовательности движений. Но система движения устроена так, что нисходящего возбуждения самого по себе недостаточно для полноценного узнавания двигательных мемстинктов низкого уровня и двигательных элементов. В результате, если активируется мемстинкт высокого уровня, содержащий воспоминания о движении, мы можем, например, увидеть это движение в нашем мысленном взоре, потому что возбуждение дойдёт до сенсорных элементов и вызовет их узнавание, но до двигательных элементов оно не дойдёт. Такое ограничение не позволяет нам совершать хаотические движения, только подумав о них, — например, мы не будем шагать в пропасть, даже если подумаем об этом. Исследования внутренней речи с использованием современных методов нейровизуализации подтверждают, что при мысленном проговаривании (нисходящее узнавание сенсорных элементов) и реальной речи (активация двигательных элементов) активны одни и те же группы нейронов — одни и те же двигательные мемстинкты низких уровней, а разница заключается именно в том, что во втором случае дополнительное возбуждение от механизмов воли запускает двигательную реакцию, произнесение звуков вслух [82].

Чтобы совершить осознанное движение, необходимо активировать механизм воли. Дополнительное возбуждение, которое поступает в моторную и близлежащие зоны, позволяет произойти полноценному узнаванию двигательных мемстинктов низких уровней, благодаря чему активируются последовательности двигательных элементов, и сигнал вниз через разные механизмы достигает мышц. Ключевым звеном здесь является прямой путь базальных ганглиев (D1-рецепторы стриатума → внутренний бледный шар → таламус), который снимает тоническое торможение с таламокортикальных моторных цепей [83].

Именно целевые, дофаминовые мемстинкты имеют преимущественный доступ к этому воротному механизму. Это означает, что мы в самом прямом нейрофизиологическом смысле делаем то, что приносит нам удовольствие (способствует передаче генетической информации), и не делаем то, что вызывает стресс (препятствует ей). Когда активирован целевой инстинкт, дофамин облегчает проведение сигналов через D1-стриарный путь, и волевой акт совершается легко — мы чувствуем мотивацию и готовность действовать. Когда же активирован стрессовый инстинкт, он тормозит VTA, снижая дофаминовую модуляцию. Целевые мемстинкты, даже будучи активными на корковом уровне, не могут «открыть ворота» базальных ганглиев. Движение блокируется на подкорковом уровне.

Таким образом, мы видим, что множество двигательных мемстинктов являются целевыми, то есть мы делаем то, что приносит нам удовольствие, в прямом смысле этого слова. С другой стороны, стрессовые инстинкты тормозят целевые, в результате чего вероятность движения сильно снижается. Получается, что мы не делаем того, что приносит нам стресс, то есть может препятствовать передаче генетической информации. Причём остановка такого движения происходит сразу же после активации стрессового мемстинкта, который реципрокно возбуждает стрессовый инстинкт, а тот, в свою очередь, автоматически тормозит целевой. Подобный механизм позволяет нам отбросить опасные действия на самых ранних этапах планирования, ещё до составления каких-либо подробных планов.

Более того, из реципрокного характера торможения следует, что мы не только делаем то, что активирует целевые инстинкты, но и то, что снижает активацию стрессовых. Поскольку стрессовые инстинкты тонически тормозят целевые, любое ослабление стрессовой активации автоматически приведёт к растормаживанию целевого контура и, как следствие, к субъективному переживанию удовольствия. Таким образом, поведение направляется двумя комплементарными векторами: стремлением к стимулам, активирующим целевые инстинкты, и стремлением к действиям, выключающим или уменьшающим активность стрессовых. При планировании мозг автоматически отдаёт предпочтение тем маршрутам, которые не только обещают награду, но и позволяют избежать или минимизировать стрессовую активацию.

2.5. Инстинкты агрессии: выход из двигательного паралича

Однако у описанной выше системы реципрокного торможения есть фундаментальная уязвимость. Если стрессовый стимул достаточно силён и продолжителен, он может полностью подавить активность целевого инстинкта, и организм окажется в состоянии поведенческого паралича — неспособным инициировать какое-либо действие. Велика вероятность возникновения ситуации, когда мышь забьётся в угол и не будет ничего делать при сильном стрессе. Это не всегда эффективно; иногда в стрессовых ситуациях, например при нападении хищника, требуется реакция типа «бей или беги».

Чтобы не возникало такого залипания, природа создала ещё один тип инстинктов — инстинкты агрессии, на схеме они обозначены буквой «А». Их частично можно ассоциировать с норадреналином. Мемстинкты из норадреналиновых регионов коры позволяют совершать движение в таких случаях. Логика взаимодействия инстинктов агрессии с двумя другими контурами прямо противоположна логике стрессового инстинкта:

· Стресс активирует агрессию. Те же стрессовые сигналы, которые через парабрахиальную область тормозят VTA, через ретикулярную формацию активируют голубое пятно [84].

