2026

Метатеория функционального наблюдателя

Часть I - Архитектура и формальная структура

Версия 1.11

АРТЕМОВ ДМИТРИЙ

Оглавление

Аннотация

Проблема сознания традиционно формулируется как проблема функционального доступа (Baars, 1988; Dehaene, 2014), интеграции причинной структуры (Tononi, 2004; Tononi et al., 2016), предсказательной регуляции (Friston, 2010; Clark, 2013) или мета-представления (Rosenthal, 2005). Тем не менее «трудная проблема сознания» (Chalmers, 1996) указывает на сохраняющийся разрыв между объяснением механизмов обработки информации и форматом данности этих состояний системе.

Настоящая работа предлагает метатеорию функционального наблюдателя (FOM), в которой сознание рассматривается как архитектурный режим организации внутренней модели агента. Агент описывается как регуляторный контур, удерживающий себя в пределах окна устойчивости посредством каузальных входов и выходов. Вводятся структурные параметры управляемости, планирования и горизонта каузального учета.

Сознание трактуется не как дополнительная каузальная сущность, а как режим организации внутренней модели через якорь реальности, само-индексацию и карту значимости. Квалиативный формат определяется как архитектурное свойство этой связки, обеспечивающее единство текущего ставочного состояния без введения новой причинности.

Проводится строгое разведение уровней реальности, модели реальности и формата данности, что позволяет переосмыслить статус «трудной проблемы» как следствия категориального смешения формата представления и представляемого содержания.

Теория задает параметры анализа и критерии операционализации, открывая возможность сравнительного исследования сознательных и несознательных архитектур в биологических и искусственных системах.

I. Базовые понятия

Реальность R определяется как функциональная сторона каузальных состояний и переходов, относительно которой **агент B** определяется как источник и получатель изменений.

Отличие агента от реальности состоит в том, что агент поддерживает собственное состояние в пределах некоторого **окна устойчивости L**, при выходе за границы которого он перестает быть данным агентом.

Границы агента и критерии его идентичности задаются операционально в зависимости от исследовательских задач. Метатеория фиксирует лишь структурные условия агентности, не предписывая конкретной реализации.

1. Обозначения воздействий

- **in** — каузальные воздействия реальности на агента
- **out** — каузальные воздействия агента на реальность

Под системой далее понимается регуляторный контур, включающий агента и его каузальные входы (in) и выходы (out).

2. Поясняющая аналогия

Венерина мухоловка служит примером минимального агента, включенного в каузальный контур **R ↔ B**.

Как регуляторная система, она:

- подвергается воздействиям среды;
- реагирует на них;
- поддерживает собственное состояние (обмен веществ, рост, выживание) в допустимых пределах.

Когда насекомое касается чувствительных волосков на поверхности листа, происходит совокупность физических процессов:

- механическое воздействие;
- изменение электрического потенциала тканей;
- запуск биохимических каскадов.

Это и есть **in** — не «восприятие» и не «сигнал» в когнитивном смысле, а совокупность физических взаимодействий, приходящих из R в B.

В ответ на эти изменения агент:

- смыкает створки;
- изменяет конфигурацию листа;
- физически воздействует на среду, удерживая насекомое.

Это и есть **out** — каузальное воздействие B на R, выраженное в изменении состояния среды.

Контур описывается схемой:

$R \rightarrow in \rightarrow B \rightarrow out \rightarrow R$

3. Применимость модели

В схеме из двух агентов ($B_1 \leftrightarrow B_2$) каждый агент выступает элементом реальности другого.

Формально это можно записать так:

- $\mathbf{B_1 \leftrightarrow R_1}$, где $R_1 \supset B_2$
- $\mathbf{B_2 \leftrightarrow R_2}$, где $R_2 \supset B_1$

Таким образом, реальность всегда задается относительно границы конкретного агента.

4. Сознание как формат доступа

Сознание Q рассматривается как формат доступа для решения задач агентом. Анализ сознания в рамках теории осуществляется без обращения к метафизическим предпосылкам.

II. Окно устойчивости (L)

1. Определение

Окно устойчивости (или окно жизни) — это область входящих каузальных воздействий (in), при которых агент сохраняется как агент, то есть продолжает существовать в той конфигурации и с теми функциями, которые позволяют считать его тем же агентом.

2. Выход за пределы L

Если воздействие **in** выходит за пределы L, возможны два исхода:

- **разрушение** — физическое уничтожение или необратимая деградация структуры;
- **утрата агента** — прекращение регуляции, разрушение контура in/out, потеря агентной идентичности.

3. Многомерность и динамичность L

В общем случае L — многомерная область в пространстве параметров, например:

- температура;
- влажность;
- химический состав среды;
- доступность ресурсов.

Параметры могут быть взаимозависимыми: изменение одного параметра способно сужать или расширять допустимый диапазон значений других параметров.

4. Примеры вариантов L

- для термостата — диапазон температур, обеспечивающий работоспособность;
- для двигателя — диапазон нагрузок и температур, исключающий разрушение;
- для организма — диапазон условий, при которых сохраняется жизнеспособность.

5. Пример удержания в пределах L

Для мигрирующих птиц сезонные изменения среды приводят к тому, что текущие входные воздействия (**in**) приближаются к границам окна устойчивости **L**. Регуляторный ответ заключается в изменении пространственного положения в реальности **R**, что приводит к удержанию входных воздействий (**in**) в окне **L**.

6. Связь L и сознания

Используя отрицательную логику, можно зафиксировать:

- чем шире и стабильнее окно устойчивости **L**, тем менее функционально оправдано введение отдельного механизма сознания;
- если система сохраняется при значительных колебаниях условий, простая реактивная регуляция оказывается достаточной;
- сознание не дает функционального выигрыша там, где удержание внутри L не требует моделирования будущих состояний.

