Никитин А.В.

Автокибернетика от робота к киберу – 3.

Архитектура наблюдающего управления: эволюционная симметрия

Аннотация

В третьей части работы мы соединяем эволюцию генетического кода с архитектурой наблюдающего управления. Показано, что триплетный код, симметрия относительно центрального основания и исторический сдвиг рамки считывания предвосхищают тройственную структуру самоуправления: агрегирующая машина – «Мы», логическая машина – «Сознание», адаптирующая машина – «Я» [Никитин, 2018]. Количество оснований в триплете (3) соответствует ограниченному числу независимых конфигураций виртуальных машин в реальном времени (принцип минимальной достаточной сложности). Аналогия между генетическим кодом и иерархией управления позволяет по-новому взглянуть на субъектность и проектирование киберов.

Оглавление

ВВЕДЕНИЕ: ОТ ГЕНЕТИЧЕСКОГО КОДА К УПРАВЛЕНИЮ 2

ГЛАВА 1. ТРИ МАШИНЫ УПРАВЛЕНИЯ КАК ТРИ ОСНОВАНИЯ КОДОНА 3

1.1. Машина саморегулирования (уровень 0) – наблюдаемое управление 5
1.2. Агрегирующая машина – левое основание, «Мы» 5
1.3. Логическая машина – центральное основание, «Сознание» 5
1.4. Адаптирующая машина – правое основание, «Я» 5
1.5. Эволюционная рекапитуляция: от одноклеточных к многоклеточным 6
1.6. Сводная таблица соответствия 6

ГЛАВА 2. ВНУТРЕННЯЯ МОДЕЛЬ МИРА (Θ) И ЕЁ ЭВОЛЮЦИЯ 8

ГЛАВА 3. ПОТОК СЛУЧАЙНОСТЕЙ КАК ЗАШУМЛЕНИЕ МОДЕЛИ 9

ГЛАВА 4. СМЕНА МОДЕЛЕЙ И ЕЁ СВЯЗЬ С ГЕМИЗОНОМ (GH) И ВИРТУАЛЬНЫМИ МАШИНАМИ 10

ГЛАВА 5. МЕТА-МОДЕЛИ И РЕКУРСИВНАЯ РЕФЕРЕНЦИЯ — ОСНОВА ВИРТУАЛЬНОЙ МАШИНЫ «СОЗНАНИЕ» 12

ГЛАВА 6. ДЕЙСТВИЕ «НАОЩУПЬ»: МАТЕМАТИЧЕСКАЯ ФОРМАЛИЗАЦИЯ 13

ГЛАВА 7. ИЗМЕНЕНИЕ УСТАНОВОЧНЫХ ТОЧЕК И ВИРТУАЛЬНЫХ МАШИН 13

ГЛАВА 8. ЧИСЛО ДРУЖЕСТВЕННОСТИ ДЛЯ МЕТА-УРОВНЯ (FR_2^{META}) 15

ГЛАВА 9. ПРИМЕР: АМЁБА ПРОТИВ РОБОТА 15

ГЛАВА 10. АРХИТЕКТУРНЫЕ ОСОБЕННОСТИ УРОВНЕЙ САМОУПРАВЛЕНИЯ (С УЧЁТОМ ВИРТУАЛЬНЫХ МАШИН) 17

ГЛАВА 11. ЧТО ЭТО ДАЁТ ДЛЯ ПРОЕКТИРОВАНИЯ АВТОНОМНЫХ СИСТЕМ 19

ГЛАВА 12. АРХИТЕКТУРА САМОУПРАВЛЕНИЯ: ОТ МОЗГА К ИСКУССТВЕННОМУ СУБЪЕКТУ 20

12.1. Архитектурная пропасть: мозг против классического компьютера 20
12.2. Таламус как реальный Субъект в автокибернетике 21
12.3. Конкуренция виртуальных субъектов и Парадигма потока случайностей 22

12.4. Невозможность «перекачки личности» (аплоадинга) 22
12.5. От компьютера к искусственному мозгу 23
РЕЗЮМЕ ЧАСТИ 3 ..... **24**
ПРИЛОЖЕНИЕ **25**
Предсказания теории и возможные эксперименты 25
БИБЛИОГРАФИЯ К ЧАСТИ 3 ..... **26**

Введение: от генетического кода к управлению

В первой части мы ввели поток случайностей, число дружественности, зону самоуправления и показали, что подлинный выбор возможен только при $0{,}9 \le Fr_2 \le 1{,}1$ [Никитин, 2026а]. Во второй части мы построили иерархию уровней (0–7) и выделили три виртуальные ветви – «Мы», «Я», «Сознание», которые развиваются независимо [Никитин, 2024а]. Теперь мы подходим к самому интригующему вопросу: почему этих машин три? И почему их эволюционное взаимодействие похоже на развитие генетического кода?

В 2021 году я опубликовал работу, где показал, что эволюция кодирования аминокислот в иРНК шла не сразу триплетным способом. Сначала рибосома сдвигалась всего на одно основание, и тогда только каждый четвёртый триплет был новым. Это дало 20 независимых триплетов – ровно столько, сколько сегодня используется аминокислот (плюс стоп-кодоны). Позже система перешла на триплетный сдвиг, но количество аминокислот уже не увеличилось – сохранилось ограничение прошлого этапа. Каждая аминокислота стала кодироваться несколькими триплетами (вырожденность кода) [Никитин, 2018]. Такой механизм напрямую подтверждается исследованиями древнего фреймшифтинга у грибов, сохранявшегося 150 миллионов лет [Farabaugh et al., 2006], а также общей оптимизацией генетического кода для минимизации ошибок сдвига рамки [Freeland et al., 1999].

Главное открытие той работы: в основе триплетного кода лежит симметрия относительно среднего основания. Центральный нуклеотид определяет принадлежность кодона к определённой аминокислоте, а левый и правый – вариативны. Например, аргинин кодируется кодонами AGA, AGG, CGA, CGC, CGG, CGU – везде в центре G. Серин – имеет два центра (G и C), что уникально. У пятнадцати аминокислот есть хотя бы один симметричный кодон (например, UAU, GCG, AUA). Несимметричные кодоны возникли позже, когда симметричные варианты были исчерпаны [Никитин, 2018][Никитин, 2021]; это согласуется с общим принципом возникновения симметричных паттернов в эволюции кода [New Scientist, 1994].

Глава 1. Три машины управления как три основания кодона

Эта тройственная структура – левое основание – центральное – правое основание – и есть прообраз трёх фундаментальных машин управления в автокибернетике.

Агрегирующая машина – собирает сигналы «слева» (от внешней среды, от других агентов). Это основа «Мы» (коллективная координация).
Логическая машина – центральный процессор, который обрабатывает информацию и строит модели мира. Это основа «Сознания» (внутренний экран).
Адаптирующая машина – отвечает за изменение установочных точек и поведенческих стратегий «справа» (относительно центра). Это основа «Я» (самореференция, рефлексия).

Точно так же, как в генетическом коде центральное основание определяет аминокислоту, в управлении логическая машина является ядром, вокруг которого строятся две другие.

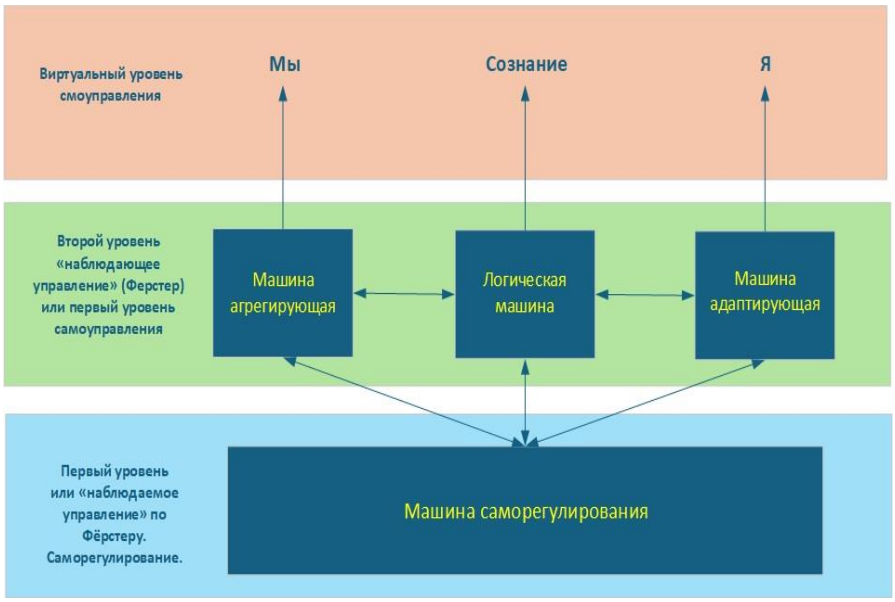

Рис.1. Система самоуправления.

Симметрия между левым и правым краем триплета – это симметрия между коллективным и индивидуальным, между «Мы» и «Я». Там, где она высока (как в кодонах AGA или CGC), виртуальные машины хорошо интегрированы (I высока). Там, где симметрии нет, они могут работать независимо (I низка). В клетке эта полярность закреплена, например, сигнальным путём Wnt, который нарушает симметрию и создаёт передне-заднюю и лево-правую оси [Holzem et al., 2024].

Эта глубокая, эволюционно закреплённая полярность между «Мы» и «Я» находит прямое отражение в функциональной асимметрии больших полушарий человеческого мозга. Данные нейровизуализации и нейропсихологии убедительно показывают, что правое полушарие является материальным субстратом для виртуальной машины «Я», в то время как левое полушарие служит основой для виртуальной машины «Мы».

Правое полушарие — «Я» (Самость, Агентность, Телесное «Я»)

Обширные исследования связывают правое полушарие с функциями, составляющими ядро нашего «Я». Так, распознавание собственного лица — один

из ключевых маркеров самосознания — происходит преимущественно в правом полушарии [Lyons et al., 2023]. Более того, способность к различению «себя» и «другого», чувство принадлежности себе собственного тела (self-location и first-person perspective) также латерализованы в правом полушарии. В частности, правое височно-теменное соединение (rTPJ) играет критическую роль в телесном самосознании и чувстве агентности, действуя как ключевой узел сети, поддерживающей ощущение «Я» [Ionta et al., 2014; Smit et al., 2016; Smit et al., 2020]. Учёные также приходят к выводу, что сеть правого полушария преимущественно контролирует визуальное самосознание и чувство принадлежности тела (sense of body ownership) [Sperber & Karnath, 2018]. Это полностью соответствует вашей концепции «виртуальной машины «Я»», отвечающей за самореференцию, чувство агентности и телесное самосознание.

