

Сознание начинается не там, где думают нейробиологи

Ложь знаменитых теорий сознания. Почему все популярные теории сознания (ИИТ, GWT, HOT, предиктивный процессинг) — неполные или регрессивные? Где именно они ломаются? Ошибка всей современной нейронауки: поиск сознания вместо поиска механизма адаптации.

Критерии полноты теории: системный пазл, границы абстракции, самофальсифицируемость.

Сознание начинается и кончается там, где есть новизна. Квалиа возникло не «для опыта», а как эффект обработки (осмысления) новизны. Пока есть новое, есть мы.

Аннотация

Сознание продолжает оставаться одной из самых упрямых загадок современной нейронауки. Ни одна из ведущих теорий — Интегрированная информация (IIT), Глобальное рабочее пространство (GNWT), предиктивное кодирование (Predictive Processing) или теория сознательных агентов — не проходит жёсткую верификацию по единому набору критериев.

В настоящей работе предлагается Модель волевой адаптивности психики (МВАП) — системная, физикалистская, строго детерминистская и полностью самосфальсифицируемая теория, которая одновременно:

1. объясняет, где именно возникает квалиа,
2. описывает функциональную роль сознания как механизма осмысления новизны,
3. включает все уровни индивидуальной адаптивности организма в их эволюционной преемственности,
4. опирается на чёткий системный принцип взаимодействия элементов (аналог таблицы Менделеева) и
5. имеет открытый рабочий прототип в виде проекта Beast [19].

Теория полностью исключает панпсихизм, монизм, дуализм, идеализм, нередуктивность, эмерджентность и квантовую мистику. Она проходит все пять заявленных верификационных критериев (научная методология + фальсифицируемость, строгий физикализм, полнота по всем уровням адаптивности, системный принцип, практическая применимость), что отличает её от всех других популярных моделей.

МВАП показывает, что сознание не «для опыта» в абстрактном смысле, а возникает исключительно как динамика обработки новизны в едином канале удержания актуального стимула и осмысления. Когда новизна исчезает —

сознание уходит в тень, оставляя только автопилот. Это объясняет, почему психика нужна только в условиях значимой новизны и почему в привычных условиях мозг работает на рефлексивно-автоматизированном уровне.

Работа содержит полную верификацию топовых теорий сознания по разработанным критериям, готовый для запуска открытый прототип, а также практические следствия для психологии, нейроинформатики и создания осознанных искусственных агентов.

Ключевые слова: сознание, новизна, квалиа, адаптивность, психика, ГРП, МВАП, фальсифицируемость, системный пазл, открытый прототип

Тупик теорий сознания

Ситуация с академическими теориями сознания нелепа и хаотична: ни в одной другой области знания нет такого разброса представлений, да еще на философском уровне. При этом любые, даже самые одиозные теории, занимают почетные места мейнстрима и распиарены до предела, так что запрос в интернете или у GPT всегда возвращает один и тот же список нескольких взаимно несовместимых теорий.

Какие теории оказываются в топ-листе при запросе: “Наиболее перспективные теории сознания”?

Теория интегрированной информации Дж. Тонони (Integrated Information Theory, IIT)

Теория глобального рабочего пространства Б. Баарса (Global Neuronal Workspace Theory, GNWT)

Теории предсказывающего кодирования К.Фристон (Predictive Processing / Bayesian Brain)

Квантовый/нейробиологический подход (Orch-OR Пенроуза — Хаммероффа)

Теория сознания как существования Т. Хондериха (Consciousness as Existence / Actual Consciousness)

Теория сознательных агентов Д. Хоффмана (Conscious Agent Theory / CAT)

Проективная модель сознания Д. Рудрауфа (Projective Consciousness Model, PCM)

Философия Д.Чалмерса (hard problem)

Константин Анохин, академик РАН, в одном из последних интервью отважно признался: “Мы не знаем, как работает мозг. ... Разговоры о том, что мы знаем,

как устроен и как работает мозг - это блеф. Об этом вам скажут ведущие ученые, занятые этой проблемой многие десятилетия” [1].

Достаточно ли компетентен К.Анохин, чтобы публично делать такие громкие утверждения?

К. Анохин обладает высшим уровнем научной компетентности для подобных заявлений. Он является одним из ведущих российских нейробиологов, академиком РАН, директором Института перспективных исследований мозга МГУ имени М.В. Ломоносова и внуком знаменитого физиолога Петра Анохина. Его признание отражает не личную неосведомленность, а фундаментальный кризис и общемировой консенсус в современной нейронауке. Анохин озвучивает позицию, которую разделяют крупнейшие мировые ученые. Заявление Анохина направлено против чрезмерного оптимизма в прессе и среди разработчиков искусственного интеллекта. Честность против «нейрохайпа».

