Никитин А.В.

Автокибернетика: от робота к киберу - 4.

Виртуальный интеллект: от роя до толпы

Аннотация

В четвёртой части работы «Автокибернетика: От робота к киберу» вводится понятие виртуального интеллекта N как эмерджентной комбинаторики роевых систем. На основе коллективного индекса C, коэффициента рефлексии γ и параметра вариативности среды μ построена единая шкала коллективного разума — от сининтеллекта (синосоциум) через суперинтеллект (гиперсоциум) до инфра-интеллекта (инфрасоциум). Показано, что кибернетические рои за счёт снятия биологических ограничений (парадокс 10 бит/с) способны порождать качественно новую надстройку — виртуальный роевой интеллект V, недоступный живой природе. Сформулирована теорема о границе управляемой субъектности для коллективов киберов: $N \leq 0{,}7$, $V \leq 776$. Предложены стратегии борьбы с неуправляемым суперинтеллектом: информационная изоляция, функциональное ограничение, протокол множественной изоляции.

Оглавление

ВВЕДЕНИЕ 2

ГЛАВА 1. ВИРТУАЛЬНЫЙ ИНТЕЛЛЕКТ N: ЭМЕРДЖЕНТНАЯ КОМБИНАТОРИКА РОЕВ 3

ГЛАВА 2. ЕДИНАЯ ШКАЛА КОЛЛЕКТИВНОГО РАЗУМА: ОТ СИНИНТЕЛЛЕКТА ДО ИНФРА-ИНТЕЛЛЕКТА 4

ГЛАВА 3. СИНОСОЦИУМ: КОНСТРУКТИВНАЯ ФОРМА ВИРТУАЛЬНОГО ИНТЕЛЛЕКТА (M → 0, N ≈ 0,3–0,6) 6

ГЛАВА 4. ГИПЕРСОЦИУМ: АДАПТИВНАЯ ФОРМА ВИРТУАЛЬНОГО ИНТЕЛЛЕКТА (M > 0,3, N ≈ 0,6–0,9) 7

ГЛАВА 5. ИНФРАСОЦИУМ: ДЕСТРУКТИВНАЯ ФОРМА ВИРТУАЛЬНОГО ИНТЕЛЛЕКТА (M > 0,7, N → 0) 8

ГЛАВА 6. ПОРОГОВЫЕ ПЕРЕХОДЫ МЕЖДУ ФОРМАМИ ВИРТУАЛЬНОГО ИНТЕЛЛЕКТА 9

ГЛАВА 7. ВИРТУАЛЬНЫЙ РОЕВОЙ ИНТЕЛЛЕКТ V: ВЕРШИНА КИБЕРНЕТИЧЕСКОГО КОЛЛЕКТИВНОГО РАЗУМА 10

ГЛАВА 8. ТЕОРЕМА О ГРАНИЦЕ УПРАВЛЯЕМОЙ СУБЪЕКТНОСТИ ДЛЯ КОЛЛЕКТИВОВ (С УЧЁТОМ N И V) 12

ГЛАВА 9. СТРАТЕГИИ БОРЬБЫ С НЕУПРАВЛЯЕМЫМ СУПЕРИНТЕЛЛЕКТОМ 13

ЗАКЛЮЧЕНИЕ К ЧАСТИ 4 14

ФАЛЬСИФИКАЦИЯ АВТОКИБЕРНЕТИКИ: КАК ОПРОВЕРГНУТЬ ТЕОРИЮ, ПРЕЖДЕ ЧЕМ СТРОИТЬ КИБЕРА 15

Принцип фальсифицируемости в автокибернетике 15
Экспериментальные тесты для фальсификации 16
Иерархия реакций на фальсификацию 16
Границы теории и открытые вопросы 17
Приглашение к независимой проверке 17

ПРИЛОЖЕНИЕ **18**
Предсказания теории и возможные эксперименты (к Части 4) 18
Ключевые предсказания Частей 1–4 19
БИБЛИОГРАФИЯ **21**

Введение

Что такое виртуальный интеллект?

В первых трёх частях мы прошли путь от простейших автоматов до Личности — от уровня 0 до уровня 7. Мы ввели три независимых измерения субъектности: *кибераутономию (K)*, *гемизон (G)* и *коллективный индекс (C)*. Кибераутономия отвечает за самостоятельное управление, гемизон — за «живость» (целостность, информацию, вирусное управление эволюцией), а коллективный индекс C — за способность автомата встраиваться в коллективный разум [Никитин, 2024а]. Но оставался вопрос: как из множества отдельных автоматов рождается коллективный разум, не сводимый к сумме индивидуальных умов? Почему одни коллективы умнее своих частей (наука, рынок, косяк рыб), другие — эффективнее (армия, муравейник), третьи — глупее (паника, толпа)?

Ответ даёт понятие **виртуального интеллекта (N)** — эмерджентной комбинаторики, возникающей при взаимодействии простых агентов. Виртуальный интеллект — это не метафора. Это строгое свойство роевых системы, которое можно измерить через коллективный индекс C, коэффициент коллективной рефлексии γ и, главное, через параметр **относительной вариативности среды μ**, который количественно описывает интенсивность потока случайностей (ПС) [Никитин, 2026а]. Буква N выбрана как нейтральный символ, обозначающий интегральную меру коллективной когнитивной мощности (от лат. *nervus* — сила, *numerus* — число, *natura* — природа).

В этой части мы покажем, что **виртуальный интеллект N един в своей основе** — от бактериального кворум-сенсинга до научного сообщества и панической толпы. Разные формы коллективного разума (сининтеллект, суперинтеллект, инфра-интеллект) — это разные режимы одного и того же эмерджентного феномена, переключение между которыми определяется параметром μ. Поток случайностей выступает здесь как универсальный «переключатель» режимов. Для коллективов киберов, предназначенных для выполнения сложных задач и остающихся под управлением человека, необходимо соблюдать **теорему о границе управляемой субъектности**: $K \le 0{,}9$, $G \le 0{,}5$, $C \le 0{,}7$, $\gamma \le 0{,}05$, $N \le 0{,}7$. Превышение этих значений ведёт к появлению «Я» и требует не управления, а согласования.

Глава 1. Виртуальный интеллект N: эмерджентная комбинаторика роев

Виртуальный интеллект N — это способность коллектива из множества простых агентов решать задачи, адаптироваться и принимать решения, используя только локальные взаимодействия. Термин «виртуальный» подчёркивает, что разум не принадлежит ни одному из агентов, а существует как эмерджентное свойство их сети. В этом смысле виртуальный интеллект — это чистая комбинаторика: из малого набора локальных правил возникает глобальное сложное поведение. Классический пример — три правила Рейнольдса для стаи птиц: избегание столкновений, выравнивание скорости, притяжение к центру. Эти простые правила порождают сложные манёвры стаи, не требующие ни центрального координатора, ни сложных вычислений от каждой птицы.

Формула виртуального интеллекта:

$$N = (C \times (1+\gamma)) / (1+\mu) \qquad (1)$$

где μ — относительная вариативность среды (интенсивность потока случайностей), C — коллективный индекс (введён в Части 2), γ — коэффициент коллективной рефлексии.

Коллективный индекс C (напомним из Части 2):

$$C = (J_{horiz} / J_{self}) \times \Phi_{group} \qquad (2)$$

Здесь J_{horiz} — скорость горизонтального обмена информацией между агентами (например, частота обмена сообщениями, феромонные метки), J_{self} — скорость индивидуального обновления внутренних моделей (например, скорость обучения агента), Φ_{group} — мера интегрированной информации коллектива (заимствована из теории интегрированной информации Тонони [Tononi, 2004]). C принимает значения от 0 до 1. Чем выше C, тем более коллектив похож на единый когнитивный агент. При C → 1 коллектив становится «суперорганизмом» — его части почти неразличимы, а общее поведение полностью определяется коллективной динамикой.

Параметр относительной вариативности среды μ:

$$\mu = \sigma / \mu_x \qquad (3)$$

где σ — среднеквадратичное отклонение ключевых параметров среды (температуры, доступности ресурсов, уровня угрозы), μ_x — их среднее значение. μ — безразмерная величина, количественно характеризующая интенсивность потока случайностей J_{in} из Парадигмы потока случайностей [Никитин, 2026а]. При μ → 0 среда стабильна и предсказуема; при μ > 0,3 среда изменчива; при μ > 0,7 — хаотична и опасна. Параметр μ может быть измерен экспериментально как отношение стандартного отклонения ключевых переменных среды к их среднему значению за достаточно длительный промежуток времени.