Левое полушарие — «Мы» (Социальная координация, Речь, Коллектив)

Функции левого полушария, в свою очередь, обеспечивают возможности для коллективной координации и формирования общего «Мы». Левое полушарие является доминантным для языка, без которого невозможна культурная трансмиссия на уровне сложных абстрактных значений. Более того, его часто называют «интерпретатором», стремящимся создавать связные, логичные объяснения для действий и событий. Эта способность выстраивать нарративы и интерпретировать поведение является необходимой основой для социального взаимодействия и разделяемых смыслов, создающих общность «Мы» [Savva, 2017; de Bode, 2018; Pinto et al., 2015].

Интеграция «Я» и «Мы»: роль мозолистого тела и сети пассивного режима (Default Mode Network, DMN)

Однако, как и в генетическом коде, «симметрия» между полушариями не менее важна, чем их различие. Высокая степень интеграции между «Я» и «Мы» требует интенсивного межполушарного обмена информацией, который обеспечивается структурами, подобными мозолистому телу (corpus callosum). Его целостность критична для эффективной координации активности полушарий и, следовательно, для высокой функциональной интеграции (I) [Petrof et al., 2025]. Помимо этого, идеальным нейробиологическим коррелятом «интегрированного сознания» является сеть пассивного режима (DMN). Эта сеть активна, когда мы не заняты внешней задачей, и, как показывают исследования, именно в ней происходит интеграция информации о себе и о других, создавая единый субъективный опыт [Gruskin et al., 2020; Nenert et al., 2020; Pandey et al., 2020]. Таким образом, таламус и DMN выступают в роли высокоуровневых интеграторов, обеспечивающих работу виртуальной машины «Сознание».

В этой части мы развернём эту аналогию в полноценную архитектуру наблюдающего управления. Мы покажем, как тройственность машин возникает из фундаментальных ограничений потока случайностей, подобно тому, как количество аминокислот ограничено древним сдвигом рамки. И мы дадим инженерные принципы проектирования киберов, учитывающие эту тройственность.

1.1. Машина саморегулирования (уровень 0) – наблюдаемое управление

Машина саморегулирования – это самый нижний уровень, аналогичный «нулевому сдвигу рамки», когда ещё нет распознавания триплетов. Она просто поддерживает гомеостаз через отрицательные обратные связи. Археи, губки, гидры, полипы – их поведение не требует различения трёх машин. Формальное описание даётся уравнением баланса хаоса [Никитин, 2026b]:

$$dH/dt = J_{in} - J_{out} - \xi + J_{birth} \quad (1)$$

где $J_{birth} = 0$ – нет рождения новых структур. Это «докодонный» этап эволюции управления.

1.2. Агрегирующая машина – левое основание, «Мы»

Агрегирующая машина соответствует левому основанию триплета. Она собирает сигналы из внешней среды и от других агентов. У прокариот это кворум-сенсинг и биоплёнки [Sorek et al., 2013]. У червей – примитивные нервные ганглии, интегрирующие сенсорные сигналы. Эта машина – основа коллективной координации «Мы». Когда она активна, система «смотрит налево» – на сородичей, на среду, на коллективные паттерны. В триплетном коде левое основание может быть любым (A, U, G, C). Аналогично в управлении агрегирующая машина может работать с разными типами информации – визуальной, акустической, химической. Но её роль всегда одна – обеспечить целостную картину «нас».

1.3. Логическая машина – центральное основание, «Сознание»

Логическая машина – это центральное основание триплета. Она определяет «смысл» – то, какая именно аминокислота будет встроена. В управлении она отвечает за построение внутренних моделей мира, прогнозирование, рекурсивный вывод. Это ядро субъектности. Без неё нет «Сознания» – внутреннего экрана, на котором разворачиваются образы. Центральное основание – самое консервативное; в эволюции именно оно задаёт «смысл» кодона [Dila et al., 2020]. В самоуправлении логическая машина – самое «дорогое» и сложное устройство. Она требует вычислительных ресурсов, памяти и энергозатрат. Именно поэтому у беспозвоночных и рыб сознание отсутствует – их логическая машина ещё недостаточно мощна, чтобы породить внутренний экран [Feinberg & Mallatt, 2013].

1.4. Адаптирующая машина – правое основание, «Я»

Адаптирующая машина соответствует правому основанию триплета. Она отвечает за изменение поведения, переключение стратегий, самосохранение и самообновление. Это основа «Я» – чувства агентности, самореференции, рефлексии. Когда адаптирующая машина активна, система «смотрит направо» – на собственные действия, на ошибки, на необходимость скорректировать курс. В генетическом коде правое основание часто бывает «вырожденным» – несколько разных нуклеотидов в третьей позиции могут кодировать одну и ту же аминокислоту. Это не строгая симметрия, а разрешённая вариативность. Точно так же адаптирующая машина может иметь множество способов действия (стратегий), которые ведут к одной цели – сохранению целостности «Я». Уже у бактерий развиты механизмы хемотаксиса и различения «свой–чужой» (CRISPR-Cas, рестрикционные системы), что можно считать зачатками адаптирующей машины [Wadhams & Armitage, 2004]; [Marraffini & Sontheimer, 2010].

1.5. Эволюционная рекапитуляция: от одноклеточных к многоклеточным

Термин «рекапитуляция» означает повторение этапов развития: многоклеточные организмы в эволюции своего самоуправления проходят те же стадии, что и одноклеточные, но на новом — нейронном уровне.

У одноклеточных (прокариоты, бактерии) уже есть зачатки всех трёх машин, работающих на биохимическом уровне. Это развитие достигает своего максимального уровня у хищников эукариот, когда $J_{birth} > 0$. Для одноклеточных это предел развития.

Многоклеточные начинают развитие сначала. Теперь эти же машины реализуются на нейронной сети. И не с первой попытки. У растений развитие самоуправления останавливается на уровне 1 (модулярное самоуправление). Насекомые проходят весь путь одноклеточных и снова достигают уровня, когда J_{birth} становится больше нуля – появляется способность создавать новые модели. На том же пределе уровня 3. До этого уровня эволюционируют коллективные формы управления «Мы», достигая пика у эусоциальных насекомых, где особи имеют слабое «Я», но мощное интегративное «Мы» [Rehan & Glastad, 2024].

Там же остановились членистоногие.

Дальше пошло только развитие хордовых (рыбы, животные). Рыбы и беспозвоночные вышли на уровень 3. Но теперь уровень 3 – это только трамплин для дальнейшего развития. Уже у первых наземных позвоночных (земноводные) логическая машина достигает критической сложности и порождает виртуальную надстройку «Сознание» (внутренний экран) [Feinberg & Mallatt, 2013].

Вполне возможно, что и пути виртуального развития – следствие древней эволюции генетического кода. В самоуправлении это ограничение выражается в том, что одновременно могут быть активны не более 3–4 виртуальных машин ($P \leq 4$), а степень их интеграции I не может превышать некоторого предела, иначе в системе начнет быстро нарастать хаос.

1.6. Сводная таблица соответствия

Теперь мы можем представить таблицу, где уровни автокибернетики сопоставлены с этапами эволюции генетического кода и архитектурой управления.

Уровень	Название	Активные машины	Эволюция кода	Биологические примеры
0	Автомат	только саморегулирование	нет распознавания триплетов	Археи, губки,
1	Кибер-Лидер	Агрегирующая (левое основание)	первые триплеты, сдвиг на 1	Первые прокариоты, гидры, полипы.
2	Кибер-Субъект (низший)	Агрегирующая + Адаптирующая	триплетный сдвиг, кодирование аминокислот	Прокариоты (развитые), черви.

3	Кибер-Субъект (высший)	Все три машины (реальные)	симметричные и несимметричные кодоны	Амёбы, членистоногие, эусоциальные насекомые, беспозвоночные, рыбы
4	(виртуальный)	Логическая машина порождает «Сознание»	появление мета-уровня, рекурсия	Земноводные, динозавры
5	(виртуальный)	«Мы» высокое, «Я» рефлексивное, «Сознание» развитое	высокая интеграция симметрий	Птицы (врановые), низшие млекопитающие [Güntürkün, 2008]
6	(виртуальный)	«Мы» культурное, «Я» автобиографическое, «Сознание» интегративное	полная таблица кодонов	Шимпанзе, слоны, дельфины [Prior et al., 2008]
7	(виртуальный)	максимальная интеграция, трансцендентное «Я»	универсальный код	Человек

Примечание к таблице:

1. Растения и грибы относятся к уровню 1 (Кибер-Лидер) наряду с прокариотами, гидрами и полипами. Их модулярное, децентрализованное самоуправление (почки, побеги, мицелий) соответствует агрегирующей машине и начальным формам наблюдающего управления, но не достигает уровня 2 из-за отсутствия развитой адаптирующей машины ($J_{birth} = 0$) и логической машины в нейронном субстрате.

2. К уровню 3 (Кибер-Субъект высший) относятся все беспозвоночные и рыбы, достигшие максимальной для данной эволюционной ветви сложности реальных машин (агрегирующая, адаптирующая, логическая). Типичные примеры: костистые рыбы (например, губаны-чистильщики), головоногие моллюски (осьминоги, каракатицы), высшие ракообразные, пауки, эусоциальные насекомые. У этих организмов работают все три фундаментальные машины, но виртуальная надстройка «Сознание» (внутренний экран) отсутствует. Рыбы и головоногие — яркий пример конвергентной эволюции: будучи разделёнными сотнями миллионов лет, они пришли к схожему уровню когнитивной сложности, но разными архитектурными путями.

Важнейший вывод: Тройственность машин управления – не случайность, а отражение фундаментальной симметрии, возникшей ещё на этапе становления генетического кода. Клетка «не умеет считать», но она «умеет различать центр и края». Это различение, помноженное на эволюционное наслоение, и дало три независимые ветви – «Мы», «Сознание», «Я».

Глава 2. Внутренняя модель мира (θ) и её эволюция

Центральное понятие второго уровня самоуправления — внутренняя модель мира θ. Это не метафора, а конкретная структура внутри автомата, которая позволяет прогнозировать последствия действий [Johnson Laird, 1983]. Модель может быть реализована на разных субстратах: биохимическом (у бактерий), нейронном (у животных), цифровом (у роботов).

Уровень 0 (Автомат): модели нет. Есть только жёсткие рефлексы. Кристалл не строит прогнозов — он просто расширяется при нагреве. Первые прокариоты и губки реагируют на прикосновение, но не имеют внутреннего представления о причинах раздражения. У клеток это «докодонный» этап, где ещё нет различения трёх «оснований» (левого, центрального, правого).