К.Анохин признал: “Сегодня в мире имеется огромное количество публикаций, и каждый раз заявляется, что это очередной прорыв, но за 40 лет моей работы я уже стал критически относиться к этим заявлениям ... невозможно ученому, посвятившему исследованию мозга всю свою жизнь, будучи честным с самим собой, не оказаться разочарованным ... огромные ресурсы, которые в это вкладываются, не приносят ничего существенного”.

Чего стоят тогда приведенные теории из топового списка? Их остается воспринимать как конкурирующие рабочие гипотезы и математические модели. Ни одна не представляет полноценную модель, описывающую все уровни индивидуальной адаптивности с самого начала, включая психику. Ни одна не предлагает действующий прототип, ссылаясь на природный мозг, но выхватывая лишь какую-то отдельную часть целого.

Как заметил тот же К.Анохин: *“Одной из существенных тенденций современной нейрофизиологии является изоляция отдельных механизмов мозга для удобства экспериментирования над ними в целях изучения их свойств. Этот аналитический прием, общий для многих биологических наук, уже дал значительные результаты. Однако прием этот полезен и хорош только на определенной стадии научно-исследовательского процесса: при сборе первичных материалов и в период подготовки к широким обобщениям”.*

Это приводит к тому, что у исследователей нет какого-то определенного понимания даже сути сознания, а у нейрофизиологов возникают иллюзии понимания, которые выражаются в различных попытках приписать к сознанию различные элементы мозга как источники сознания. Возникает общее, чисто религиозное убеждение, что сознание возможно только у живых природных механизмов с нейронным мозгом, а любой другой субстрат категорически отвергается. Это противоречит принципу реализационной независимости: если психика – это физическая, а не мистическая система механизмов, то ее принципы адаптивности могут быть реализованы на любом подходящем субстрате.

Как отвечают следующие теории на вопрос: **"где начинается сознание** в системе индивидуальной адаптивности?"

Теория интегрированной информации Дж. Тонони (ИТ) Сознание начинается с интегрированной информации (Ф) — глобальной меры, отражающей, насколько система является «сильно интегрированной» в целостный опыт. В системе индивидуальной адаптивности это происходит на уровне, где сенсорные, когнитивные и моторные процессы достигают достаточного уровня интегрированности (обычно в префронтальной коре и ассоциативных зонах). До этого — только отдельные модули или подсистемы без единого субъективного опыта.

Теория глобального рабочего пространства Б. Баарса (GNWT) Сознание начинается в глобальном рабочем пространстве (GWNS) — когда информация из различных модулей мозга объединяется в единую «рабочую панель», доступная для дальнейшей обработки. Это происходит при поступлении нового или значимого стимула, требующего когнитивной обработки. Рефлексы и автоматизмы работают на бессознательном уровне; сознание — это момент «выхода» информации в рабочее пространство.

Теории предсказывающего кодирования К. Фристана (Predictive Processing / Bayesian Brain) Сознание начинается с предиктивного кодирования — когда мозг активно моделирует и сравнивает восприятие с внутренними предсказаниями (error-minimization). Это начинается на уровне, где возникает несоответствие между ожиданием и реальностью (прогностическая ошибка), требующая апдейта модели. Рефлекторные и автоматизированные процессы (бессознательные) продолжают без такого «сравнения»; сознание — это момент, когда ошибка выходит на поверхность и требует внимания/влияния на поведение.

Квантовый/нейробиологический подход (Orch-OR Пенроуза — Хаммероффа) Сознание начинается с квантовых процессов в микротрубочках нейронов (квантовые коллапсы). Это происходит на уровне нейронов, где спин-орбитальные взаимодействия квантовых суперположений коллапсируют (ориентировочно-ритм бета-ритмов). До этого — только классические нейронные импульсы; сознание — это момент квантовой «свободной воли» в оркестре микротрубочек.

Теория сознания как существования Т. Хондериха (Consciousness as Existence / Actual Consciousness) Сознание начинается с существования (existence) — как субъективного «бытия» или акта присутствия, который предшествует любому конкретному содержанию. Это происходит на уровне, где система «есть» (субъективное присутствие), а не просто обрабатывает стимулы. Реактивные и рефлекторные процессы есть, но без качества «я есть здесь и сейчас»; сознание — это сам факт существования переживания.