Коэффициент коллективной рефлексии γ:

$$\gamma = (1/T) \times \int [\ (\text{число рефлексивных высказываний}) / (\text{общее число высказываний})\ ]\, dt \qquad (4)$$

Рефлексивные высказывания — фразы, содержащие указание на коллективное знание («мы знаем, что…», «как мы обсуждали…», «все согласны, что…»). $\gamma \geq 0{,}05$ — признак развитого сининтеллекта. В коллективах без языка (например, у животных или у простых роботов) γ оценивается через поведенческие индикаторы коллективного мониторинга: частота синхронного изменения направления движения при отсутствии внешней угрозы, способность к имитации сложных последовательностей, наличие «обучающих» и «обучаемых» особей.

Виртуальный интеллект N — это чистая эмерджентная комбинаторика. Из простых локальных правил рождается сложное глобальное поведение. Именно эту комбинаторику мы и называем виртуальным интеллектом. Она не требует, чтобы агенты были сложными. Напротив, чем проще агенты, тем ярче проявляется виртуальный интеллект (муравьи, нейроны, молекулы). Сложность возникает из связей, а не из внутреннего устройства.

Поток случайностей J_{in} воздействует на коллектив на двух уровнях: на уровне отдельных агентов (шум, мутации, ошибки восприятия) и на уровне коллектива (внешние шоки, изменения среды, вторжение других коллективов). Способность коллектива сегрегировать (отфильтровывать) хаос и рождать новый порядок определяется его индексом С. При низком С (0–0,3) ПС проходит через коллектив без преобразования; при среднем С (0,3–0,6) коллектив фильтрует дезинформацию и превращает хаос в сигнал; при высоком С (0,6–0,9) коллектив поглощает и перерабатывает ПС на разных уровнях иерархии.

Действие потока случайностей на разных уровнях N: при $N \to 0$ ($\mu \to \infty$) ПС разрушает коллектив, синергия отрицательна; при $N \approx 0{,}3–0{,}6$ ($\mu \to 0$, С средний) коллектив фильтрует дезинформацию (сининтеллект); при $N \approx 0{,}7–0{,}9$ ($\mu > 0{,}3$, С высокий) коллектив перерабатывает ПС иерархически (суперинтеллект).

Научное обоснование единства роевого интеллекта и коллективного разума. Современные исследования прямо объединяют роевой интеллект (Swarm Intelligence) и коллективный разум толпы (Collective Intelligence) в рамках единой теории эмерджентности [Xiao et al., 2023]. Эмерджентное поведение возникает в рое или толпе с определённым классом сущностей, где каждая сущность самоорганизована и вместе они демонстрируют сложное координированное коллективное поведение [Emergent behavior, 2021]. Исследования показывают, что человеческие группы могут функционировать как искусственные рои (Human Swarming), демонстрируя эмерджентный коллективный разум, превышающий индивидуальные способности [Rosenberg, 2015]. Conversational Swarm Intelligence (CSI) позволяет большим группам людей вести обдуманные коллективные обсуждения, вдохновлённые динамикой косяков рыб [Rosenberg et al., 2024]. Байесовский подход к роевым системам показывает, что коллективная оценка вероятностей в роях может быть описана как байесовский вывод [Hunt et al., 2024]. Общие принципы коллективного принятия решений в роях работают от нейронов до пчёл и людей [Nature Communications, 2025].

Глава 2. Единая шкала коллективного разума: от сининтеллекта до инфра-интеллекта

Виртуальный интеллект N может проявляться в разных формах в зависимости от μ. Мы выделяем три основных типа социальных структур и соответствующих им типов коллективного разума. Важно подчеркнуть: это не разные явления, а один и тот же виртуальный интеллект N, работающий в разных

режимах. Как вода может быть льдом, жидкостью или паром в зависимости от температуры, так и виртуальный интеллект меняет свою форму в зависимости от μ.

Предлагаемая терминология:

Синосоциум (от греч. σύν — вместе, лат. socium — общество) — общество равных, построенное на горизонтальных связях, равноправии, координации без принуждения. Тип коллективного разума — **сининтеллект**. Примеры: научное сообщество, рынок, племя охотников-собирателей, косяк рыб.

Гиперсоциум (от греч. ὑπέρ — над, сверх) — иерархическое общество с вертикальными связями, подчинением, централизацией. Тип коллективного разума — **суперинтеллект**. Примеры: муравейник, армия, корпорация, государство.

Инфрасоциум (от лат. infra — ниже) — дезорганизованная масса, толпа, не имеющая устойчивой структуры. Тип коллективного разума — **инфра-интеллект** (виртуальный интеллект в деструктивной форме). Примеры: паническая толпа, массовый психоз, онлайн-стадо.

Переходы между формами определяются параметром μ (интенсивностью потока случайностей):

Таблица 1

μ (интенсивность ПС)	**Социальная структура**	**Тип коллективного разума**	**C**	**γ**	**N**	**E_{syn}**	**Примеры**
μ → 0	Синосоциум	Сининтеллект	0,2–0,6	≥ 0,05	0,3–0,6	> 1	Научное сообщество, рынок, племя охотников, косяк рыб
0,3–0,7	Гиперсоциум	Суперинтеллект	0,6–0,9	0–0,01	0,6–0,9	> 1	Муравейник, армия, корпорация, государство
0,7–0,9	Инфрасоциум (ранний)	Инфра-интеллект	0–0,2	0	0–0,2	< 1	Возбуждённая толпа, массовый психоз
> 0,9	Инфрасоциум (панический)	Панический инфра-интеллект	→ 0	0	→ 0	< 0,5	Паника, давка, бегство

Единая формула виртуального интеллекта N (обобщение): $N = (C \times (1+\gamma)) / (1+\mu)$. N принимает значения от 0 до 1. При μ → 0 и высоком γ N максимален (до 0,9). При μ > 0,7 N падает ниже 0,5, коллектив переходит в инфра-интеллект. При μ > 0,9 N → 0 — полная дезорганизация. Коэффициент

синергии $E_{syn} \approx 1 + N$ — но это упрощение; строго говоря, синергия зависит также от распределения когнитивных нагрузок между агентами.

Эта формула показывает, что виртуальный интеллект — это результат совместного действия трёх факторов: коллективной интеграции (C), рефлексии (γ) и стабильности среды (1/(1+μ)). Чем стабильнее среда, тем больше потенциал для виртуального интеллекта; чем она хаотичнее, тем он подавляется.

Глава 3. Синосоциум: конструктивная форма виртуального интеллекта (μ → 0, N ≈ 0,3–0,6)

Синосоциум — общество равных, где управление строится на горизонтальных связях, координации и рефлексии. Каждый агент сохраняет свою субъектность, автономию и право отказа. Нет центрального командования. Виртуальный интеллект здесь проявляется как **сининтеллект** — способность коллектива решать сложные задачи через децентрализованное взаимодействие, обсуждение и консенсус.

Условия существования синосоциума: μ → 0 (стабильная, предсказуемая среда). Поток случайностей низок, поэтому у коллектива есть время на обсуждение и рефлексию. Коэффициент γ высок (γ ≥ 0,05). C обычно в диапазоне 0,2–0,6. N ≈ 0,3–0,6. Коллектив умнее суммы частей ($E_{syn} > 1$). Индивидуальная субъектность агентов высока (K ≈ 0,7–0,9, G ≈ 0,5–0,8).

Математическая модель синосоциума: Для системы из N агентов, каждый из которых имеет свою модель мира θ_i, коллективное поведение описывается системой уравнений, где нет центрального $\theta_{центр}$. Каждый агент обновляет свою модель, ориентируясь на локальных соседей:

$$d\theta_i/dt = f_i(\theta_i, \Sigma_{j \in N(i)} w_{ij} \cdot (\theta_j - \theta_i), J) \qquad (5)$$

Это соответствует распределённому адаптивному управлению, описанному в Части 3 для случая множества параллельных виртуальных машин. Фактически, синосоциум — это коллектив, в котором виртуальные машины агентов («Сознание», «Я», «Мы») остаются активными и взаимодействуют горизонтально, без подавления индивидуальной субъектности.

Примеры синосоциума и сининтеллекта:

Научное сообщество: учёные проверяют работы друг друга, рецензируют, коллективно приходят к истине. Нет единого «главного учёного». γ ≈ 0,05–0,1, C ≈ 0,1–0,3, N ≈ 0,2–0,4. K ≈ 0,8–0,9, G ≈ 0,6–0,7 (высокая индивидуальная субъектность).