И наоборот:

- чем уже и динамичнее окно **L**;
- чем выше риск выхода за его границы;
- чем выше неопределенность последствий действий (частичная наблюдаемость, длинный горизонт, необратимость);

тем более необходимыми становятся механизмы, которые:

- реализуют регуляцию через внутренние представления;
- позволяют оценивать возможные будущие состояния;
- увеличивают вероятность удержания агента внутри L.

В этом смысле сознание рассматривается не как универсальное свойство агентов, а как частное архитектурное решение, возникающее там, где простая регуляция перестает обеспечивать устойчивость агента.

III. Структурная соразмерность (C)

Структурная соразмерность (или управляемость) C — структурное соотношение разнообразия значимых входных состояний (in) и доступного пространства управляющих воздействий (out).

Обозначается приближенно как:

C ≈ in / out

Чтобы C имела смысл, in и out должны быть приведены к единой оси значимости, связанной с устойчивостью агента. Одна и та же система может быть описана различными способами при условии, что параметры соответствуют задаче регуляции.

Конкретная формализация относится к прикладным дисциплинам: теория управления, инженерия, биология или искусственный интеллект.

Структурная соразмерность может быть выражена, например, как:

- variety в духе Эшби;
- controllability в формализме теории управления;
- пропускная способность канала in/out.

Во всех случаях **C** остается рабочим параметром структурной согласованности.

1. Предельные случаи управляемости

1.1. in = 0 — отсутствие необходимости в управлении

Если на агента не поступают каузальные воздействия со стороны реальности, его состояние не зависит от внешней среды. В этом случае контур разомкнут.

1.2. out = 0 — отсутствие возможности управления

Если агент не оказывает каузальных воздействий на реальность, он может подвергаться внешним влияниям, и его состояние может изменяться, однако регуляторного ответа не возникает. В этом случае контур разомкнут. Типичный пример — пассивный объект, например камень: на него действуют температура, давление, механические удары, но он не поддерживает собственное состояние в пределах допустимого диапазона посредством активного воздействия на среду.

1.3. C ≈ 1 — согласованная регуляция

Случай **C ≈ 1** соответствует ситуации, в которой разнообразие входящих воздействий (**in**) соразмерно разнообразию доступных выходных воздействий (**out**). Важно подчеркнуть: **C ≈ 1 не означает оптимальности,** а означает лишь структурную соразмерность входа и выхода.

2. Отрицательный фильтр

Чем сильнее управляемость C отклоняется от согласованного диапазона, тем менее устойчивыми становятся формы регуляции системы, включая режимы, связанные с сознанием. В этом смысле C может использоваться как отрицательный фильтр при анализе условий возможности сознания, но не как его достаточный критерий.

3. Пример

Рассмотрим регулятор **B**, состоящий из емкости с водой и задвижки:

- **in** — уровень воды в диапазоне от 0 до 5 метров;
- **out** — степень открытия задвижки.

Функционально значимо не все формальное пространство **out**, а лишь та его часть, которая реально влияет на уровень воды:

- задвижка расположена на высоте 1 метр, ниже которого уровень не может быть снижен;
- при превышении 4 метров срабатывает аварийный перелив.

В результате:

- пространство значимых состояний **in** охватывает диапазон от 0 до 5 метров;
- пространство реально эффективных воздействий **out**, приведенных к влиянию на уровень воды, соответствует диапазону примерно от 1 до 4 метров.

Таким образом, **in** и **out** приводятся к одной оси значимости, связанной с удержанием устойчивого уровня воды. В рамках этой модели управляемость может быть **иллюстративно охарактеризована** как соотношение ширины пространства возможных состояний и ширины пространства доступных управляющих воздействий:

$C \approx 5 / 3$

Численное значение в данном примере носит исключительно поясняющий характер и не претендует на универсальность.

IV. Планирование (P)

1. Определение

Планирование **P** возникает в тех случаях, когда удержание системы **B** в пределах окна устойчивости **L** невозможно обеспечить исключительно мгновенной реакцией **out** на текущие воздействия **in**. Иными словами, **P** появляется тогда, когда между входным воздействием и потенциальным критическим исходом существует временной зазор, в пределах которого отсутствие своевременного действия приводит к выходу системы за границы **L** в будущем.

2. Граничные условия появления P

Если мгновенной реакции достаточно для удержания системы в **L**, планирование не требуется:

$in(t_o) \rightarrow out(t_o) \rightarrow in(t_o) \in L$

Если же для сохранения системы в **L** необходимо совершить действия **заранее**, до наступления опасных условий, возникает планирование:

$out(t_o) \rightarrow in(t_1) \in L$

Таким образом, планирование — это сдвиг регуляции в режим опережающего действия: выбор $out(t_o)$ определяется требованием удержать $in(t_1)$ в L.

В рамках метатеории **P фиксируется как бинарное архитектурное свойство**:

- либо удержание в **L** возможно за счет мгновенной регуляции;
- либо требуется сдвиг регуляции во времени.

P = да / нет

3. Планирование и сознание

Наличие планирования **повышает вероятность появления сознания**, поскольку требует:

- учета будущих состояний;
- соотнесения текущих действий с отложенными последствиями;
- координации поведения вне текущего момента.

4. Примеры по возрастанию роли P

4.1. Микроорганизм в химическом потоке

Микроорганизм находится в среде, где концентрации веществ изменяются, а полезные и опасные зоны чередуются. Он реагирует на локальные химические градиенты, усиливает движение к благоприятным участкам и избегает опасных. В данном случае регуляция преимущественно реактивна: планирование минимально или отсутствует, а временной горизонт почти совпадает с текущим состоянием среды. Это пример агента, для которого P практически не требуется.

4.2. Дерево и сброс листвы

Дерево сбрасывает листву осенью:

- до наступления заморозков;
- до дефицита воды;
- до механических повреждений от налипшего на листья снега.

Процесс запускается по опережающим признакам:

- сокращение светового дня;
- изменение температуры;
- внутренние биохимические каскады.

Здесь важно, что:

- опасность еще не наступила;
- действие совершается заранее;
- регуляция ориентирована на будущие условия.