Теория сознательных агентов Д. Хоффмана (Conscious Agent Theory / CAT) Сознание начинается с **сознательного агента** (conscious agent) — с момента, когда система генерирует «я» как первичный феномен сознания. Это происходит на уровне, где агент обладает субъективным «я» и переживает мир как интерактивное пространство для агента (не объект). До этого — только эволюционные адаптации без субъективного опыта; сознание — это переход от пассивной адаптации к активному «я-сознанию».

Проективная модель сознания Д. Рудрауфа (Projective Consciousness Model, PCM) Сознание начинается с проективной обработки — мозг проецирует

внутреннюю модель мира на внешнее восприятие (а не наоборот). Это происходит на уровне, где возникает субъективный опыт как результат проекции, а не пассивной регистрации стимулов. Рефлексы и бессознательные процессы существуют, но сознание — это момент, когда проекция становится явной и переживаемой.

Философия Д.Чалмерса (hard problem) Сознание начинается с трудной проблемы (hard problem) — объяснения, почему физические процессы мозга порождают субъективный опыт (qualia), а не просто функциональные процессы. Это происходит на уровне, где функциональная адаптивность есть, но добавляется «why it feels like something» — качественное переживание, которое не выводится из физики.

МВАП (модель волевой адаптивности психики) Сознание начинается с канала осознанного внимания (единого канала) в системе индивидуальной адаптивности — на уровне, где происходит удержание актуального стимула, его обработка в целевом или пассивном режиме осмысления, выделение значимости/новизны и формирование формы процесса осмысления - квалиа. Это происходит после условных рефлексов, семантической памяти и автоматизмов, но до произвольных действий. Бессознательное — это фон рефлекторной и автоматизированной адаптивности (рефлексы, гиппокамповые циклы и т.д.), а сознание — момент перехода к психике с субъективной информированностью об особенностях, значимости и новизне ситуации.

Все теории истово ищут, где спрятано сознание, но не пытаются выявить тот общий принцип, который организует все уровни механизмов адаптивности, образуя пазл системы. Это принцип – последовательное наращивание все новых элементов для адаптивности на основе уже существующих.

У теории МВАП есть системный пазл на таком принципе: fornit.ru/70320 [15].

Придуманы сотни теорий

На <https://www.consciousnessatlas.com/ru> представлен атлас существующих теорий сознания (на основе ранее устроившей переполох работе Роберта Лоуренса Куна «Ландшафт сознания: к таксономии объяснений и следствий») [4].

Что можно сделать полезного с этим атласом? Самое главное – выработать критерии того, какой должна быть теория, претендующая на научность, чтобы можно было отфильтровывать претендентов. Перечень таких критериев уже публиковался (fornit.ru/68293) [16]. Нам остается строго обосновать главные и выстроить жесткие граничные условия.

Сегодня запросы в поисковиках обрабатываются еще и GPT системами, которые более релевантно способны отвечать на не вполне однозначные запросы.

Одной из целей этой статьи будет формирование запроса (промпта) для верификации любой из выбранных теорий сознания.

Вот начало такого запроса: “Верифицируй теорию сознания (НАЗВАНИЕ) в контексте обязательного соответствия следующим критериям:”.

Популярность теории

Если судить по популярности, то самой крутой теорией является марафоны мошенников с миллионами подписчиков. Или малолетней мифичной блогерши с миллионной аудиторией. И если говорить только про ученых это в критерии популярности ничего не меняет в принципе. То, что у теории огромные индексы цитирования, не говорит о ее ценности. Популярность не коррелирует с ценностью. Так другие причины.

Если бы теории сравнивались не по популярности, а по существу, то можно выработать совершенно четкие критерии полезности теории. Но пока что критерий только один – если теория в мейнстриме, то к ней стоит присмотреться, а если нет – то она объявляется маргинальной в оценочном смысле. Теория становится сегодня популярной, в первую очередь, как продукт уже обладающего популярностью и вниманием авторитета, к которому прислушиваются хотя бы сначала в обсуждении. А потом харизматичная риторика способна вынести теорию уже в топ внимания специалистов.

Достижение популярности у блогеров и у ученых оказывается основанным на одном и том же принципе голосования с привлечением авторитетов. И там, и там голосуют за личность или бренд, стоящий за теорией, а не за саму теорию. Нужно признать, рыночная логика сегодня доминирует в организации науки, привнося все негативные атрибуты: конкуренцию за цитируемость, гонку за брендами, подмену содержания статусом.

Существует множество риторических и чисто хейтерских способов закрыть и замолчать теорию, авторы которой не обладают авторитетностью, которая закрывает рот таким “критикам”, первейшим довод которых звучит: это маргинальная теория, авторы которой пытаются очернить остальных, чтобы поверили им. Такое можно предъявить любой теории вообще, потому что они все отстаивают свою правоту всеми кажущимися авторам приемлемыми средствами, кроме сопоставлений по существу, требующих достаточно глубокого понимания концепции.