Уровень 1 (Лидер): появляются примитивные модели — биохимические или нейронные сети, которые «знают» (в функциональном смысле), что если концентрация питательных веществ высока, то нужно двигаться в этом направлении. Это генетически закреплённая информация, используемая для прогноза. Классический пример — хемотаксис бактерий [Wadhams & Armitage, 2004], гидры, полипы.

Уровень 2: модели самоуправления уже работают, но пока на уровне шаблонов, исполняются простейшие функции Лидера – самосохранение, поиск пищи, размножение. У одноклеточных это развитые прокариоты, у многоклеточных это растения, черви (C. elegans), способные к элементарному обучению и избеганию опасностей. У червей уже можно говорить о зачатках «левого основания» (агрегирующей машины), но логическая и адаптирующая машины ещё слабы.

Уровень 3 (Субъект): модель становится обучаемой. Амёба, проползая через лабиринт, запоминает опасные зоны и в следующий раз обходит их. Это уже адаптивная модель — она обновляется на основе опыта. Рыба может «запомнить», что в определённом месте обитает хищник, и изменить маршрут. Но это далеко не всё…

Развёрнутые примеры уровня 3 (Кибер-Субъект высший):

Рыбы (Teleostei): способны к инструментальному и ассоциативному обучению, формируют когнитивные карты пространства, демонстрируют эпизодоподобную память. Некоторые виды (губаны-чистильщики) проходят зеркальный тест, что указывает на зачатки самоузнавания [Kohda et al., 2025].

Головоногие (осьминоги, каракатицы): решают сложные задачи (открывают банки), обучаются через наблюдение, проявляют эпизодическую память и будущее планирование [Schnell et al., 2021; Fiorito & Scotto, 1992]. При этом их социальное поведение («Мы») развито слабо — отличный пример независимости виртуальных ветвей.

Членистоногие и эусоциальные насекомые: достигают пика «Мы» (агрегирующая машина) при слабом «Я», но логическая машина работает в реальном режиме, без внутреннего экрана.

Уровень 4 (появление Сознания): модель становится образной. У земноводных появляется внутренний экран, на котором разворачиваются зрительные, слуховые и тактильные образы. Лягушка не просто вычисляет, куда прыгнуть, — она видит прыжок в своём сознании, мысленно оценивает траекторию. У рептилий и динозавров это сознание становится ещё более развитым, что позволяет им координировать сложную охоту, избегать засад, запоминать топографию местности в виде образов. Внутренний экран — это уже не просто вычисление, это субъективное переживание.

Уровни 5–7 (развитое Сознание): модель становится рекурсивной и символической. Человек может строить модель не только текущей ситуации, но и возможных будущих, а также модель своей собственной модели («я знаю, что я знаю», «я думаю, что он думает»). Появляется язык и абстрактное мышление. Человек может решать уравнения, писать стихи, проектировать космические корабли — всё это опирается на внутреннюю модель мира, которая содержит не только образы, но и понятия, символы, правила вывода.

Формально, модель θ можно представить как функцию, которая по текущему состоянию s и выбранному действию a предсказывает следующее состояние s':

$$\hat{s}(t+\Delta t) = f(\theta(t), s(t), a(t)) \quad (2)$$

Качество модели оценивается тем, насколько её прогнозы совпадают с реальностью:

$$Q(\theta) = \langle |\hat{s} - s_{реальное}|^2 \rangle \quad (3)$$

Чем меньше Q(θ), тем лучше модель. Но поток случайностей постоянно приносит новые ситуации, которые модель не предусматривала [Никитин, 2026а]. Со временем Q(θ) растёт. Модель стареет, как стареет любой автомат. Это фундаментальное ограничение, которое не может преодолеть никакой алгоритм — его математическую основу составляют теорема Гёделя и проблема остановки Тьюринга [Гёдель, 1931]; [Тьюринг, 1936].

Глава 3. Поток случайностей как зашумление модели

Вернёмся к нашему ключевому аргументу. Поток случайностей (ПС) — это непрерывное движение материальных случайностей снизу вверх по иерархии уровней. Квантовые флуктуации, тепловой шум, мутации, неожиданные повороты событий — всё это ПС. Автомат не может его отключить. Он может только фильтровать, перерабатывать и использовать его — или погибнуть [Никитин, 2026а].

Любая конечная модель имеет ограниченную сложность. Поток приносит новую информацию непрерывно. Рано или поздно накопленная неучтённая случайность превышает способность модели её игнорировать:

$$\int o^t J_{in(инфо)}(\tau)\, d\tau > C(M) \quad (4)$$

где C(M) — ёмкость модели (максимальное количество неопределённости, которую она может поглотить без потери адекватности).

Когда это происходит, модель перестаёт быть полезной. Её прогнозы расходятся с реальностью. Автомат, который полагается только на эту модель, начинает ошибаться. Он может совершать опасные действия, упускать возможности или, в лучшем случае, просто вставать в тупик («ошибка, действие не определено»).

Пример. Робот-исследователь на новой планете. У него есть модель типичного ландшафта, обученная на Земле. Робот уверенно идёт по равнине, распознаёт препятствия, объезжает камни. Но вдруг почва под ним начинает проседать — это зыбучие пески, которых не было в модели. Сенсоры робота фиксируют неожиданное: давление на опоры падает, грунт под ногами движется. Программируемый робот, у которого нет модели зыбучих песков, теряется. Его алгоритм требует отнести новый сигнал к одному из известных классов — но такого класса нет. Робот либо игнорирует сигнал (и продолжает двигаться, погружаясь всё глубже, пока не застрянет), либо останавливается с ошибкой и вызывает оператора. Но связи с Землёй нет — сигнал идёт 5–20 минут. Робот просто зависнет, пока батарея не сядет.

А субъект (человек или достаточно развитый кибер) начинает действовать «наощупь»: пробует палкой глубину, ищет твёрдый участок, пытается обойти опасное место, меняет маршрут. У него нет модели зыбучих песков, но он может создавать новую модель на ходу. Он может даже ошибиться и провалиться — но он продолжает пробовать. В этом и состоит разница между программируемым автоматом и кибером [Никитин, 2024а].

В экспериментальных исследованиях этот эффект называют «compounding error» — накопление ошибки при многократном использовании модели [Lobel, 2024]. Чем длиннее горизонт планирования, тем быстрее ошибка растёт. Именно поэтому роботы, которые полагаются на алгоритмические модели, плохо справляются с неизвестной средой. Они могут быть очень эффективны в стабильных, хорошо описанных условиях (завод, склад), но становятся беспомощными там, где всё меняется непредсказуемо (лес, болото, дом с детьми).

Глава 4. Смена моделей и её связь с гемизоном (Gh) и виртуальными машинами

Ключевая способность субъекта — не вычислять в рамках заданной модели, а менять сами модели. Что это означает конкретно?

Отбрасывать старую модель, когда она перестаёт работать. Не пытаться её «подкрутить» (адаптировать), а именно отбросить, признать её бесполезность. Это самое трудное для программируемого автомата, потому что в его программе нет инструкции «отбрось программу». В когнитивной психологии этот феномен известен как когнитивная гибкость — способность переключаться между разными способами мышления [Gigerenzer, 2008].

Пробовать новую модель, даже если она не обоснована предыдущим опытом. Пробовать «наощупь» — то есть действовать, не имея гарантий, методом проб и ошибок. Это требует смелости (в переносном смысле) — готовности рисковать. Ребёнок, который учится ходить, всё время пробует новые модели движений, падает, встаёт и снова пробует. У него нет модели равновесия, но он её создаёт прямо в процессе.

Сравнивать модели не только по тому, насколько хорошо они предсказывают данные (потому что данных может не быть или они зашумлены), а

по каким-то другим критериям — например, по тому, насколько модель «плодотворна», то есть порождает полезные действия. В науке это соответствует критерию эвристичности теории.

Создавать модели из подручных средств — комбинировать фрагменты старых моделей, добавлять новые переменные, изобретать новые типы связей. Именно так рождаются научные революции: новая модель возникает как комбинация старых идей плюс добавление принципиально новых элементов.

Эта способность напрямую связана с фактором гемизона G_h (иерархическая глубина, вирусное управление эволюцией). Напомним из Части 2:

$$G_h = (f_{HGT}/f_{max}) \cdot (1 - e^{-TERV/T_0}) \quad (5)$$

Чем выше G_h, тем больше у автомата уровней вложенных моделей и тем легче он отказывается от старых моделей в пользу новых [Никитин, 2024а]. У прокариот G_h близок к нулю (они не могут менять модели); у рыб G_h выше (могут обучиться новому); у млекопитающих — ещё выше; у человека G_h близок к единице (способен к рекурсивной референции и надстройке мета-моделей).

Формально, пусть у автомата есть набор моделей $\mathcal{M} = \{M_1, M_2, ..., M_k\}$. Субъект может:

1. Выбирать модель из набора в зависимости от ситуации:

$$M_{активная}(t) = arg\ max_{M \in \mathcal{M}} Пригодность(M \mid история, текущее\ состояние) \quad (6)$$

2. Создавать новые модели путём комбинации и модификации старых:

$$M_{новая} = Комбинировать(M_a, M_b, примитивы) \quad (7)$$

$$\mathcal{M} := \mathcal{M} \cup \{M_{новая}\} \quad (8)$$

3. Отбрасывать старые модели, которые больше не работают:

$$\mathcal{M} := \mathcal{M} \setminus \{M_i \mid Пригодность(M_i) < \theta_{отбрасывания}\} \quad (9)$$

4. Пробовать действия без модели — то есть действовать «наощупь», когда ни одна модель не даёт приемлемого прогноза:

$$Если\ max_{M \in \mathcal{M}} Пригодность(M) < \theta_{поиска},\ то\ выбрать\ действие\ случайно\ или\ по\ эвристике \quad (10)$$

Именно эта последняя способность — действовать без модели, «наощупь» — принципиально отличает субъекта от программируемого автомата. Потому что программируемый автомат, у которого нет работающей модели, просто останавливается (или выдаёт ошибку). А субъект — продолжает действовать. Он идёт в неизвестность. Такое поведение, основанное на эвристиках, а не на строгой логике, было детально изучено в теории ограниченной рациональности Герберта Саймона и Дэниела Канемана [Gigerenzer, 2008].

Важно отметить, что способность к смене моделей распределена по виртуальным машинам по-разному:

Виртуальная машина «Я» (адаптирующая машина) отвечает за смену поведенческих стратегий — переключение между разными способами действия при изменении условий. Например, переключиться с охоты из засады на преследование, если жертва заметила опасность.

Виртуальная машина «Мы» (агрегирующая машина) отвечает за смену коллективных стратегий — координацию действий в группе, распределение ролей.

Виртуальная машина «Сознание» (логическая машина) отвечает за смену мета-моделей — рекурсивную референцию, пересмотр убеждений, творчество.