Это пример случая, в котором **P присутствует**, однако сознание не требуется.

4.3. Зверек и приближение ночи

Зверек прячется до наступления ночи, поскольку ночью:

- ухудшается видимость;
- возрастает риск встречи с хищниками;
- времени на поиск укрытия может оказаться недостаточно.

Если реакция будет основана только на текущих входных воздействиях (**in**), момент для безопасного действия может быть упущен, а последствия окажутся необратимыми. В данном случае планирование становится критическим: ошибка во времени равносильна выходу за пределы окна устойчивости **L**, и возрастает значение учета будущих состояний.

V. Горизонт планирования (T)

1. Определение

Горизонт планирования **T** — это архитектурная характеристика системы, задающая диапазон каузальных связей, которые должны быть учтены, чтобы действия агента **B**

out → … → in

приводили к удержанию системы в пределах окна устойчивости **L**.

T включает:

- временную дистанцию между действием и функционально значимым эффектом;
- количество промежуточных каузальных звеньев;
- степень устойчивости или нестабильности связей между **out** и будущими состояниями **in**.

2. Малый горизонт планирования T

При малом **T**:

- последствия действий проявляются быстро;
- причинно-следственные связи локальны и устойчивы;
- текущие **in** надежно предсказывают ближайшие **in'**.

В таких условиях:

- реактивные агенты оказываются достаточными;
- мгновенная регуляция **in → out** обеспечивает удержание в **L**.

Архитектуры, требующие внутренней модели или опережающего планирования, не дают функционального преимущества.

3. Большой горизонт планирования T

При большом **T**:

- между действием и значимым результатом возникает длинная цепочка промежуточных факторов;
- локальные **out(t_0)** перестают быть надежным предиктором **in(t_1)**;
- причинно-следственные связи становятся размытыми и нестабильными;
- корреляция между текущими действиями и будущими состояниями снижается.

В предельном случае:

- динамика становится практически непредсказуемой;
- прогнозирование теряет ценность;
- регуляция перестает обеспечивать устойчивое удержание в **L**.

В таких условиях:

- ни одна архитектура (включая сознание) не гарантирует стабильной регуляции;
- поведение может быть лишь статистическим или адаптивным в ограниченном смысле.

4. Рабочий диапазон T: плоские архитектуры

Между этими крайностями существует диапазон **T**, в котором:

- причинно-следственные связи сохраняются;
- планирование дает выигрыш.

В этом диапазоне эффективны плоские архитектуры регуляции, например:

- биологические системы (нейронные сети без выраженной иерархии);
- искусственные системы (классические нейросети, модели с весами и узлами);
- физические регуляторы (аналоговые схемы, химические реакции).

Такие архитектуры:

- аппроксимируют нелокальные зависимости;
- используют прошлый опыт;
- функционируют за счет статистики.

5. Узкий диапазон T: экспоненциальная вариативность

Внутри рабочего диапазона **T** существует более узкий диапазон, где:

- число возможных траекторий действий растет экспоненциально;
- вариативность превышает возможности прямой аппроксимации;
- простое сопоставление входов и выходов перестает масштабироваться.

В этой зоне:

- плоские архитектуры теряют эффективность;
- требуется механизм: отбора, фокусировки, временного сжатия пространства вариантов.

Именно в этом диапазоне сознание не становится обязательным, но может оказаться более эффективным по сравнению с чисто плоскими регуляторными схемами.

VI. Сознание как архитектурная альтернатива

Ключевая особенность плоской архитектуры заключается в том, что она решает задачу регуляции за счет масштабирования ресурсов, а не за счет изменения архитектурной организации.

В рамках данной теории **сознание** рассматривается как **архитектурный ответ на ограниченность ресурсов**. В этом контексте сознание можно описать как механизм, который:

- не хранит и не обрабатывает все пространство вариантов одновременно;
- не активирует всю архитектуру постоянно;
- работает через временное выделение, фокусировку и упрощение.

Это не делает регуляцию оптимальной в абсолютном смысле, но позволяет снизить ресурсные затраты при заданных ограничениях среды и архитектуры. Сознание представляет собой компромисс между глубиной каузального учета и ограниченностью вычислительных ресурсов агента.

1. Биология как частный случай

Сознание хорошо известно нам прежде всего потому, что оно стало устойчивым решением в биологической эволюции. Теория не утверждает, что сознание возможно только в биологической форме. В принципе, теория не отрицает, что сознательные архитектуры могут быть реализованы и в небиологических системах.

Однако в биологической эволюции действуют ограничения, при которых:

- энергетические ресурсы ограничены;
- нейронная ткань дорога в производстве и поддержании;
- постоянное увеличение плотности и масштаба плоских архитектур оказывается энергетически невыгодным.

Именно поэтому в биологии сознание стало устойчивым решением — не потому, что оно «особое», а потому, что оно **достаточно эффективно при данных ограничениях**. В иных условиях — при других ресурсах, иной динамике среды и иной стоимости вычислений — вполне возможны более эффективные архитектуры регуляции без сознания.

2. Принципиальное разграничение

Сознание в данной теории не рассматривается как обязательный этап эволюции агентов, как высшая форма регуляции или как универсальный критерий адаптивности. Оно описывается исключительно функционально — как один из возможных архитектурных ответов на специфический набор ограничений.

3. Психологическая стена

В исследованиях сознания часто предполагается наличие качественного разрыва между физическими регуляторными процессами и феноменом субъективности первого лица.

Дополнительное затруднение возникает из-за интуитивной иерархии «сложности», в рамках которой сознание рассматривается как качественно иной уровень по сравнению с химической регуляцией или реактивным поведением организмов. Термин «сложность» используется здесь как отражение субъективной интуиции наблюдателя и не рассматривается как строгий теоретический параметр.

Интуиция скачка возникает из иерархического способа классификации процессов — от химических реакций к организму и далее к сознанию. В данной теории такая иерархия рассматривается как инструмент описания, а не как фундаментальная структура реальности. Соответственно, переход к сознанию интерпретируется как изменение режима регуляции, а не как качественный онтологический разрыв.