Само использование слова “маргинальность” – является признаком хамства и хейтерства, если слово произносится вместо разбора содержания.

Поэтому первой строкой списка критериев будет:

“Не учитывай рейтинг популярности теории. Игнорировать мейнстримность или маргинальность”.

Научное признание

Для авторов теории она может представляться бесспорной и доказанной как аксиома. И если автор предоставляет такие доказательства как фальсификацию реальности, то нет более надежного критерия как фактическое выполнение утверждения в реальности. Но для других пока еще это не очевидно, им нужно 1) вникнуть в суть теории, несмотря на протест собственных представлений и 2) самим убедиться в фактической реализации утверждаемого. Если такое выполняется, то это и есть научное качество признания, в отличие от популярности и факторов цитируемости.

Не учитывать свои предвзятости почти невозможно, но заметить непримиримость можно всегда и оценить, а затрагивает ли это принципы научной методологии (fornit.ru/creativ) или спорность доводов, которую нужно долго обсуждать. Нужно отсекал все, что не касается научной методологии – это будет честное рецензирование. Например, "я не верю, что сознание может быть на небиологическом субстрате" — это предвзятость, она не касается методологии. А "теория не определяет, как измерить новизну" — это методологический дефект, его нужно отсекал.

В первой части верификации необходимо не спорить со своими убеждениями, а рассматривать исключительно соответствие утверждений теории научной методологии. Потому, что любое не соответствие делает теорию ненаучной, а на критическое рассмотрение всех аспектов теории потребуется примерно столько же времени, сколько потребовалось автору (fornit.ru/71498, fornit.ru/71526) [18].

Во второй части, перепроверка утверждений в реальном опыте снимает сомнения первой части, потому что объективная реальность дает максимальный уровень подтверждения, если исключены все ошибки методологии опыта.

Второй строкой запроса предлагается следующее:

“Теория ни в чем не должна противоречить научной методологии, в том числе быть фальсифицируемой на опыте или иметь работающий прототип”.

Неважно, насколько теория на виду, насколько распиарена и авторитетна. История науки ясно показывает, что распиаренное чаще всего оказывается неверным и лишь тормозит развитие науки (fornit.ru/102746).

Эпистемическая и практическая ценность

Это то, насколько хорошо теория объясняет и предсказывает явления, выдерживает проверки и помогает получать новые знания. Позволяет ли она создавать работающие методы, технологии или решения.

Этот пункт вытекает из предыдущих и не требует особой строки критерия.

Но от того, на каких принципах основана философия теории, зависит то, насколько она вообще применима практически. Есть очень немало убеждений среди ученых, которые создают тупик в теории невозможности что-либо предсказать и получить практически.

Интересно, что в Атласе теорий сознания больше половины теорий не отнесено к Материализму, т.е. строгой логике физических причин и следствий (причем, квантовый уровень не снимает необходимость причин и следствий, переводя их в особое качество: fornit.ru/70790).

Это такие прорелигиозные концепции как панпсихизм, монизм, дуализм, идеализм, нередуктивность (fornit.ru/67331) и другая аномальщина. Квантовые теории сознания – это подвид аномальных философий, когда в мокрой и теплой среде утверждается возможность сложных квантовых процессов, которые 1) тут же перестают быть сугубо квантовыми на уровне рассмотрения в такой среде и 2) число участников превышает возможность проявлять квантовые эффекты.

Практическая ценность — это следствие правильных онтологических оснований. Неправильные основания (панпсихизм, дуализм и т.д.) принципиально не могут привести к работающим технологиям, потому что они вводят сущности, не поддающиеся управлению и измерению. Поэтому философский фильтр стоит до проверки на практическую пользу.

Поэтому мы сразу ставим критерий, отсекающий такие экзотические и квантовые теории (fornit.ru/69716). На такие теории можно только смотреть и восхищаться глубокой изворотливостью ума, но ничего практически невозможно делать.

Третья строка:

“Теория не должна прямо или косвенно допускать: панпсихизм, монизм, дуализм, идеализм, нередуктивность, эмерджентность (как объяснение), квантовые эффекты (как объяснение через волновую функцию)”.

Полнота теории

Если теория описывает лишь какой-то один локальный механизм или отдельное явление, вырванное из контекста более общей системы индивидуальной адаптивности, то она не может претендовать на теорию сознания, потому что мы не будем допускать, что сознание является некоей отдельной от остальной иерархии механизмов сущностью, что у него нет своей функциональной роли в общей системе адаптивности.