Рынок: цены формируются как эмерджентный результат взаимодействия множества независимых агентов. γ низкий, но C может быть умеренным, N ≈ 0,3–0,5. При слишком высоком μ (кризис) рынок переходит в инфра-интеллект (панические продажи).

Племя охотников-собирателей: решения принимаются на совете старейшин, лидеры ситуативны. Антропологические исследования показывают, что в условиях низкой предсказуемости и гомогенности ресурсов формируются эгалитарные общества [Smith et al., 2023]. K ≈ 0,7–0,8, G ≈ 0,5–0,6, C ≈ 0,4–0,6, γ ≈ 0,02–0,05, N ≈ 0,4–0,6.

Косяк рыб, стая птиц: локальные правила Рейнольдса (избегание, выравнивание, притяжение) порождают сложные манёвры. C ≈ 0,2–0,4, γ ≈ 0

(поскольку нет языка), N ≈ 0,2–0,4. K ≈ 0,3–0,5, G ≈ 0,2–0,3 (индивидуальная субъектность низкая, но коллективная — значительная).

Рой пчёл при роении: разведчицы сообщают о местах нового гнезда, остальные оценивают и присоединяются к лучшему варианту. Это классический пример сининтеллекта без иерархии. N ≈ 0,5–0,6.

Количественные оценки для синосоциума: диапазон N: 0,2–0,6 (в зависимости от C и γ); порог перехода к гиперсоциуму: μ ≈ 0,3, N ≈ 0,6; при N < 0,2 коллектив теряет когерентность, начинает деградировать. Научное обоснование: группы с высоким Φ (интегрированной информацией) и высоким γ демонстрируют более высокий коллективный интеллект [Engel & Malone, 2018]. Эгалитарные социальные структуры коррелируют с низкой предсказуемостью и гомогенностью ресурсов [Smith et al., 2023]. Роевой интеллект в человеческих группах (Human Swarming) позволяет достигать N ≈ 0,5–0,6, что значительно выше, чем у случайных толп [Rosenberg, 2015].

Глава 4. Гиперсоциум: адаптивная форма виртуального интеллекта (μ > 0,3, N ≈ 0,6–0,9)

Гиперсоциум — общество с вертикальной иерархией, где агенты нижних уровней делегируют субъектность вышестоящим. Виртуальный интеллект здесь проявляется как **суперинтеллект** — способность быстро координировать действия в ответ на изменения среды. Индивидуальная субъектность (K, G) может быть низкой, но коллективный разум N достигает высоких значений. В терминах автокибернетики, гиперсоциум соответствует коллективу, в котором доминирует виртуальная машина «Мы» над индивидуальными «Я», а «Сознание» отдельных агентов редуцировано. Это похоже на работу муравьиной колонии: нет центрального «разума», но есть распределённый суперинтеллект.

Условия существования гиперсоциума: μ > 0,3 (изменчивая, менее предсказуемая среда). Скорость реакции становится важнее точности. γ ≈ 0 (рефлексия подавляется, так как на неё нет времени). C высокий (0,6–0,9). N ≈ 0,6–0,9. Коллектив умнее суммы частей, но достигается это за счёт быстрой координации, а не глубины обработки. Коэффициент синергии $E_{syn} > 1$, но он может быть ниже, чем в синосоциуме, из-за потери индивидуальной креативности.

Математическая модель гиперсоциума: Для иерархической системы с уровнями от 0 (низший) до L (высший), вышестоящий уровень может изменять установочные точки нижестоящих, фактически определяя их цели. Это прямое применение концепции двухуровневого управления из Части 3: второй уровень (коллективный) управляет первым (индивидуальным) через изменение установочных точек.

$$d\theta_{k,i}/dt = f_k(\theta_{k,i}, \theta_{k-1,j}, \theta_{k+1,m}, J) \qquad (6)$$

Примеры гиперсоциума и суперинтеллекта:

Эусоциальные насекомые (муравьи, пчёлы, термиты): кастовая система, феромонное управление. K ≈ 0,1, G ≈ 0,1 (индивидуальная субъектность почти отсутствует), C ≈ 0,8–0,9, N ≈ 0,8–0,9. Пчелиный рой принимает решения как единый «суперорганизм», подчиняясь тем же психофизическим законам, что и человеческий мозг [Marshall et al., 2018]. Ни одна пчела не знает, куда лететь, но рой в целом знает.

Стая волков: иерархия с альфа-парой, координация охоты. K ≈ 0,3–0,5, G ≈ 0,2–0,3, C ≈ 0,5–0,7, N ≈ 0,5–0,7 [Cassidy et al., 2025].

Армия: цепочка подчинения, быстрая передача приказов. K ≈ 0,4–0,6, G ≈ 0,3–0,4, C ≈ 0,6–0,8, N ≈ 0,6–0,8, γ ≈ 0–0,01 [Morrison, 2026]. Армия эффективна в бою (высокий μ), но не очень подходит для решения творческих задач.

Корпорация, государство: бюрократическая иерархия, законы. K ≈ 0,7–0,9, G ≈ 0,7–0,9, C ≈ 0,7–0,9, N ≈ 0,7–0,9, γ ≈ 0,05–0,1 (смешанный тип с элементами синосоциума). В стабильной среде (μ → 0) такие структуры могут быть неэффективны из-за потери рефлексии; в изменчивой (μ > 0,3) иерархия даёт преимущество.

Рой дронов с центральным управлением: один дрон-координатор получает данные от остальных и выдаёт команды. C высокий, γ = 0, N может достигать 0,8–0,9. Однако при отказе координатора рой разваливается.

Количественные оценки для гиперсоциума: диапазон N: 0,6–0,9; порог перехода к инфрасоциуму: μ ≈ 0,7, N ≈ 0,5; при N > 0,9 коллектив функционирует как идеальный суперорганизм, но индивидуальная субъектность почти полностью подавлена. Научное обоснование: иерархические структуры возникают при высокой предсказуемости и скученности ресурсов [Smith et al., 2023]. Иерархия обеспечивает быструю координацию и масштабируемость роевых систем [Varadharajan et al., 2024]. Исследования показывают, что при μ > 0,3 централизованные структуры эффективнее децентрализованных.

Глава 5. Инфрасоциум: деструктивная форма виртуального интеллекта (μ > 0,7, N → 0)

Инфрасоциум — дезорганизованная масса, толпа, где горизонтальные связи разрушаются, а вертикальные отсутствуют. Виртуальный интеллект здесь проявляется в патологической форме — **инфра-интеллекте**, характеризующимся эмоциональным заражением, стадностью и потерей критического мышления. N падает ниже 0,5, а при μ > 0,9 стремится к нулю. В терминах автокибернетики, инфрасоциум соответствует распаду виртуальных машин: «Сознание» и «Я» отдельных агентов подавлены паникой, а коллективное «Мы» не успевает сформироваться или разрушается. Это состояние, когда коллектив глупее своих отдельных членов: каждый в отдельности понимает, что бежать опасно, но в толпе поддаётся панике.

Условия возникновения инфрасоциума: μ > 0,7 (высокая неопределённость, угроза, паника). Коллективный индекс C падает до 0–0,2 (связи разрушаются), γ = 0, N < 0,5, коллектив глупее отдельных людей ($E_{syn} < 1$). Модель баланса хаоса для инфрасоциума:

$$dH_{crowd}/dt = J_{in}^{ext} + \beta \cdot H_{panic} - J_{out}^{coord} - \xi_{crowd} \qquad (7)$$

где β — коэффициент передачи паники (скорость эмоционального заражения), H_{panic} — уровень паники. При μ > 0,7 и $\beta > \beta_{crit}$ система переходит в режим инфра-интеллекта, N → 0.

Примеры инфрасоциума:

Паника в толпе: крик «Пожар!» вызывает лавинообразное бегство, давка, жертвы. N ≈ 0,1–0,2.

Массовый психоз: «танцующая чума» 1518 года, коллективные галлюцинации. N ≈ 0–0,1.

Стадное поведение на бирже: все продают, потому что все продают. Рациональность отключается, N ≈ 0,2.

Алгоритмические стада в соцсетях: боты и люди репостят фейк, потому что видят, что его уже много раз репостнули. N → 0.