4. Методологический статус теории

Метатеория функционального наблюдателя задает направление анализа сознания через структурные параметры регуляции — прежде всего **L**, **C** и **T**. Эти параметры определяют архитектурные условия, при которых возникает необходимость в определенных режимах регуляции, включая планирование и сознание.

При этом теория не задает:

- конкретных единиц измерения L, C и T;
- универсальных способов их вычисления;
- прикладных моделей их количественной оценки.

Разработка методов измерения, формализации и эмпирической верификации этих параметров относится к области прикладных дисциплин — теории управления, нейронауки, когнитивной науки, инженерии и искусственного интеллекта.

Метатеория задает структурную рамку и параметры анализа, в конкретные способы их операционализации остаются задачей специализированных научных направлений.

VII. Модель мира (W)

1. Определение

Для регуляции в реальности **R** агент **B**, обладающий сознанием **Q**, формирует внутреннюю модель, обозначаемую **W**. W вводится как архитектурный компонент агента B и не рассматривается как отдельная онтологическая область.

Принципиально важно:

- **W ≠ R**: внутренняя модель не является копией реальности;
- W не обязана сохранять масштаб, метрику или структуру R;
- структура W определяется задачами регуляции и ресурсными ограничениями агента B.

Модель W формируется в результате обработки входных воздействий (**in**) и внутренних состояний B и используется для регуляции, планирования и оценки. W не обладает самостоятельным каузальным выходом в R. Ее влияние на реальность осуществляется исключительно опосредованно через регуляторные процессы агента B. Все действия, изменяющие R, выполняются агентом B. В этом смысле W является каузально односторонней по отношению к R: она может изменяться и реконфигурироваться внутри B, но не может напрямую воздействовать на R.

2. Операционное отождествление: W ≈ R’

В процессе регуляции система **B** использует внутреннюю модель **W** как операциональный заместитель реальности.

Формально:

W ≈ R’**,**

где **R’** — та часть или версия реальности, которая доступна системе в рамках текущих входных воздействий, различима, сопоставима и пригодна для выбора действий.

Это соответствие:

- локально — ограничено текущей задачей и доступным горизонтом планирования;
- функционально — необходимо для осуществления регуляции;
- операционально — не предполагает онтологического тождества W и R.

Регуляторные операции не могут выполняться при постоянном пересмотре соответствия модели реальности. Поэтому система действует так, как если бы W являлась релевантной реальностью для текущего выбора действий, не выводя из этого утверждений о структуре R как таковой.

3. Вложенность и ограниченность модели W

Внутренняя модель **W** может включать несколько уровней описания.

Например:

1. **Первый уровень** — модель текущей ситуации: агент различает окружающий мир и свое положение в нем.
2. **Второй уровень** — модель возможного перемещения: агент способен представить себя в иной локации или ином состоянии и оценить последствия такого изменения.
3. **Третий уровень** — модель собственной модели: агент способен учитывать, как изменится его восприятие и оценка мира при переходе в другую ситуацию.

Глубина вложенности теоретически может продолжаться далее, однако на практике она:

- конечна,
- ограничена ресурсами B (время, память, энергия),
- не образует независимых «параллельных миров», а остается частью единой регуляторной архитектуры.

Рост глубины вложенности и вариативности представлений приводит к экспоненциальному росту пространства возможных траекторий действий. При превышении доступных ресурсов простая регуляция и плоские архитектуры перестают масштабироваться: система утрачивает возможность прямого перебора и полной аппроксимации этих траекторий.

VIII. Симуляция и действие

1. Постановка различия

Агент **B** использует внутреннюю модель **W** в двух принципиально различных режимах:

- **симуляция** — моделирование возможных сценариев внутри W без прямого каузального выхода в реальность;
- **действие** — выбор поведения агентом B, приводящего к формированию выхода **out**, который необратимо изменяет дальнейшее взаимодействие с реальностью **R**.

2. Якорь реальности (A)

В архитектуре агента выделяется особый режим, обозначаемый **A**.

A (якорь реальности) — это канал каузальной фиксации действий, в котором поведение агента приобретает необратимый характер.

Это означает:

- каждое действие, реализованное через **A**, изменяет последующее взаимодействие с реальностью **R**;
- последствия могут быть компенсированы или частично скорректированы, но не могут быть отменены в строгом каузальном смысле;
- любая попытка исправления последствий является новым действием и входит в каузальную цепь, порождая дальнейшие изменения.

3. Симулятивные модели и их роль

Наряду с якорным каналом A агент B может использовать внутреннюю модель W в симулятивном режиме, формируя различные конфигурации представлений, например: планы, гипотезы, воспоминания и фантазии.

Общее свойство симулятивного режима состоит в том, что он не приводит к формированию каузального выхода out в реальность R.

В этом режиме:

- допускается многократное проигрывание сценариев;
- возможен возврат к предыдущим состояниям модели;
- альтернативы могут сравниваться без необратимой фиксации результата.

Симуляции выполняют подготовительную и оценочную функцию: они позволяют перераспределить вероятность успешного удержания в пределах L, однако сами по себе не изменяют ход взаимодействия с реальностью.

Каузальная фиксация происходит только при переходе к действию через канал A.

4. Активная модель WA

В каждый момент времени среди множества внутренних моделей W агент B использует одну модель как операционный заместитель реальности при формировании выхода out. Обозначим эту активную модель как WA.

Тогда:

R → in → B → WA → B → out → R

Это означает:

- агент B может поддерживать множество моделей W в симулятивном режиме;
- однако только одна модель в данный момент участвует в формировании каузального выхода;
- именно через WA формируется последовательность действий out, имеющих необратимые последствия.

5. Почему A-канал единственный

Ограничение «не более одного A» не относится к вычислениям и не запрещает параллельные процессы внутри B.