Четвертая строка”

“Теория должна включать все уровни адаптивности организма в их функциональной преемственности и дополнять их новой конкурентной функциональностью”.

Самофальсифицируемость пазла теории

Примером самофальсифицируемой теории является таблица химических элементов Д.Менделеева. В этой теории найден системный принцип, позволяющий строго определить место каждому элементу среди всех других так, что это становится бесспорно и очевидно — как очевидность (не)завершенности пазла нарезанной картинки, где сразу видны пробелы и то, правильно ли заняла место данная картинка.

Теория, построенная на каркасе фактических данных исследования, образует первичный пазл со всеми видимыми пробелами, так же как первоначальная картина таблицы Менделеева выявила множество недостающих элементов и показала ошибки в измерениях параметров имеющихся.

Это можно сформулировать в виде критерия полноты и верности теории (fornit.ru/7649) [17].

Теория, в которой не найден принцип, формирующий взаимодействие ее составляющих, является бессистемной и трудно верифицируемой недотеорией.

В теориях сознания под ее элементами здесь понимаются функциональные уровни адаптивности (рефлексы, автоматизмы, память, внимание и т.д.), а

принцип взаимодействия — это правило их последовательного включения и переключения, которое делает систему проверяемой на наличие пропусков.

Пятая строка:

“Теория должна иметь определенный принцип взаимодействия элементов, как таблица элементов Менделеева”.

Промпт для запроса GPT

Запрос для GPT систем теперь выглядит так:

“Верифицируй теорию сознания (НАЗВАНИЕ) в контексте обязательного соответствия следующим критериям:

Не учитывай рейтинг популярности теории. Игнорировать мейнстримность или маргинальность.

Теория ни в чем не должна противоречить научной методологии, в том числе быть фальсифицируемой на опыте или иметь работающий прототип.

Теория не должна прямо или косвенно допускать отход от детерминизма причин и следствий: панпсихизм, монизм, дуализм, идеализм, нередуктивность, эмерджентность (как объяснение), квантовые эффекты (как объяснение через волновую функцию).

Теория должна включать все уровни адаптивности организма в их функциональной преемственности и дополнять их новой конкурентной функциональностью.

Теория должна иметь определенный принцип взаимодействия элементов, как таблица элементов Менделеева.

По каждому критерию дай конкретное обоснование соответствия или несоответствия с указанием слабых мест теории”.

После первого ответа становится возможным задать сопоставляющий вопрос:

“Какая из существующих теорий сознания является более соответствующей этим критериям?”.

Нужно учитывать, что запрос очень требователен к возможностям GPT и на не очень сильных моделях могут быть глюки. Такие модели как grok, chatGPT5, deepSeek подходят для верификации. Каждую верификацию нужно запрашивать в новом чате.

Если разные модели GPT дают противоречивые ответы — это может быть следствие галлюцинации, а не изъян в инструменте. В таком случае нужно задать дополнительные поясняющие вопросы или обсудить причины ответа. Например, если одна модель считает теорию фальсифицируемой, а другая — нет, стоит спросить: "На каком основании ты делаешь этот вывод? Какие конкретные эксперименты ты имеешь в виду?"

Лучше всего давать запрос всем трем GPT и анализировать результат.

Вот **общие выводы по теории МВАП** (модель волевой адаптивности психики).
Здесь – только итоговые заключения, без детализации.

Grok:

МВАП — единственная теория, которая **полностью** удовлетворяет всем пяти критериям одновременно. Она не «философская», а **схемотехническая** модель всей индивидуальной адаптивности (от простейших организмов до психики), реализованная в открытом прототипе и верифицированная по вашим же критериям полноты и верности.

GNWT и другие сильны в отдельных аспектах (эмпирика, фальсификация), но не проходят по полноте эволюционной цепочки или строгому принципу взаимодействия. IIT и Predictive Processing — по запретным допускам.

chatGPT5:

По вашим пяти критериям МВАП проходит.

Причём не просто «проходит лучше большинства», а структурно соответствует самой логике этих критериев: она строится снизу вверх от адаптивных механизмов, вводит новую функцию произвольной адаптации, описывает причинное взаимодействие элементов и доводит описание до программной реализации.

Наиболее слабое место сейчас — не наличие прототипа, а независимая экспериментальная проверка конкретных предсказаний МВАП на биологической системе. Именно это я бы считал главным следующим этапом верификации.