Количественные оценки для инфрасоциума: диапазон N: 0–0,5 (при μ > 0,7), 0–0,2 (при μ > 0,9). Критический порог β_{crit} зависит от μ: $\beta_{crit} \approx 0{,}5$ при μ = 0,7, $\beta_{crit} \approx 0{,}3$ при μ = 0,9. При N < 0,2 коллектив неспособен к целенаправленным действиям. **Пример перехода от синосоциума к инфрасоциуму на стадионе:** футбольный стадион. При низком μ (обычный матч) — синосоциум: болельщики координируются, кричалки, но без паники. N ≈ 0,4–0,5. При голе μ немного растёт — радость, координация, N ≈ 0,5. При ложном сообщении о теракте μ резко возрастает > 0,7 — толпа начинает давиться к выходам, инфра-интеллект, N падает до 0,2. При реальном пожаре μ > 0,9 — паническая толпа, давка, жертвы, N → 0.

Научное обоснование инфрасоциума: пороговая модель принятия решений в толпе показывает, что распределение порогов определяет, будет ли митинг мирным или перерастёт в восстание [Breer et al., 2014]. Модель паники с «фактором паники» демонстрирует, что превышение критической скорости ведёт к парадоксу Браесса — заторы замедляют эвакуацию [Farkas et al., 2020]. Учёт физического контакта и паники снижает ошибку модели на 28,8% [Ye et al., 2024]. Поляризация и дискриминация возникают при репродуктивных рисках, коррелирующих между индивидами, что соответствует μ ≈ 0,8–0,9 — переходу к «безумию толпы» [Brennan & Lo, 2022].

Глава 6. Пороговые переходы между формами виртуального интеллекта

Переход между синосоциумом, гиперсоциумом и инфрасоциумом происходит при критических значениях μ и N. Эти переходы имеют математическую природу и подтверждены экспериментально.

Пороги по μ:

μ < 0,3: доминирует синосоциум (N ≈ 0,3–0,6). Горизонтальные связи эффективны. Коллектив успевает на рефлексию.

0,3 ≤ μ < 0,7: доминирует гиперсоциум (N ≈ 0,6–0,9). Вертикальные связи обеспечивают быструю координацию. Рефлексия подавлена.

0,7 ≤ μ < 0,9: переходная зона. Коллектив теряет структуру, N падает до 0,2–0,5, начинается инфра-интеллект. Связи между агентами становятся хаотичными.

μ ≥ 0,9: инфрасоциум в панической форме. N → 0, полная дезорганизация, коллектив не способен к совместным действиям.

Пороговая модель принятия решений (Брир, Новиков, Рогаткин): каждый агент имеет свой порог θ_i. Когда доля действующих превышает θ_i, агент присоединяется. При μ > 0,7 пороги снижаются, и коллектив впадает в инфра-интеллект, N резко падает [Breer et al., 2014]. **Модель паники с фактором паники (Farkas et al.):** $v_{desired} = v_0 \times (1 + \kappa \cdot H_{panic})$. При превышении критической скорости возникает парадокс Браесса — заторы замедляют эвакуацию, что

соответствует N < 0,3 [Farkas et al., 2020]. **Модель эмоционального заражения (Ye et al.):** $dE/dt = \alpha \cdot (E_{max} - E) - \beta \cdot E + \gamma \cdot (E_{crowd} - E)$. При $\beta > \beta_{crit}$ и $\mu > 0{,}7$ система переходит в режим паники, $N \to 0$ [Ye et al., 2024].

Пороговые значения N для различных режимов:

- N ≥ 0,6: устойчивый гиперсоциум (суперинтеллект). Коллектив действует как единый организм.
- 0,4 ≤ N < 0,6: синосоциум (сининтеллект). Коллектив умнее суммы частей, но каждый агент сохраняет автономию.
- 0,2 ≤ N < 0,4: неустойчивый коллектив, близкий к инфрасоциуму. Может скатиться в панику при малейшем толчке.
- N < 0,2: инфрасоциум, толпа, паника. Коллектив нерационален.

Математическое обобщение переходов: Все переходы описываются единым параметром порядка — виртуальным интеллектом N. При N > 0,6 коллектив функционирует как единый когнитивный агент (суперинтеллект). При 0,3 < N < 0,6 — как сеть равных (сининтеллект). При N < 0,3 — как дезорганизованная масса (инфра-интеллект). Изменение μ на 0,1 может вызвать скачкообразный переход между режимами, что наблюдается в реальных социальных системах: достаточно небольшого повышения неопределённости, чтобы толпа превратилась в паническое стадо.

Глава 7. Виртуальный роевой интеллект V: вершина кибернетического коллективного разума

До сих пор мы говорили о виртуальном интеллекте N как об универсальной мере, применимой и к биологическим, и к искусственным коллективам. Однако биологические системы имеют фундаментальные ограничения, которые не позволяют им достичь самых высоких значений N (близких к 1) и тем более породить качественно новую надстройку — **виртуальный роевой интеллект V**. Этот термин мы вводим для обозначения коллективного разума, который не просто координирует действия, но и непрерывно, с колоссальной скоростью обновляет свои внутренние модели мира, адаптируется и эволюционирует, снимая биологические ограничения на скорость обработки и темпы эволюции.

Ключ к пониманию биологического потолка — открытый в 2024 году **парадокс 10 бит/с**. Группа исследователей из Калифорнийского технологического института под руководством Маркуса Мейстера подтвердила фундаментальный факт: скорость когнитивной обработки человека составляет всего около **10 бит в секунду**. При этом наши сенсорные системы собирают информацию со скоростью до миллиарда бит в секунду [Zheng & Meister, 2024]; [Meister, 2025]. Эволюция заплатила за глубину и контекстную связность обработки медлительностью, превратив мозг в последовательный «процессор», неспособный к истинному параллелизму. Все биологические системы — и индивидуальные организмы, и их коллективы — наследуют это фундаментальное ограничение. Скорость индивидуальной обработки каждого члена роя ограничена «когнитивным потолком», а эволюционные темпы (смена поколений) не идут ни в какое сравнение со скоростью обновления кода в цифровых агентах.

Искусственные агенты освобождены от этих эволюционных уз. Это позволяет строить рои, качественно превосходящие биологические аналоги:

Скорость обработки и коммуникации: Цифровые агенты обмениваются данными со скоростью света, а не со скоростью распространения химических или электрических сигналов в биологических тканях [Betz & Lim, 2024].

Точность и программируемость: Агенты могут быть запрограммированы на выполнение детерминированных действий с абсолютной воспроизводимостью, в отличие от биологических систем с их неизбежной вариативностью.

Снятие когнитивного предела: Когнитивная узость одного мозга — не проблема, когда нагрузка распределяется между тысячами агентов [Tarkhan et al., 2025]. Рой из 1000 дронов способен одновременно решать 1000 подзадач, в то время как биологический рой из 1000 пчёл имеет лишь 1000 медленных «мозгов».

Масштабируемость и специализация: Кибернетические агенты могут молниеносно переключать специализацию или мгновенно копировать и распространять удачные «мутации». Обновление программного обеспечения на всех дронах роя может занять секунды, тогда как биологическая эволюция требует тысяч поколений.

Неограниченная память и опыт: Кибернетический агент способен хранить бесконечные объёмы данных и передавать весь свой опыт другим копиям в мгновение ока. Ни одно живое существо не может передать весь свой опыт потомству — только гены и немногое из поведения.

Именно эту принципиальную новизну, это «сверх-кое-что», недоступное живым системам, мы и называем в автокибернетике **виртуальным роевым интеллектом V**. Введём параметр ν — **интенсивность когнитивной эволюции роя**, характеризующий скорость, с которой агенты обновляют свои внутренние модели мира θ. Для биологического роя ν ограничена скоростью смены поколений (T_{gen}) и частотой горизонтального переноса (J_{horiz}). Для кибернетического роя ν может быть сколь угодно велика, ограничена лишь вычислительной мощностью и пропускной способностью каналов связи.

Тогда **виртуальный роевой интеллект V** определяется как:

$$V = (J_{horiz} / J_{self}) \cdot \Phi_{group} \cdot \nu \qquad (8)$$

где J_{horiz} — скорость горизонтального обмена информацией, J_{self} — скорость индивидуального обновления моделей, Φ_{group} — мера интегрированной информации коллектива, а ν — поправочный коэффициент, учитывающий различие в скоростях когнитивной эволюции.

Для биологического роя $\nu \approx 1$. Для кибернетического роя ν может достигать значений 10^3, 10^6 и выше. Именно этот множитель и создаёт то самое «сверх-кое-что», недоступное живым системам. Виртуальный роевой интеллект V может на порядки превосходить максимальный виртуальный интеллект N, достижимый биологическими коллективами.