· Агрессия тормозит стресс. Активированное голубое пятно через $\alpha 2$-ауторецепторы и проекции к гипоталамусу ограничивает дальнейший выброс гистамина и кортиколиберина, снижая интенсивность стрессовой реакции.

· Агрессия возбуждает целевые инстинкты. Норадреналин через β-рецепторы в VTA и прилежащем ядре потенцирует выброс дофамина, временно растормаживая целевой контур [85].

Субъективно этот каскад переживается как переход от страха и оцепенения к ярости и готовности к активному сопротивлению или бегству. Это и есть классическая реакция «бей или беги». Как мы видим из схемы, инстинкты агрессии возбуждаются стрессовыми, но сами тормозят стрессовые; с целевыми они находятся во взаимном возбуждении. Другие типы инстинктов также могут взаимно возбуждать или тормозить друг друга, благодаря чему реализуется ещё один вариант сложности поведения животных.

Следует отметить, что стрессовые инстинкты и инстинкты агрессии могут совместно активироваться благодаря частичному совпадению последовательностей сигналов от рецепторов, которые их запускают. Это совпадение не является полным. Например, укачивание или наличие в крови маркеров воспаления и инородных агентов активирует стрессовый инстинкт, но практически не затрагивает инстинкт агрессии. Однако боль способна активировать оба контура одновременно — именно поэтому болевой стимул в зависимости от интенсивности и контекста может вызывать как реакцию избегания и замирания, так и яростное сопротивление.

Аналогичным образом могут частично совпадать активирующие последовательности для целевых инстинктов и инстинктов агрессии. В данном контексте необходимо подчеркнуть, что термин «агрессия» используется здесь в сугубо физиологическом, а не социально-оценочном смысле. Речь идёт об активации норадренергической системы, обеспечивающей мобилизацию организма, повышение мышечного тонуса, сужение фокуса внимания и готовность к энергичным действиям. Многие активные и экстремальные виды спорта, соревновательная деятельность, охота и тому подобное как раз и представляют собой ситуации, в которых одновременно активируются и целевые инстинкты (удовольствие от процесса и результата), и инстинкты агрессии (мобилизация, преодоление сопротивления, концентрация усилий). Именно это сочетание даёт характерное переживание «драйва» и азарта.

2.6. Планирование как внутреннее проигрывание: отбрасывание опасных вариантов

Трёхконтурная система мотивации взаимодействует не только с моторным выходом, но и с самим процессом мышления. Поскольку сигнал в системе мемстинктов может двигаться как вверх, так и вниз, активация мемстинкта высокого уровня способна последовательно активировать — то есть вызывать узнавание — входящих в него мемстинктов более низких уровней, вплоть до сенсорных элементов. Этот процесс нисходящего узнавания и составляет нейрофизиологическую основу планирования — мысленного проигрывания последовательности событий до их реального совершения.

И здесь описанная выше система реципрокного торможения инстинктов проявляет себя в полной мере. Представьте, что вы обдумываете маршрут до магазина. Один из вариантов предполагает пройти мимо дома соседа, с которым вчера произошла ссора. Как только этот фрагмент плана активируется в префронтальной коре, он возбуждает соответствующий стрессовый мемстинкт. Стрессовый мемстинкт имеет двусторонние связи со «своим» подкорковым инстинктом (как будет подробно разобрано в Главе 3). Он реципрокно активирует TMN, что приводит к торможению VTA и подавлению дофаминовой поддержки всего текущего плана.

План, содержащий опасный элемент, разрушается на самых ранних стадиях мысленного проигрывания. Он просто перестаёт «казаться привлекательным», и мышление спонтанно переключается на другой, обходной маршрут. Замечательно, что этот процесс происходит автоматически и не требует осознанного анализа опасности. Более того, мы склонны немедленно изобретать рациональные обоснования для уже совершённого подсознанием выбора: «Я пойду дальней дорогой, потому что заодно посмотрю на строящееся здание».

Этот механизм принципиально отличен от того, как обучаются современные системы искусственного интеллекта на основе обучения с подкреплением (Reinforcement Learning, RL). В классической RL-парадигме отрицательная награда (negative reward) просто снижает вероятность повторения всей цепочки действий, которая к ней привела. Отрицательный опыт, по сути, стирается, а положительный — закрепляется. Такая система просто не даёт сохраниться негативному опыту и позволяет сохранять только правильные паттерны поведения. В биологическом мозге, напротив, и негативный, и позитивный опыт сохраняются в виде мемстинктов. Регуляция же осуществляется не на уровне запоминания/стирания, а на уровне динамического торможения одного мотивационного контура другим.