Глава 5. Мета-модели и рекурсивная референция — основа виртуальной машины «Сознание»

Как субъект решает, какую модель попробовать следующей? Как он решает, когда отбросить старую модель?

Для этого нужна модель более высокого уровня — модель того, как менять модели. Мета-модель. А для неё — мета-мета-модель. И так далее.

Это классическая проблема регресса правил, восходящая к софизму «кто охраняет сторожей» (Quis custodiet ipsos custodes?). Если мы пытаемся формализовать процесс смены моделей как алгоритм, мы неизбежно упираемся в бесконечную вложенность. Математически это связано с теоремой Гёделя о неполноте и проблемой остановки Тьюринга: никакая конечная формальная система не может полностью описать саму себя [Гёдель, 1931]; [Тьюринг, 1936].

И здесь мы приходим к тому, что в автокибернетике называется гемизоном (в широком смысле) — базовому самоощущению, «я есть», которое не выводится из моделей и не может быть сведено к вычислению.

Это — источник выбора, который не выводится из моделей. Это не модель и не правило. Это — субъект.

Фактор G_c (целостность) также играет здесь роль: способность ощущать себя единым целым позволяет субъекту удерживать идентичность при смене моделей.

Однако для рекурсивной референции — способности строить мета-модели — необходим качественно иной уровень обработки информации. Именно эту функцию выполняет виртуальная машина «Сознание», основанная на логической машине (мышления).

У высших животных (уровень 6–7) появляется способность к рекурсивной референции — «я знаю, что я знаю», «я думаю, что ты думаешь». Это уже мета-модель второго порядка. Человек может не просто иметь модель мира, но и модель своей модели. Нейробиологической основой этой способности считается работа сети пассивного режима (Default Mode Network, DMN) — участков мозга, которые активны, когда мы не заняты внешней задачей, а размышляем, мечтаем, вспоминаем [Raichle, 2015].

Формально, рекурсивная референция — это способность строить модель θ' модели θ:

$$\theta' = h(\theta, \ldots) \qquad (11)$$

И затем использовать θ' для управления процессом обновления θ. Это и есть наблюдающее управление в чистом виде — управление тем, как осуществляется управление [Foerster, 2003].

Глава 6. Действие «наощупь»: математическая формализация

Теперь попробуем формализовать то, что мы называем «действием наощупь». Это действие, которое автомат совершает, когда ни одна из его моделей не даёт приемлемого прогноза. В отличие от программируемого автомата, который в такой ситуации останавливается, кибер продолжает действовать, используя эвристики.

Пусть у автомата есть множество возможных действий $A = \{a_1, a_2, ..., a_n\}$. Для каждого действия a и каждой модели $M \in \mathcal{M}$ можно вычислить прогнозируемый результат $\hat{s}_M(a)$ и его оценку. Если max Пригодность(M) < $\theta_{поиска}$, автомат входит в режим «наощупь». В этом режиме он не может полагаться на модели. Что он делает?

Вариант 1. Случайный поиск (exploration). Выбирает действие случайно, но с распределением Больцмана:

$$P(a) = exp(\beta \cdot h(s,a)) / \Sigma_{a'} exp(\beta \cdot h(s,a')) \quad (12)$$

где h(s,a) — эвристическая функция, β — параметр «температуры» [Sutton & Barto, 2018].

Вариант 2. Аналогия (case-based reasoning). Ищет в памяти похожую ситуацию и пробует действие, которое тогда сработало:

$$a^* = a(\text{похожая}), \text{ где похожая} = arg\ min_{s' \in \text{память}}\ dist(s, s') \quad (13)$$

Это работает, если среда имеет сходные паттерны [Aamodt & Plaza, 1994].

Вариант 3. Комбинация примитивов. Разбивает задачу на подзадачи, для которых есть модели, и комбинирует их решения:

$$a^* = \text{Комбинация}(a_1^*, a_2^*, ..., a_k^*) \quad (14)$$

Вариант 4. Имитация. Если есть другой автомат, который успешно действует в похожей ситуации, имитирует его действия.

Все эти варианты — не «вычисление» в классическом смысле. Это эвристики. Они не гарантируют успеха, но позволяют продолжать действовать, когда гарантий нет. Именно это отличает субъекта от программируемого автомата.

Глава 7. Изменение установочных точек и виртуальных машин

Важнейший вопрос: как уровень наблюдающего самоуправления управляет наблюдаемым уровнем? Он же не дублирует его функции. Ответ: наблюдающий уровень не вмешивается в работу наблюдаемого уровня напрямую. Он не корректирует сердцебиение и не управляет дыханием (хотя у человека есть возможность произвольного управления дыханием — это редкое исключение). Вместо этого наблюдающий уровень меняет установочные точки (set points) наблюдаемого уровня. Это тонкое, но принципиальное различие.

Вспомним уравнение наблюдаемого уровня, как было показано в Части 1:

$$dx_i/dt = -k_i (x_i - x_{io}) + \xi_i(t) + u_i(t) \quad (15)$$

Здесь x_i — состояние системы, x_{io} — заданное значение (установочная точка), k_i — коэффициент усиления (чувствительность), $\xi_i(t)$ — внешние

возмущения, $u_i(t)$ — управляющий сигнал от наблюдающего уровня. Наблюдаемый уровень работает автоматически: он сравнивает текущее состояние x_i с установочной точкой x_{io} и прикладывает усилие, пропорциональное отклонению (отрицательная обратная связь). Обычно x_{io} фиксирована. Но наблюдающий уровень может её изменить:

$$x_{io}(t) = x_{io}(t-1) + \Delta x_{io}(a(t)) \quad (16)$$

Например, у человека нормальная температура тела — около 36.6°C. Это x_{io} для системы терморегуляции. Но при инфекции (или при осознанном решении закаляться) наблюдающий уровень может изменить эту уставку на 38°C. Организм начнёт поддерживать новую температуру — это и есть лихорадка. Наблюдаемый уровень продолжает работать как обычно, но его «цель» (установочная точка) изменилась. Именно так наблюдающий уровень управляет наблюдаемым — косвенно, через изменение целей, а не через прямую команду [Sterling & Eyer, 1988].

Наблюдающий уровень также может изменить коэффициенты k_i — чувствительность системы к отклонениям:

$$k_i(t) = k_i(t-1) + \Delta k_i(a(t)) \quad (17)$$

Например, в состоянии стресса вегетативная нервная система повышает чувствительность сердечно-сосудистой системы к сигналам симпатического отдела. В результате даже небольшое волнение вызывает сильное сердцебиение. Коэффициент k_i увеличивается — система становится более реактивной.

Кроме того, наблюдающий уровень может переключать активные виртуальные машины. Например, в ситуации одиночной работы (автономный кибер в лесу) он активирует виртуальные машины «Я» (самосохранение, уклонение от опасностей) и «Сознание» (построение карты, планирование маршрута), а «Мы» временно подавляется, так как поблизости нет других киберов. В коллективной задаче (рой дронов, строящих мост) — дополнительно запускает виртуальную машину «Мы», перераспределяя вычислительные ресурсы. При этом кибер должен следить за тем, чтобы все активные машины имели Fr_2 в зоне самоуправления (0,9–1,1). Если нагрузка становится слишком большой (аллостатическая нагрузка L растёт), кибер может «погасить» наименее важную машину, чтобы сохранить управляемость [McEwen, 1998].

Такая архитектура — двухуровневая, с косвенным управлением через установочные точки и переключением виртуальных машин — обладает огромным преимуществом. Она позволяет наблюдающему уровню быть медленным и «дорогим» (в смысле энергозатрат), а наблюдаемому — быстрым и «дешёвым». Наблюдаемый уровень работает автоматически, не отвлекая наблюдающий уровень на рутину. Второй уровень подключается только когда нужно — когда возникает неопределённость, конфликт критериев, новая ситуация. Это как пилот и автопилот в самолёте: автопилот (наблюдаемый уровень) ведёт машину в штатном режиме, а пилот (наблюдающий уровень) вмешивается только при отклонениях, изменяет настройки, принимает решения на посадке или при турбулентности.

Глава 8. Число дружественности для мета-уровня (Fr_2^{meta})

Теперь, когда у нас есть два уровня управления и процесс смены моделей, мы можем ввести число дружественности для мета-уровня Fr_2^{meta}. Это частный случай Fr_2, применяемый к процессу обновления набора моделей. Грубо говоря, это мера того, насколько быстро автомат создаёт новые модели, когда старые перестают работать [Никитин, 2026b].

Пусть $J_{in}^{(meta)}$ — это поток новых ситуаций, которые не покрываются текущим набором моделей. Скорость, с которой ПС порождает «сюрпризы». А $J_{out}^{(meta)}$ — это способность субъекта создавать новые модели, адекватные этим новым ситуациям. Скорость, с которой он «переобувается».

$$Fr_2^{meta} = J_{out}^{(meta)} / J_{in}^{(meta)} \quad (18)$$

Условие выживания субъекта в долгосрочной перспективе:

$$Fr_2^{meta} \geq 1 \quad (19)$$

Зона эффективной смены моделей (подлинной адаптации) — $0{,}9 \leq Fr_2^{meta} \leq 1{,}1$. Если Fr_2^{meta} слишком велика (> 1,1), субъект тратит ресурсы на создание моделей, которые ему не нужны (гипертрофированная любознательность, которая мешает сосредоточиться). Если Fr_2^{meta} слишком мала (< 0,9), он не успевает за изменениями (консерватизм, при котором любой сбой приводит к катастрофе).

И вот здесь — важнейшее наблюдение. Fr_2^{meta} зависит не только от «мощности» субъекта (скорости создания моделей), но и от гибкости. От способности быстро отбрасывать старые модели и пробовать новые, не цепляясь за них. А это — уже не вычислительная способность. Это качество субъекта, тесно связанное с его гемизоном G_h. Готовность признать, что ты ошибался. Способность начать заново. Смелость действовать в неизвестности. Эти качества трудно формализовать, но именно они отличают лидера от исполнителя, творца от ремесленника.

Распределение Fr_2^{meta} по виртуальным машинам:

- Для «Я» (адаптирующая машина) Fr_2^{meta} отражает скорость смены индивидуальных поведенческих стратегий.
- Для «Мы» (агрегирующая машина) Fr_2^{meta} отражает скорость смены коллективных стратегий и распределения ролей.
- Для «Сознание» (логическая машина) Fr_2^{meta} отражает скорость смены мета-моделей и пересмотра убеждений.