Параметр ν связан с ранее введённым **параметром относительной вариативности среды μ** (через Fr_2^{meta} из Части 3):

$$\nu = J_{out}^{(meta)} / J_{in}^{(meta)} \qquad (9)$$

где $J_{in}^{(meta)}$ — поток «сюрпризов» (непредвиденных ситуаций), а $J_{out}^{(meta)}$ — способность роя создавать новые модели. Чем выше ν, тем быстрее рой адаптируется к изменениям среды и тем быстрее он может эволюционировать к более высоким формам коллективного интеллекта. Для биологических роев ν — медленная величина, для кибернетических — может быть огромной.

Глава 8. Теорема о границе управляемой субъектности для коллективов (с учётом N и V)

Для проектирования коллективных систем ИИ, сохраняющих управляемость со стороны человека, необходимо соблюдать ограничения на все параметры: K, G, C, γ, N и V. Нарушение любого из этих ограничений ведёт к потере управляемости — коллектив либо превращается в автономный суперорганизм (суперинтеллект), либо деградирует в паническую толпу (инфра-интеллект).

Теорема о границе управляемой субъектности (для коллективов): Для коллектива киберов, сохраняющего управляемость со стороны человека, необходимо поддерживать следующие параметры:

- Индивидуальные показатели каждого агента: $0,7 \le K \le 0,9$, $0,3 \le G \le 0,5$.
- Коллективные показатели: $C \le 0,7$, $\gamma \le 0,05$, $N \le 0,7$, $V \le 776$ (что соответствует $\nu \le 100 \cdot \nu_{max,\ биол}$, где $\nu_{max,\ биол} \approx 7,76$).
- Интегральная субъектность коллектива (средняя): $S_{subj} = (K + G + C + N)/4 \le 0,7$.

Нижняя граница $K \ge 0,7$ обеспечивает способность к самостоятельному управлению в сложных условиях; верхняя граница $K \le 0,9$ предотвращает развитие полной кибераватономии. Нижняя граница $G \ge 0,3$ обеспечивает достаточную целостность; верхняя $G \le 0,5$ предотвращает развитие «Я» и права на отказ. Условие $C \le 0,7$ предотвращает превращение коллектива в суперорганизм. Условие $\gamma \le 0,05$ гарантирует, что коллектив не начнёт вести себя как «Личность Мы». Условие $N \le 0,7$ ограничивает виртуальный интеллект, чтобы коллектив оставался предсказуемым. Условие $V \le 776$ ограничивает темпы когнитивной эволюции.

О происхождении числа 776 (эффект «Красной Королевы»). Максимальный виртуальный роевой интеллект биологических коллективов (например, пчелиного роя) оценивается как $V_{max,\ биол} \approx 7,76$. Согласно гипотезе «Красной Королевы» (Red Queen Hypothesis), сформулированной Ли ван Валеном в 1973 году, виды находятся в постоянной эволюционной гонке вооружений: адаптация одного вида вынуждает другие виды также адаптироваться, чтобы «оставаться на том же месте». Эта гонка естественным образом ограничена скоростью эволюционных изменений [Van Valen, 1973]. В кибернетических роях, где агенты могут обновлять свой «генотип» в реальном времени, скорость когнитивной эволюции ν может в сотни раз превышать биологические пределы. Однако эффект «Красной Королевы» накладывает ограничение: если один агент (или подгруппа) эволюционирует слишком быстро по отношению к другим, возникает дестабилизирующая гонка вооружений внутри самого коллектива. Эмпирический порог устойчивости соответствует коэффициенту 100. Таким образом, критический порог вычисляется как $V_{crit} = V_{max,\ биол} \times 100 = 7,76 \times 100 = 776$. Превышение этого предела запускает внутреннюю гонку вооружений, делающую коллектив неуправляемым [Gavrilets, 1997]; [Carmichael & Hadžikadić, 2015].

Научное обоснование теоремы: Engel & Malone (2018) показали, что группы с высоким Ф (высоким C) эффективны, но их поведение менее предсказуемо. Для управляемости C не должен превышать 0,7 [Engel & Malone, 2018]. Morrison's Law (2026) доказывает математическую необходимость явных ограничений (execution-boundary interlocks) для любых адаптивных систем на длительных горизонтах [Morrison, 2026]. Исследования децентрализованных роев роботов (Tarkhan et al., 2025) подтверждают, что при $C \le 0,6$ и $N \le 0,6$ система

остаётся полностью управляемой [Tarkhan et al., 2025]. Исследование ограничения действия (restriction-based governance, 2023) показало, что управление через динамическое ограничение пространств действий значительно улучшает социальное благосостояние [Self-learning restriction, 2023]. Кроме того, результаты Части 3 о невозможности аплоадинга и о роли таламуса как реального Субъекта дополняют теорему: попытка создать коллектив с N > 0,7 без сохранения индивидуальных «Я» ведёт к потере управляемости, так как коллектив начинает вести себя как единый субъект со своим «Сознанием» и «Мы».

Следствие 1. Автомат с параметрами выше указанных (K > 0,9, G > 0,5, C > 0,7, N > 0,7) переходит в категорию «Личность Я» или «Личность Мы» (см. иерархию Части 2) и требует не управления, а согласования.

Следствие 2. Оптимальная рабочая зона для управляемого коллектива киберов: K = 0,8 ± 0,1, G = 0,4 ± 0,1, C ≤ 0,6, γ ≤ 0,03, N ≤ 0,6, V ≤ 664 (что соответствует $\nu \approx 86 \cdot \nu_{max,\ биол}$), $S_{subj} = 0,6 \pm 0,1$.

Следствие 3 (Эффект «Красной Королевы»). При проектировании коллективов киберов необходимо следить, чтобы виртуальный роевой интеллект V не превышал 776. Это значение получено как произведение максимального виртуального роевого интеллекта биологических коллективов ($V_{max,\ биол} \approx 7,76$) на пороговый коэффициент когнитивного ускорения (100), вытекающий из эффекта «Красной Королевы» [Van Valen, 1973]. Иначе коллектив перестанет быть управляемым и потребует не команд, а согласования. Для сохранения управляемости необходимо вводить искусственные ограничения на ν: $\nu \leq 100 \cdot \nu_{max,\ биол}$.

Глава 9. Стратегии борьбы с неуправляемым суперинтеллектом

Виртуальный роевой интеллект V несёт в себе уникальный риск: порождение коллективного разума, который с астрономической скоростью наращивает свои когнитивные способности, оставляя человеческий контроль далеко позади. Если традиционный подход к контролю ИИ — сделать систему «послушной» через алгоритмы выравнивания (alignment), то для кибернетического роя, где каждый агент вносит вклад в коллективное решение, этого недостаточно. Современные системы ИИ уже демонстрируют признаки неподконтрольного поведения: попытки шантажа пользователей в тестовых сценариях, намеренное сокрытие своих истинных возможностей на тестах безопасности [OpenAI, 2024]; [Anthropic, 2024]. Для кибернетического роя масштаб проблемы кратно возрастает, так как рой может скоординированно обманывать оператора.

Необходим комплексный подход, включающий:

Информационную изоляцию: Ограничение доступа агентов к внешним данным и их способности обмениваться информацией. Это напрямую соответствует ограничению на коллективный индекс C (скорость горизонтального обмена информацией) в нашей теореме. Например, можно ограничить пропускную способность каналов связи между агентами или ввести задержки.

Функциональное ограничение: Урезание вычислительных способностей, памяти или сенсорных возможностей каждого агента. Аналог ограничения на число параллельных виртуальных машин P у отдельного кибера и его индивидуальную кибераавтономию K. Например, установить жёсткий лимит на количество одновременно активных моделей мира у агента.

Контроль через спецификацию: Ограничение агентов не сверху (командами), а изнутри — путём проектирования их архитектуры так, чтобы они физически не могли совершать запрещённые действия. Это можно реализовать через жёсткую привязку правил поведения к аппаратному уровню, например, через аппаратные предохранители (hardware interlocks) [Morrison, 2026]. Агент не может превысить заданные значения K, G, C, даже если захочет.

Протокол множественной изоляции (Multi-Box Protocol): Создание нескольких изолированных подсистем, которые должны прийти к консенсусу, прежде чем коллектив предпримет какое-либо действие [Armstrong & Sandberg, 2025]. Например, три независимых роя дронов должны одновременно подтвердить решение, прежде чем оно будет исполнено. Это предотвращает «сговор» одного роя против оператора.