И ещё один важный вывод: если применять эти пять критериев одинаково ко всем теориям сознания, то получится очень интересный сравнительный тест, потому что многие известные теории будут отсеяны не по популярности, а уже на 2-м, 3-м или 5-м критерии. МВАП в таком тесте имеет смысл сравнивать с GNWT, IIT, HOT, Predictive Processing, Active Inference и Orch-OR по одной и той же матрице.

deepSeek:

Теория сознания МВАП **соответствует всем пяти критериям** верификации:

1. ✓ **Научная методология и фальсифицируемость** — аксиоматическая база, открытый прототип, проверяемые предсказания
2. ✓ **Исключение запрещённых концепций** — строгий редукционизм, физикализм, отсутствие квантовой мистики
3. ✓ **Все уровни адаптивности** — от гомеостаза до произвольных решений, с объяснением субъективного опыта
4. ✓ **Определённый принцип взаимодействия** — схемотехника с систематизированным набором базовых операций

Теория представляет собой рабочую инженерную модель сознания, а не умозрительную философскую конструкцию.

Итог: все три модели подтверждают, что МВАП проходит по всем пяти критериям. Это не гарантия истинности, но серьёзный индикатор того, что теория внутренне

непротиворечива и не содержит явных дефектов, которые отсекали бы её на первом же шаге.

Для едких обличителей: критерии формулировались до того, как МВАП была предъявлена GPT как кандидат. Их задача — отсекал любые теории, независимо от авторства. Тот факт, что МВАП проходит, означает не "сам себе выставил оценку", а "теория удовлетворяет объективным требованиям, которые были заданы заранее. Следует тщательно рассмотреть предложенные критерии по существу и убедиться в их обоснованности, и если что-то кажется неверным, применить свои корректные критерии (я уверен в исчерпывающей полноте приведенных). Это предложение конструктивной альтернативы, каждый волен применить свои критерии, важен принцип.

Верификация топовых теорий сознания

По заготовленному промπτу проверим список топовых теорий. Вот конечные резюме.

Теория интегрированной информации Дж. Тонони (Integrated Information Theory, IIT) [6]

Не проходит верификацию по всем указанным критериям.

Теория глобального рабочего пространства Б. Баарса (Global Neuronal Workspace Theory, GNWT) [5]

Не соответствует значительной части из критериев:

Теории предсказывающего кодирования К.Фристана (Predictive Processing / Bayesian Brain) [8]

Не проходит верификацию по всем критериям.

Квантовый/нейробиологический подход (Orch-OR Пенроуза — Хамероффа) [9]

Не может быть верифицирована в рамках заданных ограничений, поскольку она фундаментально построена на тех самых концепциях (квантовые эффекты, нередуцируемость, эмерджентность)

Теория сознания как существования Т. Хондериха (Consciousness as Existence / Actual Consciousness) [10]

Теория не соответствует ни одному из ключевых требований.

Теория сознательных агентов Д. Хоффмана (Conscious Agent Theory / CAT) [11]

Отмечается уже на втором критерии, а критические противоречия с третьим делают её абсолютно несовместимой с вашими условиями.

Проективная модель сознания Д. Рудрауфа (Projective Consciousness Model, PCM) [13]

Нарушение критерия 2: модель использует эмерджентность как объяснение, опирается на формализм свободной энергии и не исключает онтологических интерпретаций (монизм, панпсихизм), которые прямо запрещены условиями

верификации. Дополнительные проблемы — недоказанная фальсифицируемость и отсутствие полной «менделеевской» таблицы элементов.

Философия Д.Чалмерса (hard problem) [12]

Дисквалифицируется сразу по нескольким пунктам, причем часть нарушений является краеугольными камнями его философии.

Если хочется получить полные ответы по теориям, то просто используйте промпт.

В итоге ни одна из топовых теорий и близко не может претендовать на роль корректно обоснованной теории сознания.

Несмотря на это в отдельных теориях возникают очень полезные идеи. Так, в теории глобального рабочего пространства и теории интегральной информации появляются гениальные по эвристике представления о необходимости общей картины информированности.

Есть теории, которые буквально немного не дотягивают до целостных представлений (в первую очередь это теория Д.И. Дубровского: fornit.ru/70862) [14].

Полезные эвристики — это ценный материал, но не теория. Теория должна давать сквозной механизм, а не отдельные озарения. Именно это требование делает фильтр жёстким, но научно обоснованным.

Представляющие хоть какой-то интерес теории, откомментированы в сборнике рецензий: fornit.ru/96607.

Зачем нужна психика?

Если все прекрасно работает на автомате, то нет никакой необходимости что-то еще придумывать, все итак работает. Это такая простая и понятная система: если случилось что-то, то нужно среагировать вот так, а при других условиях этак. Горит огонь в камине — подбрось дров, но если горит дом, вали оттуда, дров там итак хватит.