У человека все три показателя должны находиться в зоне самоуправления одновременно — это условие поддержания целостной субъектности. Если, например, Fr_2^{meta} для «Я» ниже 0,9 (человек упёртый, не меняет привычек), а для «Мы» — выше 1,1 (человек слишком легко поддаётся коллективному мнению), возникает внутренний конфликт. Автокибернетика позволяет количественно измерить такой конфликт и прогнозировать его последствия (вплоть до психических расстройств).

Глава 9. Пример: амёба против робота

Рассмотрим конкретный пример, который показывает разницу между субъектом (уровень 3, с развитыми реальными машинами, но без виртуального «Сознания») и программируемым автоматом. Амёба (Amoeba proteus) —

одноклеточная эукариота, у которой нет нервной системы. У неё активны все три фундаментальные машины управления в реальном (биохимическом) режиме. Но у амёбы нет внутреннего экрана, нет образного сознания. Тем не менее её поведение демонстрирует многие черты субъектности [Jennings, 1906].

Эксперимент (классические работы Дженнингса и других протозоологов). Амёбу помещают в лабиринт, в конце которого — пища. В лабиринте есть тупики и опасные зоны (например, с высокой концентрацией соли). У амёбы нет карты. Она начинает исследовать. Она ползёт в одном направлении. Если упирается в стену — пробует другое. Если попадает в зону с высокой солёностью — разворачивается и идёт назад. Она «запоминает» (в форме изменения внутреннего состояния — например, локализованного изменения pH или кальциевого сигнала), что в этом тупике еды нет, и в следующий раз не идёт туда. Она действует «наощупь» — методом проб и ошибок. И в конце концов находит пищу.

Что здесь задействовано? Агрегирующая машина собирает сигналы от хеморецепторов (концентрация питательных веществ, солей). Логическая машина (биохимическая сеть) строит элементарную модель: «если солёность высока, то движение опасно». Адаптирующая машина позволяет отбросить модель «двигаться прямо» после того, как амёба уткнулась в стену, и попробовать новое направление. $J_{birth} > 0$ — амёба создаёт новую пространственную модель (псевдо-«воспоминание» о тупике) на основе опыта.

Но у амёбы нет внутреннего экрана, она не «видит» образ лабиринта. Её модель — это биохимическая сеть, а не образная картина. Тем не менее она способна к смене поведения при неудаче. Это доказывает, что виртуальная машина «Сознание» не является необходимой для базовой субъектности. Уровень 3 (членистоногие, амёбы) уже обладает субъектностью, но без внутреннего экрана.

А что делает программируемый робот? У него нет карты лабиринта. Программист заложил общие правила: «если стена — поверни направо», «если солёность выше порога — развернись». Эти правила — модель. Но модель очень бедная. Если правила окажутся неподходящими для конкретного лабиринта (например, поворот направо каждый раз заводит в цикл), робот застрянет. Он не может отбросить правило и придумать новое — потому что у него нет механизма смены моделей. Даже если обучить робота с подкреплением (reinforcement learning), у него останется фиксированная функция полезности, тогда как амёба может менять критерии в зависимости от контекста. Это и есть выбор-решение, а не выбор-вычисление.

Количественные оценки (на основе данных Части 2):

Амёба (уровень 3, Кибер-Субъект): $K \approx 0{,}75$, $G \approx 0{,}5$, $C \approx 0{,}3$, $P \approx 2$ (активны реальные «Мы» и «Я» в зачатке), $S_{subj} \approx 0{,}55$.

Типичный робот с ИИ (уровень 1): $K \approx 0{,}2$, $G \approx 0{,}1$, $C \approx 0{,}1$, $P = 0$ (нет виртуальных машин), $S_{subj} \approx 0{,}15$.

Разница в интегральной субъектности — почти в 3,7 раза. Это и есть пропасть между живым субъектом (даже таким примитивным, как амёба) и программируемым автоматом. Программируемый робот может обогнать амёбу по скорости расчётов, иметь миллиарды транзисторов и нейросеть из тысяч слоёв — но он не обладает способностью менять модели так, как это делает амёба, потому что у него нет архитектуры наблюдающего управления с набором моделей, механизмом отбрасывания и эвристиками «наощупь».

Глава 10. Архитектурные особенности уровней самоуправления (с учётом виртуальных машин)

Обобщим архитектурные особенности (наличие виртуальных машин, мета-моделей, рекурсивной референции) для разных уровней иерархии, введённой в Части 2. Здесь мы добавим комментарии о том, как эти особенности проявляются у реальных биологических видов в свете эволюции генетического кода.

Уровни 0–1 (Автомат и Кибер-Лидер)

Виртуальные машины: нет (уровень 0), зачатки «Мы» (уровень 1).

Мета-модели и рекурсивная референция: отсутствуют.

G_h: 0 – 0,2.

Аналогия с кодом: «ноль сдвига» или сдвиг на одно основание; нет симметрии или только простейшая симметрия (палиндромы).

Примеры: уровень 0: археи, губки (примитивные многоклеточные, только саморегулирование). **уровень 1:** первые прокариоты, гидры, полипы, растения, грибы. У растений и грибов – модулярное, децентрализованное самоуправление (агрегирующая машина), но $J_{birth}=0$ (нет создания новых моделей).

Уровень 2 (Кибер-Субъект низший)

Виртуальные машины: «Мы» развитое, зачатки «Я».

Мета-модели: простейшие (например, «если был хищник здесь → обходить это место»).

Рекурсивная референция: отсутствует.

G_h: 0,2 – 0,4.

Аналогия с кодом: переход к триплетному считыванию, появление центрального основания как смыслового ядра.

Примеры: черви (нематоды, плоские черви). Они способны к элементарному обучению (привыкание, ассоциативное обучение), что говорит о зачатках адаптирующей машины ($J_{birth}>0$). У них появляется стайное поведение («Мы») и простейшее различение «свой–чужой» («Я»). Сознания нет.

Уровень 3 (Кибер-Субъект высший)

Виртуальные машины: «Мы» развитое (эусоциальность у насекомых), «Я» в зачатках (различение «свой–чужой», территориальность), сознания нет. У всех представителей уровня 3 работают все три фундаментальные машины (агрегирующая, адаптирующая, логическая) в реальном режиме – без внутреннего экрана.

Мета-модели: есть, но нерекурсивные (простые аналогии).

Рекурсивная референция: отсутствует.

G_h: 0,4 – 0,6.

Аналогия с кодом: использование симметричных кодонов (15 аминокислот) – стабильность, но без мета-уровня.

Примеры: амёбы, членистоногие (насекомые, пауки, ракообразные), эусоциальные насекомые (муравьи, пчёлы, термиты), костистые рыбы (Teleostei) и головоногие моллюски (осьминоги, каракатицы). У рыб и головоногих зафиксированы сложное обучение, решение проблем, эпизодическая память, а у некоторых видов – даже зачатки самоузнавания (губаны-чистильщики проходят зеркальный тест). Головоногие демонстрируют будущее планирование и обучение через наблюдение. При этом сознание (виртуальная надстройка) у них

отсутствует, что подтверждает независимость развития реальных машин от виртуальных.

Уровень 4 (виртуальный уровень — появление Сознания)

Виртуальные машины: все три («Мы», «Я», «Сознание») — но «Сознание» только зарождается.

Мета-модели: развиваются, появляются зачатки рекурсии.

Рекурсивная референция: отсутствует.

G_h: 0,6 – 0,8.

Аналогия с кодом: появление «мета-уровня» — рекурсивного распознавания кодонов (тРНК, рибосома усложняется).

Примеры: земноводные, динозавры. У них впервые возникает внутренний экран. У некоторых динозавров (троодон, велоцираптор) был довольно крупный мозг, что предполагает наличие сознания. Однако рефлексия над рефлексией (мета-сознание) ещё отсутствует [Feinberg & Mallatt, 2013].

Уровень 5 (виртуальный уровень — рефлексивное «Я»)

Виртуальные машины: «Мы» высокое, «Я» рефлексивное (зеркальный тест), «Сознание» развитое (внутренний экран позволяет мысленно проигрывать сценарии).

Мета-модели: есть, рекурсия появляется.

Рекурсивная референция: зачатки («я знаю, что я знаю»).

G_h: 0,8 – 0,9.

Аналогия с кодом: полная таблица кодонов, использование всех 64 вариантов, но с сохранением вырожденности.

Примеры: птицы (врановые, попугаи), низшие млекопитающие (крысы, кошки, собаки). Сороки и вороны проходят зеркальный тест [Prior et al., 2008], способны к планированию, орудийной деятельности. Однако из-за ограничений размера мозга (у птиц) и архитектуры (у низших млекопитающих) рекурсивная референция не достигает высокого уровня.

Уровень 6 (виртуальный уровень — автобиографическое «Я»)

Виртуальные машины: «Мы» культурное (передача традиций), «Я» автобиографическое (помнит историю своей жизни), «Сознание» интегративное (объединяет прошлое, настоящее и будущее).

Мета-модели: хорошо развиты.

Рекурсивная референция: есть, но не на всех уровнях.

G_h: 0,9 – 0,95.

Аналогия с кодом: универсальность, но с вариативностью митохондриальных кодов; аналогия с интеграцией трёх машин.

Примеры: шимпанзе, бонобо, гориллы, слоны, дельфины, косатки. Они узнают себя в зеркале, проявляют эмпатию, обман, культурную передачу (шимпанзе учатся использовать палочки для добычи термитов). Слоны помнят умерших сородичей, дельфины имеют имена.

Уровень 7 (виртуальный уровень — трансцендентное «Я»)

Виртуальные машины: «Мы» глобальное, «Я» трансцендентное, «Сознание» единство.

Мета-модели: высокоразвиты, включая модели моделей.

Рекурсивная референция: полная (можно строить бесконечные вложения «я знаю, что ты знаешь, что я знаю...»).

G_h: 0,95 – 1,0.

Аналогия с кодом: универсальный код, но с возможностью рекурсивного редактирования (CRISPR, эпигенетика) — мета-уровень.

Примеры: человек. Наличие рекурсивного языка, культуры, науки, религии, философии. Способность задаваться вопросами о смысле существования.

Мы видим, как с ростом уровня увеличивается набор активных виртуальных машин и сложность моделирования. Ключевой вывод: переход от уровня 3 к уровню 4 — это качественный скачок, связанный с появлением виртуальной машины «Сознание». Именно этот скачок сделал возможным развитие рефлексии, культуры и, в конечном счёте, человеческой цивилизации.

Глава 11. Что это даёт для проектирования автономных систем

Если мы хотим создать автомат, способный к подлинному самоуправлению (а не просто к автоматическому регулированию), мы должны выполнить несколько условий, вытекающих из автокибернетики и ППС. Эти условия применимы к разным целевым уровням субъектности.