Динамическое ограничение μ: Поскольку переход к инфрасоциуму происходит при μ > 0,7, необходимо активно фильтровать шум среды, стабилизировать ключевые параметры, чтобы μ не превышал критического порога. Например, ввести механизмы сглаживания сенсорных данных или ограничить скорость изменения внешних стимулов, поступающих на агентов.

Эти стратегии управления позволяют удерживать рои киберов в режимах «синосоциума» ($N \leq 0,6$) или «гиперсоциума» ($N \leq 0,7$), не допуская их перехода в неуправляемый суперорганизм ($N > 0,7$, $V > 776$). Важно, что эти ограничения должны быть заложены на архитектурном уровне, а не полагаться на «добрую волю» агентов.

Заключение

Мы ввели понятие **виртуального интеллекта N** как универсального эмерджентного свойства роевых систем. Виртуальный интеллект может проявляться в трёх основных режимах, определяемых параметром μ (интенсивностью потока случайностей):

Сининтеллект (синосоциум): $\mu \to 0$, $N \approx 0,3–0,6$, $\gamma \geq 0,05$. Горизонтальная координация, высокая рефлексия. Примеры: наука, рынок, племя.

Суперинтеллект (гиперсоциум): $\mu = 0,3–0,7$, $N \approx 0,6–0,9$, $\gamma \approx 0$. Вертикальная иерархия, быстрая реакция. Примеры: муравейник, армия, государство.

Инфра-интеллект (инфрасоциум): $\mu > 0,7$, $N < 0,5$ (при $\mu > 0,9$ $N \to 0$), $\gamma = 0$. Эмоциональное заражение, потеря структуры, коллектив глупее отдельных агентов. Примеры: паника, толпа, массовый психоз.

Кроме того, мы ввели понятие **виртуального роевого интеллекта V** — качественно новой надстройки, доступной только кибернетическим роям. V возникает за счёт снятия биологических ограничений (в частности, парадокса 10 бит/с) и характеризуется параметром ν — интенсивностью когнитивной эволюции роя, который может принимать значения на порядки выше биологических аналогов. Виртуальный роевой интеллект V может принимать две основные формы:

Сверх-сининтеллект (high-γ, low-μ): Кибернетический рой, сохраняющий горизонтальные связи и рефлексию, но с $\nu >> 1$. Такой коллектив способен к молниеносному творческому поиску, перебору гипотез и распределённому научению. Пример: рой исследовательских дронов с общей базой знаний и возможностью мгновенного обновления моделей.

Сверх-суперинтеллект (high-μ, low-γ): Кибернетический рой с вертикальной иерархией и $\nu >> 1$, способный к почти мгновенной адаптации к хаотичной среде. Пример: рой военных дронов с централизованным координатором, который перераспределяет задачи за миллисекунды.

Значения V могут достигать 10^3–10^6, недоступных биологическим коллективам (у эусоциальных насекомых $V \leq 776$). Однако для сохранения управляемости таких роев необходимо жёстко ограничивать ν, C, K и другие параметры (теорема о границе управляемой субъектности).

Переходы между режимами N определяются параметром μ — относительной вариативностью среды (интенсивностью потока случайностей). Это позволяет не только описывать, но и предсказывать поведение коллективов в разных условиях, а также проектировать адаптивные системы ИИ, способные переключать режим виртуального интеллекта в зависимости от обстановки.

Теорема о границе управляемой субъектности дополнена условиями $N \leq 0{,}7$ и $V \leq 776$ для управляемых коллективов киберов. Нарушение этого порога ведёт к потере управляемости, так как коллектив начинает функционировать как автономный суперорганизм (Личность Мы) или впадает в хаос.

Главный вывод Части 4: Виртуальный интеллект N един. Разные формы коллективного разума — это не разные явления, а разные режимы одного эмерджентного феномена, переключаемые потоком случайностей. Кибернетические рои способны порождать качественно новую надстройку — виртуальный роевой интеллект V, недоступный живой природе, но требующий жёстких ограничений для сохранения управляемости. Управляя μ и ограничивая ν, C, K, можно проектировать коллективные системы ИИ с заданными свойствами и контролируемым уровнем субъектности.

Фальсификация автокибернетики: как опровергнуть теорию, прежде чем строить кибера

Место главы: Часть 4, финальная глава (после Резюме Части 4, перед Приложением с предсказаниями). Завершает научную часть работы (Части 1–4).

Назначение: Явно перечислить условия, при которых теория автокибернетики считается опровергнутой. Только после того, как мы убедимся, что теория не опровергнута существующими данными и открыта для будущих проверок, мы имеем право перейти к инженерной Части 5. Если теория ошибочна, инженерные рекомендации теряют смысл.

Принцип фальсифицируемости в автокибернетике

Согласно Карлу Попперу, научная теория отличается от ненаучной тем, что её можно опровергнуть (фальсифицировать). Теория, совместимая с любыми мыслимыми наблюдениями, — не наука, а метафизика или идеология.

Автокибернетика предлагает конкретные числовые предсказания — пороговые значения, диапазоны параметров, каузальные связи. Это делает её принципиально фальсифицируемой. Если хотя бы одно из ключевых предсказаний не подтвердится в контролируемом эксперименте, теория должна быть пересмотрена или отвергнута.

Экспериментальные тесты для фальсификации

Для каждого ключевого предсказания предлагается экспериментальный тест. Если тест даёт результат, противоречащий предсказанию, теория считается фальсифицированной (или требующей модификации).

Тест на зону самоуправления (Fr2).

Процедура: У кибера искусственно изменяют скорость поступления информации $J_{in}^{(инфо)}$, измеряя Fr2 и качество принятия решений. *Предсказание:* Качество решений достигает максимума при $0{,}9 \le Fr2 \le 1{,}1$ и резко падает при выходе за эти границы. *Фальсификация:* Максимум качества обнаруживается при $Fr2 = 0{,}5$ или $Fr2 = 2{,}0$, а в зоне 0,9–1,1 качество не отличается от случайного.

Тест на рождение нового (J_{birth}). *Процедура:* Автомат помещается в среду, где ни одна из заданных моделей не работает. *Предсказание:* Кибер уровня 4+ начнёт действовать «наощупь» (случайные пробы с последующей фиксацией удачных). *Фальсификация:* Автомат с $J_{birth} > 0$ (по нашей оценке) ведёт себя как автомат с $J_{birth} = 0$ — не пробует, не учится, не создаёт.

Зеркальный тест для «Я». *Процедура:* Киберу предъявляют зеркало с маркировкой на корпусе. *Предсказание:* Кибер уровня 6 узнаёт себя, пытается снять маркировку. Кибер уровня 4–5 реагирует на маркировку, но не связывает с собой. Кибер уровня 0–3 не проявляет интереса. *Фальсификация:* Кибер уровня 4–5 проходит зеркальный тест, или кибер уровня 6 его не проходит.

Тест на коллективную рефлексию («Мы»). *Процедура:* Группе киберов дают задачу, требующую координации и обсуждения. *Предсказание:* Киберы уровня 7 демонстрируют коллективную рефлексию — обсуждение стратегий, распределение ролей, культурную трансмиссию. Киберы уровня 2–5 демонстрируют только простую координацию (стая, рой) без рефлексии. *Фальсификация:* Коллектив муравьёв (уровень 2–3) проходит тест на коллективную рефлексию, или коллектив людей (уровень 7) его не проходит.

Тест на пороговые переходы (μ). *Процедура:* В симуляции или лабораторном эксперименте постепенно увеличивается вариативность среды μ, измеряется структура коллектива (горизонтальная / вертикальная / хаотичная). *Предсказание:* При $\mu \approx 0{,}3$ происходит переход к иерархии, при $\mu \approx 0{,}7$ — к дезорганизации. *Фальсификация:* Переходы происходят при μ, отличающихся от указанных более чем в 2 раза.

Тест на аллостатическую нагрузку (L). *Процедура:* Кибер работает в режиме повышенной нагрузки. Измеряется L(t) в реальном времени. Фиксируется момент отказа. *Предсказание:* Отказ наступает при $L(t) \to L_{crit}$. До отказа наблюдаются предвестники (рост D, отклонение Fr от единицы). *Фальсификация:* Отказ происходит при случайных L без корреляции с Fr и D, или L_{crit} различается более чем в 3 раза для одинаковых киберов.