Это каузальное ограничение:

- агент B может реализовать только одну фактическую последовательность действий в реальности R;
- невозможно одновременно осуществлять несколько несовместимых каузальных историй;
- фиксация одной траектории автоматически исключает альтернативные.

Таким образом, множественность симуляций допускается, но каузальная история агента всегда единственна.

IX. Значимость (M) и нормы

1. Определение

Слой значимости **M**, функционально наложенный на внутреннюю модель **W** задает распределение:

- относительной важности состояний;
- приоритетов действий;
- допустимых и недопустимых потерь;
- условий, подлежащих сохранению.

M не изменяет фактическую структуру W, а определяет, какие элементы модели получают больший вес при выборе поведения. Интуитивно M можно представить как фильтр или оптический слой, где факты W остаются теми же, но меняется то, что выделяется, что становится критическим, что определяет направление выбора.

2. Значимость как нормативная структура

M задает не только предпочтения («что лучше»), но и ограничения («что нельзя нарушать»).

Поэтому M включает две взаимосвязанные стороны:

- **ценностную** — ранжирование состояний и траекторий (лучше / хуже);
- **нормативную** — фиксацию пределов, выход за которые считается неприемлемым.

В рамках теории действия оцениваются не только по немедленному выигрышу, но по степени соответствия структуре M. Иными словами, выбор считается допустимым или недопустимым относительно заданной конфигурации значимости.

3. Функция M при ограниченных ресурсах

Поскольку ресурсы B ограничены (время, память, энергия), система не может одинаково обрабатывать все пространство W.

M решает эту проблему архитектурно:

- усиливает обработку релевантных элементов;
- подавляет нерелевантные;
- задает порядок сравнения альтернатив;
- определяет распределение «цены ошибки» при дефиците ресурсов.

Таким образом, M делает пространство W управляемым не за счет сокращения самой модели, а за счет перераспределения приоритетов обработки и выбора.

X. Само-индексация (S) и границы агента

1. Определение

S — это функциональный оператор, вводящий в модели **W** различение:

- относящееся к агенту B / не относящееся к агенту B;
- последствия для B / последствия вне B;
- состояния, учитываемые как изменения самого B.

S определяет:

- какие элементы модели интерпретируются как состояния агента;
- к каким изменениям приписывается авторство B;
- какие последствия считаются последствиями «для системы».

S — не образ себя, не нарратив и не описание личности. Это оператор принадлежности и индексирования.

2. S в симуляциях

Оператор S может присутствовать в любой внутренней конфигурации W:

- планах;
- сценариях;
- мысленных симуляциях;
- воспоминаниях.

Это необходимо потому, что планирование требует заранее определять, какие последствия относятся к самому агенту. Даже нереализованные сценарии должны быть привязаны к B как носителю рисков и действий.

3. Разделение модели: Wb и Wr

При наличии S внутри W возникает функциональное разделение:

- **Wb** — подмодель, интерпретируемая как состояния самого B (ресурсы, повреждения, устойчивость, уязвимости);
- **Wr** — подмодель, интерпретируемая как остальная реальность (среда, предметы, другие агенты, структуры).

Граница между Wb и Wr не является жесткой. Она определяется степенью каузальной включенности элементов в контур удержания системы в L.

4. Динамичность границ

S не обязан совпадать с анатомическими границами тела и не фиксирован в пространстве. Расширение Wb происходит тогда, когда внешний элемент становится каузально включенным в регуляцию так, что:

- его повреждение функционально эквивалентно повреждению самого B;
- действия через него становятся продолжением действий агента.

Поэтому в область «относящегося к B» могут временно или устойчиво входить:

- инструменты и экипировка;
- транспорт;
- протезы и интерфейсы;
- внешние ресурсы, критически включенные в регуляцию;
- координированные роли и совместные контуры, если они каузально необходимы для удержания в L.

5. Связь S и M

В типичном случае карта значимости **M** имеет максимум на состояниях Wb — сохранности регулятора. Однако M может устойчиво максимизироваться на элементах Wr — например, на ребенке, группе, норме или идее. Тогда высказывание «он важнее меня» в рамках теории означает: в структуре значимости M соответствующие элементы Wr имеют больший вес, чем состояния Wb в данном классе сценариев. Иными словами, данные элементы Wr оказываются приоритетнее собственных состояний Wb для данного агента. Альтруизм и самопожертвование интерпретируются как частные случаи соблюдения M, а не как нарушение само-индексации S.

6. Разведение функций A, S и M

Архитектурно:

- **A** различает симуляцию и необратимое вмешательство;
- **S** вводит само-индексацию — для кого учитываются последствия;
- **M** задает структуру значимости — что считается важным.

Эти операторы независимы и могут комбинироваться в режимах симуляции и реального действия.

XI. Первое и третье лицо

1. Почему WAr структурно «от первого лица»

В якорном режиме A внутренняя модель реальности WAr формируется из контура:

R → in → B → out → R

и используется для генерации действий, имеющих каузально необратимые последствия.

Поэтому WAr неизбежно индексирована на саму систему B — в координатах ее сенсоров, эффекторов и окна устойчивости L.

Модель «от третьего лица» требует дополнительных операций:

- перехода во внешнюю систему координат;
- реконструкции скрытых параметров;
- введения абстрактного наблюдателя.

Эти операции увеличивают вычислительную нагрузку и неопределенность. В симулятивных режимах подобная реконструкция допустима. Однако в A-канале она функционально невыгодна, поскольку цена ошибки необратима. Следовательно, «первое лицо» в WAr — это не нарратив и не субъективное ощущение, а структурная форма индексирования модели на владельца in/out и на допустимые потери относительно L. Это минимизация промежуточных реконструкций между входом, состоянием системы и действием.