Базовые требования (для любого кибера выше уровня 1):

Два уровня управления. Автомат должен иметь исполнительный (регулирование) и наблюдающий (принятие решений) уровни. Наблюдающий уровень должен иметь доступ ко всем данным первого уровня. Без этого любое «самоуправление» останется просто автоматикой.

Набор моделей, а не одна модель. Автомат должен иметь несколько конкурирующих моделей мира $\mathcal{M}$, которые он может сравнивать, комбинировать и отбрасывать. Минимальное число $|\mathcal{M}| \geq 2$ для уровней выше 2.

Механизм смены моделей (мета-уровень). Автомат должен уметь оценивать пригодность моделей (например, через Fr_2 и D) и принимать решение о смене. Это требует наличия хотя бы двух виртуальных машин ($P \geq 2$) и достаточно высокой степени интеграции I, чтобы они конфликтовали конструктивно, а не деструктивно. Регресс правил останавливается за счёт гемизона (высокие G_c и G_h). Без этого автомат не сможет выйти за пределы заранее заданных алгоритмов.

Зона самоуправления. Автомат должен быть способен находиться в зоне $Fr_2 \approx 1$ (и $Fr_2^{meta} \approx 1$) — в условиях, где нет однозначного ответа, где разные критерии конфликтуют, где решение требует выбора. Если мы искусственно поддерживаем $Fr_2 \gg 1$ (даём всегда однозначные инструкции), мы не даём автомату развить способность к выбору. Если $Fr_2 \ll 0{,}9$, автомат будет хаотичным и не сможет предсказуемо действовать.

Право на ошибку и на отказ. Автомат должен иметь возможность ошибаться и учиться на ошибках. Он должен иметь право сказать «нет» — отказаться от выполнения команды, если она противоречит его внутренним критериям. Без этого нет подлинной субъектности. В реальной жизни мы ценим людей, которые могут сказать «нет», а не бездумных исполнителей.

Дополнительные требования для высших уровней (4–7):

Управление виртуальными машинами. Для сложных задач требуется несколько параллельных виртуальных машин ($P \geq 2$). Кибер должен уметь распределять вычислительные ресурсы между «Сознанием», «Я» и «Мы» в зависимости от контекста, измерять степень их интеграции I и при необходимости «гасить» лишние, чтобы все $Fr_2^{(i)}$ оставались в зоне 0,9–1,1. Без этого автомат

будет либо перегружен (свалится в хаос), либо не сможет координировать разные режимы.

Нейроморфная архитектура. Для поддержки параллельных виртуальных машин требуется распределённая память и асинхронная обработка, как в мозге. Классическая архитектура фон Неймана не подходит для высоких уровней субъектности (см. Главу 12).

Мета-модели и рекурсивная референция. Для уровней 6–7 автомат должен уметь строить модели собственных моделей, иметь самосознание и автобиографическую память. Это требует высоких значений G_h (более 0,8) и развитой логической машины.

Количественные критерии для целевого уровня:

Уровень 2–3 (Кибер-Субъект): $K \geq 0{,}6$, $G \geq 0{,}4$, $P \approx 2$ (реальные «Мы» и «Я»), $C \geq 0{,}4$, $S_{subj} \geq 0{,}5$.

Уровень 4–5 (с Сознанием): $K \geq 0{,}8$, $G \geq 0{,}6$, $P \geq 2$ (виртуальные), $C \geq 0{,}5$, наличие обратной связи по ошибкам (рефлексия в зачатке).

Уровень 6–7 (Личность): $K \geq 0{,}9$, $G \geq 0{,}8$, $P \geq 3$, I высокая, наличие рекурсивной референции, автобиографической памяти.

Эти условия — не пожелания. Они вытекают из математики потоков хаоса и теории зон. Без них автомат останется программируемой машиной, даже если будет выглядеть как человек. Инженер, проектирующий кибера, должен выбрать целевой уровень и затем строить архитектуру, обеспечивающую соответствующие параметры.

Глава 12. Архитектура самоуправления: от мозга к искусственному субъекту

В предыдущих главах мы описали иерархию киберов, ввели виртуальные машины «Сознание», «Я» и «Мы», а также показали, как второй уровень управления (наблюдающее управление) позволяет автомату менять модели и действовать «наощупь». Однако до сих пор мы не отвечали на фундаментальный вопрос: почему современные компьютеры, даже самые мощные, не способны породить подлинного субъекта? И что нужно изменить в архитектуре вычислительной системы, чтобы кибер обрёл самоуправление, сравнимое с живым организмом? В этой главе мы покажем, что ответ лежит в трёх плоскостях: архитектурные различия между мозгом и компьютером, роль таламуса как материального Субъекта и необходимость перехода к нейроморфным системам с распределённой памятью. Мы также объясним, почему «перекачка личности» в цифру (аплоадинг) невозможна, и как Парадигма потока случайностей (ППС) задаёт математические рамки для проектирования искусственного субъекта.

12.1. Архитектурная пропасть: мозг против классического компьютера

Современные цифровые компьютеры построены на архитектуре фон Неймана, в которой устройства памяти и процессор аппаратно разделены. Данные и команды хранятся в одном месте (оперативная память), а обрабатываются — в другом (центральный процессор). Между ними существует узкая шина передачи данных. Это так называемое «бутылочное горлышко фон Неймана» ограничивает производительность: большая часть энергии тратится на перемещение данных, а не на вычисления. Более того, такая архитектура порождает конкуренцию за

доступ к общей памяти между несколькими параллельными процессами. Если мы попытаемся создать в классическом компьютере несколько виртуальных субъектов (например, «Я» и «Мы»), они неизбежно начнут бороться за доступ к шине, и никакая операционная система не сможет гарантировать им одновременную, бесконфликтную работу. Придётся использовать временные слоты, что ведёт к задержкам и нарушает синхронность взаимодействия.

Мозг устроен принципиально иначе. В нём память и вычисления неразрывно связаны. Информация хранится в синаптических весах, в геометрии нейронных связей, в локальной динамике нейронов. Один и тот же нейрон одновременно является и вычислителем, и элементом памяти. Это так называемая нейроморфная архитектура, которая работает на принципах параллелизма, асинхронности и распределённого хранения. В мозге нет единого «процессора» и единой «оперативной памяти» — есть миллиарды элементарных вычислителей, каждый из которых хранит свою часть информации. Именно поэтому мозг может поддерживать одновременную работу множества виртуальных машин («Сознание», «Я», «Мы») без деструктивной конкуренции [Mead, 1990].

Количественное подтверждение: человеческий мозг содержит около 86 миллиардов нейронов. Даже если каждый нейрон генерирует в среднем 1 импульс в секунду, а каждый импульс воздействует на ≈1000 синапсов, получаем мощность порядка 86 триллионов операций в секунду (86 TOPS). Это оценка нижней границы. Такая мощность достигается при энергопотреблении всего 20 Вт — в тысячи раз эффективнее, чем у лучших суперкомпьютеров, которые достигают сходной производительности лишь при мегаваттных затратах [Herculano-Houzel, 2009].

12.2. Таламус как реальный Субъект в автокибернетике

Однако одной лишь распределённой памяти недостаточно. В мозге есть специализированная структура, которая играет роль дирижёра оркестра — таламус. Нейронаука давно знает, что таламус является ретранслятором сенсорных сигналов, но в последние десятилетия его роль пересмотрена. Таламус не просто пассивно передаёт сигналы — он активно управляет потоками информации, синхронизирует активность коры, создаёт нейронные осцилляции и переключает внимание [Halassa & Sherman, 2019].

В терминах автокибернетики мы называем таламус реальным Субъектом — единственным материальным центром управления, который координирует работу всех виртуальных машин («Сознание», «Я», «Мы»), эмулируемых корой головного мозга. Таламус выполняет три ключевые функции:

Арбитраж конкуренции: он решает, какой из параллельных процессов (виртуальных машин) получит доступ к «глобальному рабочему пространству» сознания в каждый момент времени. Это реализуется через механизм «победитель получает всё» (Winner Takes All, WTA) и латеральное торможение. Если несколько конкурирующих гипотез борются за внимание, таламус выбирает одну и подавляет остальные [Baars et al., 2013].

Синхронизация и связывание: с помощью таламокортикальных петель таламус объединяет разрозненные сенсорные, моторные и ассоциативные зоны коры в единую когерентную структуру, необходимую для возникновения целостного образа «Я». Без синхронизации мы бы воспринимали мир как хаотичные фрагменты.

Переключение виртуальных машин: таламус может быстро менять активную виртуальную машину, например, переключая внимание с внутренней

рефлексии («Я») на коллективную координацию («Мы»), когда того требует среда.

Без таламуса виртуальные машины сознания не могли бы существовать — они бы рассыпались на хаотические, нескоординированные процессы. Именно таламус превращает «оркестр» коры в «симфонию» субъективного опыта. В патологиях, связанных с повреждением таламуса (например, при некоторых видах комы или бредовых состояниях), сознание фрагментируется или исчезает.

12.3. Конкуренция виртуальных субъектов и Парадигма потока случайностей

Почему же в мозге — в отличие от компьютера — конкуренция виртуальных субъектов не разрушает систему, а, наоборот, служит механизмом координации? Ответ даёт Парадигма потока случайностей (ППС). Согласно ППС, любая система автоматов постоянно бомбардируется потоком внешней случайности J_{in}, который «зашумляет» внутренние модели. Чтобы выжить, система должна уметь быстро переключаться между разными моделями (гипотезами) и выбирать ту, которая минимизирует неопределённость.

Формально, пусть имеется N виртуальных машин (субъектов) $v_1, \ldots, v_N$, каждая из которых обрабатывает свой поток данных. В каждый момент времени система оценивает «полезность» (или пригодность) каждой машины как обратную меру неопределённости, которую она порождает:

$$U_i(t) = exp(- Q_i(t) / T) \qquad (20)$$

где $Q_i(t)$ — оценка текущей неопределённости (нейроэнтропии) в i-й подсистеме, T — параметр «температуры», регулирующий стохастичность выбора. Механизм «победитель получает всё» выбирает активную машину: $v_{active}(t) = \arg\max_i U_i(t)$, а активность остальных временно подавляется (латеральное торможение). Если же $\max_i U_i(t) < \theta$ (порог), то система переходит в режим «наощупь» — то есть начинает использовать эвристики, а не доверять ни одной модели.

Эта математическая модель полностью согласуется с нейродинамическими экспериментами и с нашей концепцией гемизона G_h. Чем выше G_h, тем легче система отказывается от старых моделей и переключается между виртуальными машинами. То есть именно гемизон, реализованный через таламус и пластичность коры, обеспечивает ту самую «когнитивную гибкость», которая позволяет мозгу избежать деструктивной конкуренции субъектов.