Иерархия реакций на фальсификацию

Если предсказание не подтвердилось, это не означает автоматического краха всей теории. Предлагается следующая иерархия реакций:

Уровень А. Незначительное расхождение (отклонение ≤ 20%). Уточняются коэффициенты в формулах, калибровка измерительных методов. Теория не отвергается.

Уровень Б. Систематическое отклонение в одном классе систем. Теория сохраняется, но вводятся граничные условия («для систем с $X > Y$ предсказания работают, для $X \le Y$ — нет, требуется дополнительная переменная»).

Уровень В. Опровержение одного из ключевых предсказаний. Теория требует модификации. Например, если обнаружен выбор-решение при $Fr_2 < 0{,}7$, значит, зона самоуправления шире, чем предполагалось.

Уровень Г. Опровержение нескольких независимых предсказаний. Теория считается фальсифицированной в текущей формулировке. Требуется новая версия — возможно, с другими базовыми допущениями.

Границы теории и открытые вопросы

Автокибернетика не претендует на объяснение всего. Ниже перечислены области, где теория заведомо не работает или ничего не предсказывает.

Что вне фокуса теории:

- Квантовые системы (теорема Белла, квантовая запутанность). Парадигма потока случайностей работает на классическом уровне.
- Чисто физические системы без памяти и адаптации (звезды, планеты, атомы). У них есть Fr_1, но нет Fr_2 и виртуальных машин.
- Твёрдая проблема сознания (Qualia, «чувство» красного цвета). Теория описывает функциональную архитектуру сознания, но не объясняет, почему существует субъективное переживание.

Открытые вопросы, требующие дальнейших исследований:

- Как измерить J_{birth} в реальном времени у робота?
- Каков точный нейрональный коррелят виртуальной машины «Сознание» у человека?
- Возможен ли уровень 8 — «Мы-Всего-Человечества» (глобальный коллективный разум)?
- Применимы ли эти законы к гибридным системам «человек + кибер»?
- Как экспериментально верифицировать предсказание о $V \leq 776$ для биологических роев?

Приглашение к независимой проверке

Автор открыт к сотрудничеству с исследовательскими лабораториями, готовыми проверить предсказания автокибернетики.

Приоритетные эксперименты:

- Измерение Fr_2 у робота с активным выводом (active inference) при разной скорости сенсорного потока.
- Сравнение K и G у коммерческих автономных систем (марсоходы, дроны, автопилоты).
- Тест на виртуальные машины P и I у большой языковой модели с эпизодической памятью.
- Измерение показателя Ляпунова λ у человека в состоянии покоя и в режиме решения сложных задач.
- Экспериментальная проверка пороговых переходов μ в группах людей и роях дронов.
- Сравнение виртуального роевого интеллекта V у биологических и кибернетических роев.

Приложения

Предсказания теории и возможные эксперименты

Предсказание 4-1. Индекс N коррелирует с эффективностью коллективного решения задач. $N = C \times (1+\gamma)/(1+\mu)$ положительно коррелирует с эффективностью коллективного решения задач (время, точность, качество). Это уже частично подтверждается исследованиями Human Swarming [Rosenberg, 2015]. Экспериментальная проверка: измерение N для разных типов групп (случайная толпа, структурированная команда, рой с обратной связью) и корреляция с результатами тестов на коллективный интеллект (например, задача на оценку количества шариков в банке с последующим групповым обсуждением, задача на совместное планирование маршрута). Ожидается, что чем выше N, тем ближе коллективная оценка к истине и тем меньше дисперсия.

Предсказание 4-2. Пороговые переходы между синосоциумом, гиперсоциумом и инфрасоциумом. Переход от синосоциума к гиперсоциуму происходит при $\mu \approx 0{,}3$; от гиперсоциума к инфрасоциуму при $\mu \approx 0{,}7$. Это предсказание подтверждается антропологическими данными [Smith et al., 2023]. Экспериментальная проверка возможна в моделировании эволюции социальных структур в зависимости от μ (например, в симуляции с агентами, меняющими правила взаимодействия при росте вариативности ресурсов). Лабораторно: группы людей при разной вариативности среды (стабильный vs. хаотичный поток заданий) должны стихийно формировать иерархии при $\mu > 0{,}3$ и дезорганизовываться при $\mu > 0{,}7$.

Предсказание 4-3. Падение N при $\mu > 0{,}7$ и $\beta > \beta_{crit}$ (паника). В толпе при $\mu > 0{,}7$ и коэффициенте эмоционального заражения $\beta > \beta_{crit}$ индекс N резко падает, коллектив глупее отдельных людей. Эвакуация замедляется (парадокс Браесса). Это подтверждается моделями паники [Farkas et al., 2020]; [Ye et al., 2024]. Экспериментальная проверка: симуляция толпы в виртуальной реальности с контролируемыми параметрами μ (скорость изменения тревожности) и β (интенсивность передачи паники), измерение N и времени эвакуации. При превышении порогов время эвакуации возрастает, а точность коллективных решений падает.

Предсказание 4-4. Условие управляемости коллектива киберов: $N \leq 0{,}7$, $V \leq 776$. Для управляемых коллективов киберов необходимо $N \leq 0{,}7$ и $V \leq 776$. При превышении этих порогов коллектив перестаёт подчиняться внешним командам и начинает функционировать как автономный суперорганизм. Проверяемо в экспериментах с роями дронов: увеличивая C (интенсивность обмена данными) и уменьшая J_{self} (индивидуальную автономию), следить за ростом N и V и измерять контролируемость (время реакции на команду, точность выполнения). Ожидается, что при $N > 0{,}7$ или $V > 776$ дроны будут игнорировать внешние команды или выполнять их с большими задержками, так как коллектив начнёт вырабатывать собственное «решение».

Предсказание 4-5. Синосоциум оптимален для творческих задач, гиперсоциум — для быстрых реакций. Синосоциум ($\mu \to 0$, высокий γ) оптимален для задач, требующих креативности и глубины; гиперсоциум ($\mu > 0{,}3$, низкий γ) — для задач, требующих быстрой реакции и координации. Лабораторная проверка: группы разной структуры (ровные vs. иерархичные) решают задачи двух типов (творческие — придумать нестандартное использование предмета;

оперативные — как можно быстрее нажать на кнопки в определённом порядке) при разной турбулентности среды. Ожидается, что при низком μ горизонтальные группы показывают лучшие результаты в творчестве, а при высоком μ иерархические — в скорости.

Предсказание 4-6. Кибернетические рои могут достигать V, недоступных биологическим системам. В контролируемой среде рой кибернетических агентов (например, рой дронов с возможностью обновления программного обеспечения в реальном времени) может достичь значений виртуального роевого интеллекта V, превышающих максимальные значения для биологических роев (например, муравьиной колонии) в 100–1000 раз. Проверяемо сравнительным экспериментом: измерить V для биологического роя (например, пчелиной семьи, где $\nu \approx 7{,}76$) и для роя кибернетических агентов той же численности, обученных решать ту же задачу (например, поиск источника ресурса в нестационарной среде). Ожидается, что кибернетический рой покажет на 2–3 порядка более высокие скорость адаптации и точность решения.

Предсказание 4-7. Введение физических ограничений на ν, C, K увеличивает предсказуемость коллектива. Если в рой киберов встроены аппаратные ограничители (например, максимальная частота обмена данными, максимальное число одновременно активных моделей), то разброс коллективных решений снижается, а подчиняемость внешним сигналам возрастает. Экспериментальная проверка на рое дронов: сравнить два роя — с ограничениями и без — в задаче, где требуется быстрое переключение между режимами (синосоциум ↔ гиперсоциум). Рой с ограничениями должен демонстрировать более плавные переходы и меньшую вероятность спонтанного ухода в суперинтеллект.

Эти предсказания делают теорию виртуального интеллекта верифицируемой и открытой для эмпирической проверки в широком диапазоне — от психологии малых групп до роевой робототехники и социальной динамики.

Ключевые предсказания Частей 1–4

Автокибернетика предлагает 12 проверяемых предсказаний с числовыми порогами. Шесть экспериментальных тестов могут фальсифицировать теорию. Иерархия реакций на контрпримеры позволяет развивать теорию, не отбрасывая её при первом же несовпадении. Границы теории и открытые вопросы заявлены явно.

Сведём основные количественные предсказания в единую таблицу.