2. «Третье лицо» как модель внутри W

Так называемый «опыт от третьего лица» не является самостоятельным режимом данности. Он всегда представляет собой опыт от первого лица, содержанием которого является модель, описанная в формате «третьего лица». Иначе говоря, «третье лицо» — это способ организации представления внутри W, а не независимый формат существования. Например, измерительный прибор может зафиксировать физическое изменение. Однако «градусник», «Солнце» и «температура в градусах» являются элементами модели W. Физический след в устройстве может существовать как состояние R. Но его статус как «измерения температуры» требует шкалы, единиц и интерпретации — то есть принадлежит уровню W. Если предположить отсутствие агентов, то исчезают не изменения в R, а их представление в терминах объектов и измеримых величин. Даже если остаются физические переходы, то исчезает их модельная организация. Таким образом, «объективность» третьего лица не устраняется, а переосмысливается как устойчивость и воспроизводимость модели W при смене агентов. Она означает стабильность структуры представления, а не существование привилегированного формата вне всякой модели.

XII. Квалиативный формат (Q)

1. Определение

В данной теории все содержимое внутренней модели W дано агенту B в квалиативном формате Q.

При этом:

- без модели W квалиативность невозможна;
- Q не является объектом, процессом или элементом W;
- вопрос «где находится Q» является категориальной ошибкой.

Q — это не дополнительная сущность, а способ данности модели системе.

Формат представления не может быть включен в состав представляемого и не подлежит локализации среди элементов содержания. Мозг, нейроны, сигналы и их динамика — элементы модели W. Q — это режим, в котором эта модель вообще может быть дана агенту.

Попытка обнаружить Q «в мозге» смешивает уровень формата и уровень содержания: ищется способ представления среди представляемых элементов.

В классической формулировке (Chalmers, 1996) трудная проблема возникает как разрыв между физическим описанием процессов и субъективным переживанием. В предлагаемой архитектуре этот разрыв интерпретируется как категориальное смешение уровня формата (Q) и уровня содержания модели (W). Подобные смешения обсуждались и в контексте различения феноменального и функционального сознания (Block, 1995), однако здесь делается шаг к формальной архитектурной реконструкции этого различия.

2. Разведение R, W и Q

Хотя W является моделью реальности R, между ними существует структурное различие:

R → B[Q] → W

где:

- R — каузально первична;
- W — модель;
- Q — формат, в котором W вообще может быть дана агенту.

Стрелки не обозначают новую причинность. Q не вводит дополнительной каузальной сущности и не вмешивается в физику. Все каузальные цепи полностью описываются через контур in/out системы B.

3. Q не является причиной

Q:

- не добавляет новых физических процессов;
- не конкурирует с нейронными описаниями;
- не является объяснительным «остатком».

Он фиксирует режим организации модели, который становится архитектурно выгодным при перегрузке плоских схем.

4. Сознательная архитектура

Сознание определяется архитектурно как:

Сознательная архитектура = (W, A, S, M) в режиме Q

где:

- W — активная модель;
- A — канал каузальной фиксации;
- S — само-индексация;
- M — структура значимости;
- Q — формат их совместной данности как единого текущего режима.

Q не добавляет новой сущности — он организует связку.

5. Ставочное состояние

В каждый момент активная модель собирается в единую ставочную конфигурацию:

- W — «что происходит»;
- A — «что станет необратимым»;
- S — «для кого»;
- M — «что важно».

Q — это формат, в котором эта конфигурация существует как единое текущее состояние, а не как распределенный набор независимых вычислений.

6. О моделях и онтологии

Наука оперирует рабочими моделями, ценность которых определяется регулятивной и предсказательной эффективностью. Атомы, поля и нейроны — элементы описания, но не окончательные онтологические утверждения. Когда модель принимается за реальность, возникают псевдопроблемы — в частности, попытка разместить сознание «поверх» уже описанного мира.

В данной теории сознание не добавляется к миру. Оно описывается как архитектурный режим организации модели относительно реальности.

7. Статус «трудной проблемы»

Трудная проблема сознания, сформулированная Дэвидом Чалмерсом (1995), ставит вопрос: почему и каким образом физические процессы в мозге сопровождаются субъективным опытом? Иначе говоря, почему система не просто функционирует, а «что-то чувствует»?

В рамках FOM предпосылка этой формулировки требует уточнения. То, что называется «физическими процессами», «нейронными вспышками», «объектами» и «причинностью», уже принадлежит уровню модели реальности W, а не непосредственно R. Мы не имеем доступа к R вне формата модели; нам даны лишь структурированные представления внутри W. Это не означает отрицания существования R, а лишь фиксирует, что все описания, включая физические, относятся к уровню модели. Следовательно, противопоставление «физических процессов» и «субъективного опыта» формируется внутри самой модели W. Проблема принимает вид требования вывести формат представления (Q) из представляемых состояний (W). Здесь возникает категориальное смешение: свойства содержания модели интерпретируются как способные породить режим ее данности. Однако формат не является элементом множества содержаний; он задает способ их организации и доступности системе.

Требование объяснить квалиативность через физические процессы тем самым предполагает вывод формата из его собственных элементов, что делает исходную постановку логически некорректной. Поэтому после разведения уровней проблема не «решается» в традиционной формулировке, а требует пересмотра исходных предпосылок. Вопрос смещается с «как физика производит опыт» к «какова архитектура, при которой модель вообще может быть дана системе».

Метатеория функционального наблюдателя не предлагает метафизического объяснения существования данности как таковой. Она ограничивается архитектурным анализом условий, при которых внутренняя модель может организовываться в квалиативный режим.

XIII. Операционализация квалиативного режима

1. Операциональное определение через q(t)

Введем операциональный коррелят квалиативного режима — состояние **q(t)**, которое будем называть *ставочным узлом*. Агент B находится в квалиативном режиме **Q** в момент времени *t*, если существует состояние q(t), удовлетворяющее следующим условиям:

(Q1) Необратимая фиксация

Интервенция в q(t) при прочих равных условиях систематически изменяет выбор действия, переходящего в режим необратимой фиксации. Иначе говоря, изменение q(t) меняет то, что будет зафиксировано в реальной каузальной истории.