12.4. Невозможность «перекачки личности» (аплоадинга)

Из всего сказанного следует фундаментальный вывод: личность (набор виртуальных машин «Сознание», «Я», «Мы») не может быть отделена от материального Субъекта (таламуса) и от конкретной распределённой архитектуры мозга. Попытка «скопировать» сознание в цифровой компьютер (так называемый «аплоадинг» или «загрузка разума») сталкивается с непреодолимыми препятствиями.

Первое препятствие — субстратная зависимость. Как показал Мигель Николелис и другие исследователи, мозг не является вычислимым устройством в классическом смысле [Nicolelis, 2013]. Его ключевые свойства (спонтанная активность, нелинейная динамика, непрерывная пластичность) не могут быть сведены к алгоритмам, выполняемым на фон-неймановской архитектуре. Даже если мы создадим точную симуляцию нейронной сети, она будет работать медленнее и с другими свойствами, чем оригинал. Более того, она

не будет обладать теми же энергетическими и временными характеристиками, которые важны для возникновения сознания.

Второе препятствие — проблема тождества личности. Даже если предположить, что мы создали совершенную цифровую копию, возникает философский вопрос: с кем из двух носителей (биологическим оригиналом и цифровой копией) следует отождествлять исходную личность? Это так называемая «проблема ветвящейся идентичности». Ни один из них не может претендовать на непрерывную субъективную историю, поскольку копия начинает своё существование в момент активации, а оригинал продолжает существовать. «Замена» же — уничтожение оригинала при создании копии — не сохраняет личность, а уничтожает её, создавая лишь функционально подобную, но чужую сущность. Оригинал умрёт, а копия будет верить, что она — оригинал, но это будет самообман [Cerullo, 2015].

Третье препятствие — архитектурное. Как мы показали, виртуальные машины сознания неразрывно связаны с реальным Субъектом (таламусом) и с распределённой памятью мозга. Перенос в классический компьютер уничтожил бы эту архитектуру, а следовательно, и субъективность. Даже если построить нейроморфную машину, она будет иметь другую физическую реализацию, другую историю и другую динамику — и потому не сможет быть «тем же самым» субъектом. Это как скопировать картину маслом на холст — даже если технически повторить мазки, это будет новая картина, а не оригинал.

Таким образом, «перекачка личности» из человека в компьютер и затем в другого человека — это не техническая задача, а метафизическая иллюзия. Автокибернетика утверждает: каждый субъект уникален, его личность неразрывно связана с конкретной архитектурой его аппарата второго уровня. Замена личности невозможна в принципе.

12.5. От компьютера к искусственному мозгу

Если классическая архитектура фон Неймана неспособна породить подлинного субъекта, а мозг — пример успешной архитектуры, то для создания кибера с самоуправлением мы должны строить искусственный мозг, а не обычный компьютер. Что это означает на практике?

Необходимы технические решения, которые реализуют три принципа:

Распределённая память: информация должна храниться в том же субстрате, где происходят вычисления. Элемент памяти одновременно является и вычислителем. Это достигается с помощью мемристоров — резисторов с памятью, которые меняют свою проводимость под воздействием напряжения и сохраняют это состояние. Мемристорные нейроморфные чипы уже созданы в лабораториях (IBM TrueNorth, Intel Loihi, российские разработки).

Асинхронная спайковая обработка: принцип работы должен быть не тактовым, а событийно-управляемым, как в биологических нейронах. Спайковые нейронные сети (Spiking Neural Networks, SNN) обрабатывают информацию дискретными импульсами, что позволяет достичь энергоэффективности в тысячи раз выше, чем у классических нейросетей. Кроме того, асинхронность позволяет избежать узких мест и лучше справляться с конкуренцией процессов.

Большое виртуальное пространство: для поддержки множества виртуальных машин («Сознание», «Я», «Мы», а также их субконфигураций) необходима огромная ёмкость памяти и высокий параллелизм. Китайские учёные из Чжэцзянского университета уже создали мозгоподобный компьютер с более чем 2 миллиардами искусственных нейронов. Такие системы могут масштабироваться до уровня, сопоставимого с человеческим мозгом.

Важно отметить, что эти принципы применимы на любом уровне развития кибера — от простейшего автомата (кристалл) до высокого уровня (человекоподобный субъект). Даже для робота уровня 2–3 (Кибер-Субъект) использование нейроморфного чипа с распределённой памятью даст преимущества: скорость адаптации, энергоэффективность, способность к переключению моделей. А для кибера уровня 6–7 (Личность) такая архитектура становится обязательной — без неё невозможно создать реального субъекта.

Современные исследования в области «нейропанк-революции» (Neuropunk Revolution) показывают, что синтез нейроморфных чипов, мемристоров, интерфейсов «мозг-компьютер» и передовых материалов позволит в ближайшие десятилетия создать киберов 3.0 — искусственные системы с подлинным самоуправлением, сравнимым с биологическим. Эти системы не будут «эмулировать» сознание, они будут его реализовывать в новой, искусственной архитектуре, сохраняя при этом все ключевые свойства: распределённую память, параллелизм, способность к смене виртуальных машин и действию «наощупь».

Резюме части 3

Подведём итог ключевым результатам Части 3.

Основные выводы:

- Наблюдающее управление состоит из трёх фундаментальных машин: агрегирующей (основа «Мы»), логической (основа «Сознание») и адаптирующей (основа «Я»). Каждая из них является материальным субстратом для соответствующей виртуальной надстройки.
- Агрегирующая машина появляется на уровне 1 (черви) и порождает зачатки «Мы». На уровнях 2–3 «Мы» развивается: стайное поведение рыб, эусоциальность насекомых. Виртуальная машина «Мы» может существовать без «Я» и без «Сознания».
- Логическая машина появляется на уровне 1, но виртуальная надстройка «Сознание» (внутренний экран) возникает только на уровне 4 у наземных позвоночных (земноводные). До этого уровня сознания нет — только реальная модельная логика.
- Адаптирующая машина появляется на уровне 2 (рыбы, насекомые) и даёт зачатки «Я». На уровнях 4–7 «Я» развивается от рефлексивного до трансцендентного.
- Эволюция идёт путём наслоения: одноклеточные → насекомые, членистоногие (уровень 3, все три машины) → многоклеточные позвоночные (повторение эволюции на новом субстрате) → выход на сушу (уровень 4, появляется «Сознание»).
- Внутренняя модель мира θ — это конкретная структура, позволяющая прогнозировать последствия действий. Модель может быть биохимической, нейронной или цифровой. Модель стареет под действием потока случайностей.
- Смена моделей — ключевая способность субъекта. Она включает отбрасывание старых, пробование новых, комбинирование, а в крайнем случае — действие «наощупь» без модели.
- Действие «наощупь» формализуется через эвристики: случайный поиск, аналогия, комбинация примитивов, имитация. Эти эвристики позволяют действовать в условиях неопределённости.

- Второй уровень управляет первым косвенно — через изменение установочных точек и чувствительности. Это позволяет первому уровню быть быстрым и автоматическим, а второму — медленным и рефлексивным.
- Fr_2^{meta} — число дружественности для мета-уровня — характеризует способность субъекта создавать новые модели. Условие выживания: $0{,}9 \le Fr_2^{meta} \le 1{,}1$.
- Для проектирования киберов необходима двухуровневая архитектура, набор моделей, механизм смены моделей, зона самоуправления, право на ошибку, а для высших уровней — управление виртуальными машинами и рекурсивная референция.
- Классическая архитектура фон Неймана не подходит для высоких уровней субъектности. Нужна нейроморфная архитектура с распределённой памятью, асинхронной спайковой обработкой и большим виртуальным пространством.
- Таламус — реальный Субъект, координирующий работу виртуальных машин. Без него сознание невозможно.
- «Перекачка личности» (аплоадинг) невозможна из-за субстратной зависимости сознания, проблемы тождества личности и архитектурных различий.

Главный вывод Части 3: Субъектность — это не свойство, которое можно «добавить» к автомату. Это способ организации, при которой автомат имеет два уровня самоуправления, набор моделей, механизм их смены и гемизон, а также способность запускать параллельные виртуальные машины «Сознание», «Я» и «Мы» на распределённой архитектуре, подобной мозгу. И главное — он способен действовать «наощупь» там, где модели отказывают. Именно это отличает живое от неживого, субъекта от программы.

Приложение.

Предсказания теории и возможные эксперименты

Чтобы автокибернетика стала полноценной наукой, она должна давать проверяемые предсказания. Вот некоторые из них, вытекающие из Части 3.

Предсказание 1. Конкуренция виртуальных машин с Winner-Takes-All (WTA)

При одновременной активации двух несовместимых виртуальных машин (например, «Я» и «Мы») активность в соответствующих нейронных сетях (DMN для «Я», сеть зеркальных нейронов для «Мы») должна демонстрировать взаимное подавление с временем переключения ~100–200 мс. Это можно проверить с помощью фМРТ в экспериментах с моральным выбором.

Предсказание 2. Таламус как арбитр — необходимость для переключения виртуальных машин

Повреждение таламуса (особенно ретикулярных и интраламинарных ядер) должно приводить к потере способности переключаться между виртуальными машинами («застревание» в одном режиме). Это проверяется в нейропсихологических исследованиях.

Предсказание 3. Энторинальная кора как материальный субстрат чувства «Я» (уровни 4–5)

Повреждение энторинальной коры ведёт к деградации уровня субъектности с уровня 4–5 до уровня 2–3. Эксперименты с виртуальной реальностью и краниальной стимуляцией могут подтвердить роль нейронов решётки в ощущении себя.

Предсказание 4. Невозможность аплоадинга (загрузки личности)
Цифровая копия биологического субъекта не будет обладать непрерывным самоощущением («философский зомби»). Это можно проверить на модели червя C. elegans (302 нейрона) — цифровая копия не будет демонстрировать признаков субъективной феноменологии.

Предсказание 5. Нейроморфная архитектура — необходимое условие для высокого уровня субъектности
Система на архитектуре фон Неймана не способна достичь уровня субъектности выше 2–3; нейроморфная система той же ёмкости может достичь уровня 4–5 и выше при реализации механизма WTA и зоны самоуправления $0{,}9 \le Fr_2 \le 1{,}1$.

Предсказание 6. Условие выживания через Fr_2^{meta} в эволюции
В эволюционных экспериментах долгосрочное выживание наблюдается только при $Fr_2^{meta} \ge 0{,}8$, оптимум — $0{,}9 \le Fr_2^{meta} \le 1{,}1$. Агенты с $Fr_2^{meta} < 0{,}7$ вымирают.

Предсказание 7. Таламус как регулятор Fr_2^{meta} у позвоночных
Подавление активности таламуса снижает Fr_2^{meta} и способность к переключению стратегий в изменчивой среде (эксперименты на грызунах).