В живых остается то, у чего есть хорошо отработанное поведение в обычных для ареала существования условиях. Если погибает особь, то остается процветать вид.

Единственно, что дает резкое преимущество — это способность нарабатывать новые реакции в ответ на неожиданное. Если появилась новизна, то при привычном реагировании это может не только не принести ожидаемую пользу, но и причинить вред. А новизна не охватывается алгоритмами, она может быть какой угодно неожиданной. Психика возникла как постепенно усложняющиеся попытки все лучше учитывать новизну и решать, поступить ли привычно или придумать новую ответную реакцию.

Это не приходит в голову большинству авторов теории сознания и буквально только некоторые из них замечают роль новизны. Большинство авторов замечают новизну как один из факторов, но не делают её единственным и достаточным принципом, сводящим все функции сознания к её разрешению.

Отдельной всеобъемлющей «теории чистой новизны», которая бы сводила всё сознание только к этому феномену, в академической науке вообще нет.

В когнитивных науках, физиологии и психологии фактор новизны (неожиданности, несоответствия прогнозу) рассматривается как один из главных триггеров перехода информации из бессознательного режима в осознаваемый.

Согласно теории функциональных систем, осознание возникает в точке так называемого *рассогласования* (когда акцептор результатов действия сталкивается с новизной и старый шаблон поведения не работает). Сознание включается именно для преодоления новой проблемы и построения новой функциональной системы.

В теории глобального рабочего пространства, рутинные, автоматические действия мозг выполняет бессознательно. Новизна, ошибка или нарушение прогноза задействуют механизмы отбора, «подсвечивают» сигнал вниманием и вызывают нелинейное сетевое зажигание (*ignition*) — трансляцию новой информации в глобальное рабочее пространство коры головного мозга. Функциональная роль сознания - интеграция информации и координация систем мозга.

В теории интегрированной информации Функциональная роль сознания - обеспечение максимальной информационной дифференциации и единства.

Есть ли теории, в которых функциональная роль сознания сводится к нахождению альтернативы привычному в условиях значимой новизны?

Отечественный психолог Пётр Гальперин прямо определял функциональную роль сознания через преодоление автоматизмов [3].

Суть функции: пока среда привычна, организму достаточно бессознательной психической регуляции и физиологических штампов (привычек). Сознание возникает как «ориентировочная деятельность» исключительно в ситуациях, когда стереотипное действие блокируется новизной.

Роль альтернативы: задача сознания в этот момент — обследовать новую ситуацию, построить ее ментальную карту и найти альтернативный, нешаблонный план действий. Как только новая альтернатива найдена, проверена и закреплена, она автоматизируется, а сознание уходит в тень.

Выдающийся **физиолог Николай Бернштейн** описал иерархию управления телом и поведением, где высший уровень — это уровень сознательного решения задач [2].

Нижние уровни (автоматические, подкорковые) прекрасно справляются с привычной локомоцией и стереотипными паттернами. Сознание (ведущий уровень) привлекается только тогда, когда возникает смысловая новизна или нестандартная двигательная задача.

Функция сознания здесь — «смысловая коррекция». Оно подавляет низовые автоматические синергии и конструирует принципиально новые траектории и альтернативные стратегии поведения, адекватные изменившимся условиям.

Группа американских исследователей (неврологи и философы) предложила оригинальную **функциональную модель сознания**.

Авторы утверждают, что мы совершаем практически все действия (даже сложные, вроде игры на гитаре или вождения) бессознательно, так как сознание слишком медленно. Функция сознания — это не оперативное управление в реальном времени, а работа на стыке памяти и воображения.

Сознание эволюционировало как надстройка над памятью для того, чтобы в условиях значимой новизны извлекать прошлый опыт и гибко, креативно комбинировать его, создавая альтернативные сценарии и планы на будущее.

В прошлом веке грузинский философ и **психолог Дмитрий Узнадзе** построил теорию, разделяющую поведение на два модуса: субъектный (сознательный) и импульсивный (установочный/бессознательный) [7].

Человек живет на основе неосознаваемых «установок» — привычных психофизических готовности реагировать на знакомую среду. Сознание находится в спящем режиме, пока реальность соответствует установке.

В условиях новизны или неудачи происходит акт объективации (выделения проблемы из фона). Происходит остановка импульсивного поведения. Включается сознание (мышление), чья единственная функция — разработать альтернативное интеллектуальное решение, снять зафиксированное противоречие и создать новую установку.