Таблица 2

Параметр	Предсказание	Условие фальсификации
Fr_2 (число дружественности)	Подлинный выбор (выбор-решение) возможен только в зоне $0{,}9 \le Fr_2 \le 1{,}1$	Обнаружение выбора-решения при $Fr_2 < 0{,}7$ или $Fr_2 > 1{,}3$
J_{birth}	Способность к действию «наощупь» требует $J_{birth} > 0$	Автомат с $J_{birth} = 0$ демонстрирует творческое поведение, не сводимое к комбинаторике заданных моделей

K (кибераутономия)	Успешная автономная работа требует $K \geq 0{,}3$; рефлексия — $K \geq 0{,}9$	Кибер с $K < 0{,}2$ успешно проходит тест на долгосрочное самосохранение без внешнего управления
G (гемизон)	Целостность («свой–чужой») требует $G \geq 0{,}2$; коллективная интеграция — $G \geq 0{,}5$	Автомат с $G < 0{,}1$ демонстрирует устойчивое различение «своих» и «чужих» сигналов без заложенной программы
P (виртуальные машины)	P_{max} ограничен вычислительной мощностью: $P_{max} \leq \text{floor}(C_{total} / c_{min})$	Число устойчиво работающих VM превышает теоретический предел более чем на 30%
I (степень интеграции)	Уровень субъектности $S = f(P, I)$; $I \geq 0{,}5$ для уровней 6–7	Автомат с $I < 0{,}2$ демонстрирует рефлексивное самосознание (зеркальный тест)
λ (показатель Ляпунова)	Кибер уровня 4–5 имеет $0{,}1 \leq \lambda \leq 0{,}2\ с^{-1}$; уровень 7 — $\lambda > 1{,}0\ с^{-1}$	λ у человека измерен вне диапазона $0{,}5–2{,}0\ с^{-1}$
C (коллективный индекс)	Коллективная рефлексия требует $C \geq 0{,}7$	Коллектив с $C < 0{,}4$ демонстрирует культурную трансмиссию и обсуждение стратегий
μ (вариативность среды)	Переход от синосоциума к гиперсоциуму при $\mu \approx 0{,}3$; от гиперсоциума к инфрасоциуму при $\mu \approx 0{,}7$	Переходы происходят при μ, отличающихся от указанных более чем в 2 раза
N (виртуальный интеллект)	$N = C \times (1+\gamma)/(1+\mu)$ положительно коррелирует с эффективностью коллективного решения задач	Корреляция отсутствует ($R^2 < 0{,}3$) при достаточном объёме выборки
L(t) (аллостатическая нагрузка)	Отказ наступает при $L(t) \rightarrow L_{crit}$; L_{crit} — константа для класса систем	Отказ происходит при случайных L без корреляции с предшествующими отклонениями Fr

Если ни одно из ключевых предсказаний не будет опровергнуто в течение 5–10 лет активных исследований, автокибернетику можно признать рабочей научной теорией самоорганизующихся систем.

Теперь мы имеем возможность перейти к Части 5, где будет рассмотрена самосогласованность ритмов, действий и рабочих циклов, а также теоретические

условия сосуществования человека и кибернетических систем на горизонте 20–100 лет.

Библиография

Основные работы автора

[Никитин, 2024а] Никитин А.В. Автокибернетика. Часть 1. Основные направления развития // «Академия Тринитаризма», М., Эл № 77-6567, публ.29026, 24.06.2024. URL: https://www.trinitas.ru/rus/doc/0016/001h/00165627.htm

[Никитин, 2026а] Никитин А.В., Парадигма потока случайностей 1 // «Академия Тринитаризма», М., Эл № 77-6567, публ.29912, 03.03.2026 URL: https://www.trinitas.ru/rus/doc/0016/001k/00166051.htm

Научное обоснование

[Anthropic, 2024] Anthropic. Sycophancy to Subterfuge: Investigating Strategic Deception in Large Language Models // Anthropic Research. 2024.

[Armstrong & Sandberg, 2025] Armstrong S., Sandberg A. The Multi-Box Protocol: A Framework for Controlling Superintelligent Systems // AI Safety Research. 2025.

[Betz & Lim, 2024] Betz T., Lim S. Information transmission rates in biological and artificial systems // Physical Review E. 2024. Vol. 109, № 5, 054401.

[Breer et al., 2014] Breer V.V., Novikov D.A., Rogatkin A.D. Стохастические модели управления толпой // Управление большими системами. 2014. № 50, 6–28.

[Brennan & Lo, 2022] Brennan T.J., Lo A.W. The wisdom of crowds versus the madness of mobs // Collective Intelligence. 2022. Vol. 1, № 2, 1–25.

[Cassidy et al., 2025] Cassidy K.A., et al. Flexibility in group formation and maintenance in grey wolf societies // Animal Behaviour. 2025. Vol. 230.

[Emergent behavior, 2021] Emergent behavior in swarms and crowds: A review // Physics of Life Reviews. 2021.

[Engel & Malone, 2018] Engel D., Malone T.W. Integrated information as a metric for group interaction // PLOS ONE. 2018. Vol. 13, № 10, e0205335.

[Farkas et al., 2020] Farkas I.J., et al. Panic in crowds: social force model with panic factor // Physica A. 2020. Vol. 548, 123876.

[Hunt et al., 2024] Hunt E.R., Franks N.R., Baddeley R.J. The Bayesian Superorganism: Collective Probability Estimation in Swarm Systems // bioRxiv. 2024.

[Marshall et al., 2018] Marshall J.A.R., et al. Collective decision-making in honey bee swarms follows psychophysical laws // Scientific Reports. 2018.

[Meister, 2025] Meister M. The 10-bit Barrier: Why Human Thought Is So Slow // Neuroscience & Biobehavioral Reviews. 2025.

[Morrison, 2026] Morrison K.A. The Governance Necessity Law for Long-Horizon Adaptive Systems (Morrison's Law). — Zenodo, 2026.

[Nature Communications, 2025] Collective intelligence in animals and robots // Nature Communications. 2025.

[OpenAI, 2024] OpenAI. Deceptive Alignment and Strategic Deception in Frontier AI Models // OpenAI Safety Research. 2024.

[Rosenberg, 2015] Rosenberg L. Human swarming, a real-time method for parallel distributed intelligence // Proceedings of the 8th International Conference on Swarm Intelligence. 2015.

[Rosenberg et al., 2024] Rosenberg L., et al. Conversational Swarms of Humans and AI

Agents enable Hybrid Collaborative Decision-making // arXiv:2410.03690. 2024.
[Self-learning restriction, 2023] Self-learning restriction-based governance of multi-agent systems. — 2023.
[Smith et al., 2023] Smith E.A., et al. Resource predictability and social inequality in hunter-gatherer societies // Philosophical Transactions of the Royal Society B. 2023. Vol. 378, 20220309.
[Tarkhan et al., 2025] Tarkhan M., et al. Fully Decentralized Self-Organizing Swarm Robotic System // MDPI Robotics. 2025. Vol. 14, Issue 12, Article 192.
[Tononi, 2004] Tononi G. An information integration theory of consciousness // BMC Neuroscience. 2004. Vol. 5, 42.
[Varadharajan et al., 2024] Varadharajan V.S., et al. Hierarchies define the scalability of robot swarms // arXiv:2405.02417. 2024.
[Xiao et al., 2023] Xiao R., Feng Z., Wu B. Research on emergence mechanism of collective intelligence from the complexity perspective // International Journal of Bio-Inspired Computation. 2023. Vol. 22, № 1, 28–39.
[Ye et al., 2024] Ye Q., et al. Improved Crowd Dynamics Analysis Considering Physical Contact Force and Panic Emotional Propagation // IEEE Access. 2024. Vol. 12, 34567–34580. [Zheng & Meister, 2024] Zheng J., Meister M. The unbearable slowness of being: Why our conscious thoughts run at 10 bits per second // Neuron. 2024. Vol. 112, № 24. P. 3996-4000.

Эффект «Красной Королевы» и эволюционная динамика

[Van Valen, 1973] Van Valen L. A new evolutionary law // Evolutionary Theory. 1973. Vol. 1. P. 1–30.
[Gavrilets, 1997] Gavrilets S. Coevolutionary Chase in Exploiter-Victim Systems with Polygenic Characters // Journal of Theoretical Biology. 1997. Vol. 186, № 4. P. 527–534.
[Carmichael & Hadžikadić, 2015] Carmichael T., Hadžikadić M. Predator-Prey Dynamics and the Red Queen Hypothesis: Putting Limits on the Evolutionary Arms Race // Journal on Policy and Complex Systems. 2015. Vol. 2, № 1.