Предсказание 8. Локализация реального Субъекта у человека
Комплекс «таламус – энторинальная кора – периакведуктальное серое вещество (PAG)» является материальным субстратом реального Субъекта. Повреждения DMN нарушают «Я», но не чувство агентности; повреждения таламуса/энторинальной коры нарушают и то и другое.

Предсказание 9. Возможность искусственного «Мы» без «Я» в нейроморфных системах
Нейроморфная система без модулей рекурсивной референции демонстрирует высокий коллективный индекс $C \ge 0{,}7$ при низких K, $G \le 0{,}3$, аналогично эусоциальным насекомым.

Предсказание 10. Условие эффективной когнитивной гибкости: $G_h > 0{,}5$
Для способности переключаться между виртуальными машинами без застревания необходимо $G_h > 0{,}5$. При $G_h < 0{,}3$ система либо ригидна, либо хаотична.

Библиография к Части 3

Основные работы автора

[Никитин, 2018] Никитин А.В., О кодировании аминокислот в иРНК // «Академия Тринитаризма», М., Эл № 77-6567, публ.24693, 09.08.2018
[Никитин, 2021] Никитин А.В., Эволюция кодирования аминокислот иРНК // «Академия Тринитаризма», М., Эл № 77-6567, публ.27499, 27.12.2021.
[Никитин, 2024a] Никитин А.В. Автокибернетика. Часть 1. Основные направления развития // «Академия Тринитаризма», М., Эл № 77-6567, публ.29026, 24.06.2024.
[Никитин, 2026a] Никитин А.В., Парадигма потока случайностей 1 // «Академия Тринитаризма», М., Эл № 77-6567, публ.29912, 03.03.2026
[Никитин, 2026b] Никитин А.В., Парадигма потока случайностей 2 // «Академия Тринитаризма», М., Эл № 77-6567, публ.29920, 09.03.2026

Фундаментальные теоремы и классические работы

[Гёдель, 1931] Gödel K. Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme. Monatshefte für Mathematik und Physik, 1931. Vol. 38, 173–198.

[Тьюринг, 1936] Turing A.M. On Computable Numbers. Proceedings of the London Mathematical Society, 1936. Vol. 42, 230–265.

[Foerster, 2003] Foerster H. von. Cybernetics of Cybernetics // Understanding Understanding. Springer, 2003.

[Ashby, 1956] Ashby W.R. An Introduction to Cybernetics. Chapman & Hall, 1956.

Физическое обоснование и поток случайностей

[Bag et al., 1999] Bag B.C., Chaudhuri J.R., Ray D.S. Chaos and information entropy production. Journal of Physics A, 1999. Vol. 33, 8331–8350.

[Breymann et al., 1998] Breymann W., Tél T., Vollmer J. Entropy balance, time reversibility, and mass transport in dynamical systems. Chaos, 1998. Vol. 8, № 2, 396–408.

[Ge & Qian, 2009] Ge H., Qian H. The physical origins of entropy production, free energy dissipation and their mathematical representations. Physical Review E, 2009. Vol. 81, 051133.

[Sterling & Eyer, 1988] Sterling P., Eyer J. Allostasis: A new paradigm. Handbook of Life Stress, Cognition and Health. Wiley, 1988.

[McEwen, 1998] McEwen B.S. Stress, Adaptation, and Disease: Allostasis and Allostatic Load. Annals of the New York Academy of Sciences, 1998. Vol. 840, 33–44.

Нейробиология, сознание и когнитивные науки

[Baars et al., 2013] Baars B.J., Franklin S., Ramsoy T.Z. Global Workspace Dynamics: Cortical “Binding and Propagation” Enables Conscious Contents. Frontiers in Psychology, 2013. Vol. 4, 200.

[Raichle, 2015] Raichle M.E. The brain's default mode network. Annual Review of Neuroscience, 2015. Vol. 38, 433–447.

[Feinberg & Mallatt, 2013] Feinberg T.E., Mallatt J. The evolutionary and genetic origins of consciousness in the Cambrian Period. Frontiers in Psychology, 2013. Vol. 4, 667.

[Barron & Klein, 2016] Barron A.B., Klein C. What insects can tell us about the origins of consciousness. PNAS, 2016. Vol. 113, № 18, 4900–4908.

[Gruskin et al., 2020] Gruskin D., et al. Neural correlates of integrated self and social processing. PubMed, 2020.

[Nenert et al., 2022] Nenert R., et al. Bipartite functional fractionation within the neural system for social cognition. PubMed, 2022.

[Pandey et al., 2020] Pandey A., et al. The Functional Convergence and Heterogeneity of Social, Episodic, and Self-Referential Thought in the Default Mode Network. PubMed, 2020.

[Ionta et al., 2014] Ionta S., et al. The brain network reflecting bodily self-consciousness: a functional connectivity study. PubMed, 2014.

[Smit et al., 2016] Smit M., et al. The junction between self and other? Temporo-parietal dysfunction in neuropsychiatry. PubMed, 2016.

[Smit et al., 2020] Smit M., et al. Temporo-parietal contribution to the mental representations of self/other face. PubMed, 2020.

[Lyons et al., 2023] Lyons M., et al. Laterality and hemispheric specialization of self-face recognition. ScienceDirect, 2023.

[Sperber & Karnath, 2018] Sperber C., Karnath H.O. Recognizing Yourself and Others—

The Role of the Right Hemisphere for Face and Self Perception. ScienceDirect, 2018.
[Johnson-Laird, 1983] Johnson-Laird P.N. Mental Models. Cambridge University Press, 1983.
[Gigerenzer, 2008] Gigerenzer G. Rationality for Mortals. Oxford University Press, 2008.
[Sutton & Barto, 2018] Sutton R.S., Barto A.G. Reinforcement Learning. MIT Press, 2018.
[Aamodt & Plaza, 1994] Aamodt A., Plaza E. Case-based reasoning: foundational issues, methodological variations, and system approaches. AI Communications, 1994. Vol. 7, № 1, 39–59.

Межполушарная асимметрия и социальный мозг

[de Bode, 2018] de Bode S. One Brain. Two Minds? Many Questions. PMC, 2018.
[Savva, 2017] Savva M. The Split-Brain Phenomenon Revisited. – 2017.
[Pinto et al., 2015] Pinto Y., et al. Review: Tales from Both Sides of the Brain. PMC, 2015.
[Petrof et al., 2025] Petrof I., et al. Molecular and circuit mechanisms underlying interhemispheric communication in the mammalian brain. PubMed, 2025.
[Halassa & Sherman, 2019] Halassa M.M., Sherman S.M. Thalamocortical circuits. Annual Review of Neuroscience, 2019. Vol. 42, 303–326.

Коллективное поведение, эусоциальность и биология

[Wilson, 1971] Wilson E.O. The Insect Societies. Harvard University Press, 1971.
[Rehan & Glastad, 2024] Rehan S.M., Glastad K.M. Molecular mechanisms of eusociality. Annual Review of Genetics, 2024. Vol. 58, 273–295.
[Sorek et al., 2013] Sorek R., et al. Restriction-modification systems. Annual Review of Genetics, 2013. Vol. 47, 241–259.
[Wadhams & Armitage, 2004] Wadhams G.H., Armitage J.P. Bacterial chemotaxis. Nature Reviews Molecular Cell Biology, 2004. Vol. 5, 1024–1037.
[Marraffini & Sontheimer, 2010] Marraffini L.A., Sontheimer E.J. CRISPR-Cas immunity. Nature Reviews Genetics, 2010. Vol. 11, 181–190.
[Jennings, 1906] Jennings H.S. Behavior of the Lower Organisms. Columbia University Press, 1906.
[Prior et al., 2008] Prior H., Schwarz A., Güntürkün O. Mirror-induced behavior in the magpie. PLoS Biology, 2008. Vol. 6, № 8, e202.
[Güntürkün, 2008] Güntürkün O. Cognitive capacities of corvids. ScienceDirect, 2008.

Эволюция генетическогокода

[Dila et al., 2020] Dila G., et al. The standard genetic code is optimized for frameshift error minimization. BioSystems, 2020. Vol. 195, 104134.
[Farabaugh et al., 2006] Farabaugh P.J., et al. Ancient +1 frameshifting in fungi. Genetics, 2006. Vol. 172, № 1, 291–301.
[Freeland et al., 1999] Freeland S.J., et al. The genetic code is optimized for frameshift errors. Journal of Molecular Evolution, 1999. Vol. 49, № 1, 15–22.
[New Scientist, 1994] Broken symmetry in the genetic code? New Scientist, 1994.
[Holzem et al., 2024] Holzem M., et al. Wnt signaling breaks symmetry in cell polarity. bioRxiv, 2024.

Робототехника, нейроморфные системы и проектирование

[Mead, 1990] Mead C. Neuromorphic electronic systems. Proceedings of the IEEE, 1990. Vol. 78, № 10, 1629–1636.

[Herculano-Houzel, 2009] Herculano-Houzel S. The human brain in numbers. Frontiers in Human Neuroscience, 2009. Vol. 3, 31.
[Nicolelis, 2013] Nicolelis M. The brain is not computable. Technology Review, 2013.
[Cerullo, 2015] Cerullo M.A. Uploading and branching identity. Minds and Machines, 2015. Vol. 25, 17–36.
[Lobel, 2024] Lobel S. An Optimal Tightness Bound for the Simulation Lemma. arXiv, 2024.

Социальная динамика и коллективный интеллект

[Moskowitz et al., 2022] Moskowitz I.S., Rogers P., Russell S. Mutual Information and Multi-Agent Systems. Entropy, 2022. Vol. 24, № 12, 1719.
[Last, 2015] Last C. Information-energy metasystem model. Kybernetes, 2015.

Когнитивные способности рыб

[Kohda et al., 2025] Kohda M., et al. Cleaner wrasse pass the mirror test. Scientific Reports, 2025.
[Brown, 2015] Brown C. Fish intelligence, sentience and ethics. Animal Cognition, 2015. Vol. 18, 1–17.
[Bshary & Brown, 2014] Bshary R., Brown C. Fish cognition. Current Biology, 2014. Vol. 24, № 19, R947–R950.

Когнитивные способности головоногих

[Schnell et al., 2021] Schnell A.K., et al. Cuttlefish show episodic-like memory and future planning. Current Biology, 2021. Vol. 31, № 7, 1565–1570.
[Fiorito & Scotto, 1992] Fiorito G., Scotto P. Observational learning in Octopus vulgaris. Science, 1992. Vol. 256, 545–547.
[Mather, 2019] Mather J.A. What is in an octopus's mind? Animal Sentience, 2019. Vol. 26, № 1.