Это даёт колоссальное преимущество при изменении ситуации: связь, например, стрессового мемстинкта с породившим его стрессовым инстинктом может быть ослаблена, и сохранённая последовательность действий уже не будет казаться такой страшной и неприятной. Представим себе ситуацию: вы живёте на левом берегу реки, на правом водятся опасные тигры, и вы, несмотря на то что когда-то нашли способ перебраться туда, совершенно не хотите этого делать. Однако на правом берегу находятся неисчерпаемые запасы еды — например, ягод и орехов, — и если тигры покинули это место, то заполненный ранее вами путь, который может привести на правый берег, может стать очень выгодным для вашего выживания. В этом случае наш мозг может ослабить связь этого воспоминания со стрессовым инстинктом, и вы перестанете бояться переходить на соседний берег. Как можно использовать этот механизм, мы будем говорить ниже (раздел 7.6, Самоформирование). Искусственный интеллект, обучающийся по классической парадигме reinforcement learning, так не умеет — ведь он в принципе не сохранил бы опасный путь, и при изменении обстановки ему потребовалось бы полное переобучение с нуля. Современные гибридные подходы, включающие модели мира (model-based RL), частично смягчают это ограничение, но принципиальное различие сохраняется: биологиче-

ский мозг хранит и позитивный, и негативный опыт одновременно и регулирует поведение через динамическое торможение мотивационных контуров, а не через переобучение.

Ещё одно фундаментальное различие между биологическим мозгом и современными системами искусственного интеллекта касается способности к упреждающей фильтрации опасных планов. Ребёнок, однажды засунувший шпильку в розетку и получивший удар током, запомнит не только саму розетку как источник боли, но и сформирует стрессовый мемстинкт, активируемый видом любых оголённых проводов, искр, специфических предупреждающих знаков и т. д. В следующий раз, обдумывая какой-либо план действий, он ещё на стадии мысленного проигрывания столкнётся с узнаванием элемента «оголённый провод», что немедленно активирует стрессовый мемстинкт, затормозит целевой инстинкт и отбросит опасный вариант. Система, не сохраняющая негативный опыт, не сможет оценить угрозу, содержащуюся в новой, ранее не встречавшейся комбинации знакомых элементов. Например, если взрослый человек в процессе планирования ремонта скомбинирует несколько известных ему операций, но в получившейся последовательности будет присутствовать «сборка проводов под напряжением голыми руками», этот план будет отброшен на стадии мысленного проигрывания, даже если сам человек никогда прежде не делал именно этой ошибки.

Именно отсутствие подобного механизма является одной из фундаментальных причин появления так называемых галлюцинаций у больших языковых моделей (Large Language Models, LLM). Многие люди в детстве бывали наказаны или испытывали стыд за ложь, заведомо неправильный ответ или грубую ошибку, а значит, имеют стрессовые мемстинкты, связанные с порождением недостоверной информации. При попытке сказать заведомую нелепость эти мемстинкты активируются и тормозят речевой акт ещё до его совершения. У LLM такой встроенной системы валидации на основе эмоционального опыта нет, поэтому проблему галлюцинаций приходится решать внешними, инженерными инструментами — reinforcement learning from human feedback (RLHF), фильтрацией обучающих данных, формальной верификацией вывода и т. д.

Как мы видим, описанная схема с сохранением и позитивного, и негативного опыта и их динамическим взаимодействием на уровне инстинктов работает намного эффективнее и надёжнее применяемых ныне в искусственном интеллекте архитектур.

2.7. Нервная система как часть целостного организма: аллостаз, эндокринная и иммунная регуляция

Далее мы будем говорить о том, что наша нервная система не находится в изоляции от остального организма. Мы обсудили поведение, которым управляет нервная система; теперь обратимся к другим аспектам регуляции, в частности к аллостазу — процессу поддержания стабильности организма через изменение его внутренних параметров в ответ на внешние и внутренние вызовы, согласно Булгаковой О. С. [86]. В регуляции аллостаза участвуют две другие важнейшие системы: иммунная и эндокринная.

Прежде чем рассматривать их взаимодействие с нервной системой, взглянем на организм целиком. Многие современные исследователи сравнивают наш организм с лесом, в котором обитает множество других живых существ. На первый взгляд мы видим лес с большими деревьями, но, присмотревшись, замечаем, что в них и вокруг них живёт огромное количество мелких организмов — животных, растений, микроорганизмов. Хотя масса клеток самого организма (деревьев) и превосходит массу биома, обитающего в нём и на нём, по количеству клеток микроорганизмы значительно превышают число клеток человека. Безусловно, микробиом взаимодействует с организмом хозяина, и тому есть множество примеров [87]. Однако, описывая общую упрощённую схему регуляции человеческого организма, мы оставим данные типы взаимодействия (в основном химические) за рамками нашего рассмотрения. Кроме того, все клетки нашего организма взаимодействуют друг с другом физически и посредством химических сигналов, но мы также не будем детально рассматривать эти механизмы, хотя обозначим их наличие на схеме (см. Рисунок 2.2).

Конец ознакомительного фрагмента.

Текст предоставлен ООО «Литрес».

Прочитайте эту книгу целиком, купив полную легальную версию на Литрес.

Безопасно оплатить книгу можно банковской картой Visa, MasterCard, Maestro, со счета мобильного телефона, с платежного терминала, в салоне МТС или Связной, через PayPal, WebMoney, Яндекс.Деньги, QIWI Кошелек, бонусными картами или другим удобным Вам способом.