(Q2) Кросс-модульная сшивка

q(t) выступает общей точкой согласования для нескольких подсистем:

- оценка ситуации;
- планирование;
- память;
- оценка риска и потерь.

Через q(t) разрешаются конфликты между альтернативами в едином режиме.

(Q3) Концентрация значимости

При дефиците ресурсов система перераспределяет вычисления так, чтобы сохранять стабильность и точность процессов, связанных с q(t), жертвуя периферийными задачами. Потенциальная цена ошибки концентрируется вокруг состояний, связанных с q(t).

(Q4) Индексация настоящего относительно A

q(t) индексирует различие между потенциальными возможностями и уже зафиксированными последствиями.

До порога необратимости система сохраняет гибкость изменения намерения, после — наблюдается смена режима коррекции.

Эти условия:

- не требуют языка;
- не требуют отчетности;
- задают проверяемый класс архитектурных свойств.

2. Разведение Q и q(t)

Важно различать:

- **Q** — инвариант уровня описания (формат данности);
- **q(t)** — функциональный коррелят в механизмах WASM.

Интервенции осуществляются не «в Q», а в механизмы, реализующие q(t). Q — свойство архитектурного режима, проявляющееся через организацию q(t).

3. Наблюдаемые сигнатуры квалиативного режима

Из условий (Q1–Q4) следует набор эмпирически проверяемых признаков.

(S1) Узкое горлышко ставок

При попытке поддерживать две независимые ставки, требующие необратимой фиксации, возникают:

- интерференция;
- рост времени реакции;
- увеличение ошибок.

(S2) Кросс-модульная причинность

Интервенция в q(t) приводит к согласованным изменениям:

- выбора действия;
- распределения внимания;
- активации памяти;
- структуры плана.

(S3) Защита ставки при перегрузке

При дефиците ресурсов система сохраняет процессы, связанные с q(t), жертвуя вторичными задачами.

(S4) Порог необратимости

Наблюдается различимый порог необратимости:

- до него намерение гибко изменяется;
- после — возникает инерция и иной стиль коррекции.

Приложение. Сравнение теорий сознания и метатеории функционального наблюдателя (FOM)

Данное приложение не ставит целью опровергнуть существующие теории сознания. Его задача — показать, на каком архитектурном уровне работает каждая из них, какие аспекты сознания она адекватно описывает и в каком смысле ее охват отличается от метатеории функционального наблюдателя (FOM).

Метатеория FOM не выступает как конкурент частных теорий, а задает рамку анализа, в которой становится видно:

- какую часть архитектуры регуляции фиксирует каждая теория;
- какие связи между компонентами остаются неявными или неописанными;
- в каких случаях происходит перенос объяснения между различными уровнями (категориальное смешение).

В терминах FOM различаются следующие архитектурные компоненты:

- **R** — каузально первичная реальность;
- **B** — регулятор (контур in/out, удержание в окне жизни L);
- **W** — внутренняя модель реальности;
- **A** — якорный канал необратимого действия;
- **S** — само-индексация («для кого», границы «своего»);
- **M** — карта значимости и норм (что важно / недопустимо);
- **Q** — квалиативный режим организации ставочного состояния.

1. Иллюзионизм

Иллюзионизм (Dennett, 1991; Frankish, 2016) утверждает, что феноменальные свойства не требуют отдельного онтологического статуса. Существуют лишь функциональные, репрезентативные и каузальные процессы, а феномен «кажется, что есть цвет/боль/вкус» не требует введения особого уровня бытия. Сознание, в этом понимании, — это определенный тип когнитивной организации, а не дополнительная сущность сверх физики.

В рамках FOM иллюзионизм оказывается частично прав.

Он корректно указывает на то, что:

- цвет, вкус, боль и другие феноменальные содержания не являются свойствами R как каузально первичной реальности;
- они принадлежат операциональной реальности R', то есть содержанию внутренней модели W;
- многие феноменальные характеристики возникают как конструктивные особенности репрезентации, а не как «свойства вещей самих по себе».

В этом смысле иллюзионизм прав против наивного реализма: содержание опыта не является прямым свойством внешнего мира. Однако в терминах FOM возникает различие, которое иллюзионизм обычно не проводит. Необходимо различать содержание модели (W) и формат, в котором эта модель вообще дана системе (Q). Иллюзия — это несоответствие внутри содержания W относительно R. Например, мираж или ложная интерпретация цвета — это ошибка репрезентации. Но сама возможность говорить об ошибке уже предполагает, что модель дана системе в некотором режиме доступа. Q в FOM — это не дополнительный объект внутри W и не «субстанция переживания», а формат организации ставочного состояния: WASM. Когда иллюзионизм объявляет квалиа иллюзией, он корректно демистифицирует содержание. Однако если вместе с этим отрицается сама необходимость формата данности, возникает напряжение.

Если сознание — это полностью «ошибка», то в каком режиме эта ошибка фиксируется? Ошибка предполагает:

- систему, для которой что-то является ошибкой;
- различие между корректным и некорректным;
- ставочное состояние, относительно которого возможно рассогласование.

То есть сама структура «кажется, что…» уже указывает на наличие некоторого режима организации модели. В FOM этот режим обозначается как Q. Таким образом, иллюзионизм может быть интерпретирован как теория конструктивности содержания W, но не как отрицание архитектурного режима Q. Отрицая квалиа как объекты, он остается убедительным. Отрицая необходимость формата данности как архитектурного свойства регулятора, он фактически вынужден опираться на тот самый режим, который стремится устранить.

Итог

- Иллюзионизм корректен как теория конструктивности и ошибочности содержания W.
- Он убедительно демонстрирует, что феноменальные свойства не являются свойствами R.
- Однако он не устраняет необходимость различать содержание модели и формат ее данности.
- В терминах FOM иллюзионизм отрицает объекты, но не отменяет режим Q как архитектурное условие возможности самой ошибки.