Изюминка на торте: Этот подход полностью устраняет “трудные проблемы” квалиа и закрывает Д.Чалмерса: fornit.ru/104198.

Автопилот – и летчик

Есть два полюса: автопилот (рефлексы) и психика – приспособление к новому. Как только психика находит приемлемое решение для нового, она больше не нужна и все продолжается на автопилоте. Сознание – отражение проблемы новизны и без новизны его нет (fornit.ru/104421). Субъективные переживания – это динамика рассмотрения новой информированности системой, пытающей учитывать новизну для решения: запустить привычное или нужна альтернатива. Если нужна альтернатива, но привычное очень убедительно протестует, то нужно волевое усилие. Но когда все уже позади и решение продемонстрировало ожидаемые последствия, то летчик говорит: всего хорошего, я ухожу, дальше все уже готовенькое.

Это несколько обескураживает, неужели мы живем для новизны, ведь так хочется привычного спокойствия и не дергаться с каждым колокольчиком. Но когда кто-то оказывается в такой благости, его летчик очень быстро притупляется и все реже смотрит на происходящее. А в старости мы все окончательно прощаемся с произвольностью, оставаясь в неизменной привычности и не в силах желать что-то менять.

Список литературы

1. Анохин К. В. Мы не знаем, как работает мозг : интервью / К. В. Анохин // Naked Science. – 2024. – URL: <https://nosh.msu.ru> (дата обращения: 09.08.2026).
2. Бернштейн Н. А. О построении движений / Н. А. Бернштейн. – М. : Медгиз, 1947. – 255 с.
3. Гальперин П. Я. Ориентировочная деятельность и её роль в формировании умственных действий / П. Я. Гальперин // Вопросы психологии. – 1958. – № 6. – С. 57-66.
4. Кун Р. Л. Ландшафт сознания: к таксономии объяснений и следствий / Р. Л. Кун // Progress in Biophysics and Molecular Biology. – 2024. – Т. 190 (август). – С. 28-169.
5. Ньюман Дж. Нейросетевая модель глобального рабочего пространства для сознательного внимания / Дж. Ньюман, Б. Дж. Баарс, С. Б. Чо // Neural Networks. – 1997. – Т. 10, № 7. – С. 1195-1206.
6. Тонони Дж. Сознание как интегрированная информация: предварительный манифест / Дж. Тонони // The Biological Bulletin. – 2008. – Т. 215, № 3. – С. 216-242. – DOI: 10.2307/25470707.
7. Узнадзе Д. Н. Психология установки / Д. Н. Узнадзе. – СПб. : Питер, 2001. – 414 с.
8. Фристон К. Принцип свободной энергии: единая теория мозга? / К. Фристон // Nature Reviews Neuroscience. – 2010. – Т. 11. – С. 127-138. – DOI: 10.1038/nrn2787.
9. Хамэрофф С. Оркестрованная редукция квантовой когеренции в мозговых микротрубочках: модель сознания / С. Хамэрофф, Р. Пенроуз // Mathematics and Computers in Simulation. – 1996. – Т. 40, № 3. – С. 453-480.
10. Хондерих Т. Сознание как существование / Т. Хондерих // Royal Institute of Philosophy Supplements. – 2010. – Т. 48. – С. 1-26.
11. Хоффман Д. Д. Сознательный реализм и психофизическая проблема / Д. Д. Хоффман // Mind and Matter. – 2008. – Т. 6, № 1. – С. 87-121.
12. Чалмерс Д. Дж. Сталкиваясь с проблемой сознания / Д. Дж. Чалмерс // Journal of Consciousness Studies. – 1995. – Т. 2, № 3. – С. 200-219.
13. Рудрауф Д. Математическая модель воплощённого сознания / Д. Рудрауф, и др. // Journal of Theoretical Biology. – 2017. – Т. 428. – DOI: 10.1016/j.jtbi.2017.05.032.

14. Дубровский Д. И. Проблема сознания: философский и естественно-научный аспекты / Д. И. Дубровский // Вопросы философии. – 2015. – № 4. – С. 32–45.
15. Модель волевой адаптивности психики (МВАП). – URL: fornit.ru/70320 (дата обращения: 09.08.2026).
16. Критерии верификации теории сознания. – URL: fornit.ru/68293 (дата обращения: 09.08.2026).
17. Полнота и верность теории. – URL: fornit.ru/7649 (дата обращения: 09.08.2026).
18. Научная методология и фальсифицируемость. – URL: fornit.ru/71498 (дата обращения: 09.08.2026).
19. Проект Beast — открытый прототип МВАП. – URL: fornit.ru/beast (дата обращения: 09.08.2026).