2. Панпсихизм

Панпсихизм в современных версиях (Chalmers, 2015; Goff, 2017) утверждает фундаментальность опыта. Панпсихизм — это семейство позиций, объединенных тезисом о фундаментальности ментального или прото-ментального в структуре реальности. Однако внутри этого семейства существуют существенно различающиеся версии.

1. Конститутивный панпсихизм

Согласно конститутивной версии панпсихизма, фундаментальные элементы реальности обладают некоторой формой прото-опыта, а сложные формы сознания возникают как композиции этих базовых элементов. Индивидуальное сознание в этом подходе мыслится как результат объединения множества микроскопических «опытов». Центральная трудность этой позиции — так называемая проблема комбинации: как множество элементарных форм субъективности складываются в единое макроскопическое сознание?

В FOM эта проблема получает иную интерпретацию. Проблема комбинации возникает потому, что сознание онтологически локализуется на уровне R — как свойство базовых элементов реальности. Тогда необходимо объяснить, почему и по какому принципу именно эти элементы образуют более высокий субъект, а другие — нет.

В онтологическом смысле в R не существует «привилегированных систем». Любая система есть результат выбора модели.

Например, металлический синий круг одновременно входит в:

- систему геометрических фигур;
- систему электромагнитных взаимодействий;
- систему химических соединений;
- систему цветовых классификаций;
- систему промышленных объектов.

Никакая из этих систем не обладает онтологическим преимуществом перед другими. Они различаются по способу описания, а не по степени реальности.

Если квалиативность приписывается элементам R как их онтологическое свойство, то необходимо ответить:

- по какому принципу именно одна из возможных систем становится носителем «единого опыта»?
- почему одни комбинации элементов образуют субъект, а другие — нет?

Проблема комбинации тем самым оказывается следствием предположения, что существует онтологически заданная иерархия систем.

В FOM такая иерархия не постулируется. Единство сознательного состояния не выводится из композиции элементов R, а определяется архитектурой регулятора B:

- наличием активной модели W;
- само-индексацией S;
- картой значимости M;
- якорем необратимости A.

Единство возникает не как онтологическая сумма, а как функциональное свойство регуляторной архитектуры. Таким образом, с точки зрения FOM проблема комбинации является

не физической загадкой, а следствием онтологизации модели. Сознательное единство не нужно «собирать» из элементов реальности. Оно возникает как результат определенной архитектуры регуляции, в которой активная модель организуется в ставочное состояние.

2. Космопсихизм (глобальный панпсихизм)

Космопсихизм (Shani, 2015) утверждает, что существует единое фундаментальное сознание (космический субъект), а индивидуальные сознания — производные или частные аспекты этого целого. В этой версии сознание не распределено по частицам, а глобально. Именно здесь возникает интересная точка соприкосновения с FOM.

Если понимать утверждение космопсихизма так: все, что доступно субъекту, существует в форме внутренней модели, то в операциональном смысле мир W действительно «внутри» квалиативного режима Q.

В FOM:

- весь доступный мир для агента существует как W;
- W дана в режиме Q;
- следовательно, весь доступный «мир» для агента существует в формате сознания.

В этом ограниченном эпистемическом смысле космопсихическая интуиция совпадает с FOM: для агента весь его мир действительно «внутри» режима данности. Однако FOM не делает онтологического вывода, что вся R является сознанием.

Различие остается:

- космопсихизм делает метафизический тезис о природе R;
- FOM делает архитектурный тезис о формате доступа к W.

3. Прото-панпсихизм

Некоторые версии утверждают не наличие полноценного сознания в фундаментальных структурах, а наличие прото-ментальных свойств — нечто, что при определенной организации дает сознание. Это попытка ослабить онтологическое обязательство.

В FOM подобный ход оказывается избыточным, поскольку:

- квалиативность не является фундаментальным свойством R;
- она возникает как режим организации регулятора B при определенной архитектуре (WASM).

То есть FOM объясняет появление квалиативного режима через архитектуру, а не через базовые свойства материи.

Где панпсихизм прав

Панпсихизм корректен как интуиция о неустранимости формы данности.

Он справедливо указывает, что:

- нельзя объяснить опыт, полностью игнорируя сам факт данности;
- сознание не выглядит как «надстройка», добавляемая постфактум.

В этом смысле панпсихизм верно чувствует, что формат данности не редуцируется к описанию объектов.

Где происходит расхождение

Различие возникает в онтологическом выводе. Панпсихизм заключает: если квалиативность неустранима, значит она свойство самой реальности (R). FOM заключает иначе: если квалиативность неустранима, значит она архитектурное свойство регулятора (B), а не онтологическое свойство R.

Таким образом:

- панпсихизм локализует квалиативность на уровне онтологии;
- FOM локализует ее на уровне архитектуры.

Итог

- Конститутивный панпсихизм сталкивается с проблемой комбинации, поскольку онтологизирует формат.
- Космопсихизм частично совпадает с FOM в эпистемическом смысле (мир как W внутри Q), но расходится в онтологическом выводе.
- Прото-панпсихизм пытается смягчить позицию, но остается в рамках фундаментализации ментального.

В терминах FOM:

- панпсихизм корректен как тезис о неустранимости формата (Q);
- но избыточен как универсальная теория онтологии реальности (R).

Он верно фиксирует, что квалиативность нельзя устранить, но делает дополнительный метафизический шаг, который FOM считает необязательным.

3. Global Workspace Theory (GWT)

Global Workspace Theory (Baars, 1988; Dehaene & Changeux, 2011) описывает архитектуру, в которой множество специализированных подсистем обрабатывают информацию параллельно, а часть этой информации становится «глобально доступной» через механизм широкого распространения (broadcasting). Сознательное состояние в GWT связано с тем, что определенное содержание получает доступ к глобальному рабочему пространству и тем самым становится координирующим центром поведения.

GWT тем самым объясняет:

- почему сознательные содержания обладают широкой доступностью;
- почему они координируют память, внимание и принятие решений;
- почему существует ограничение на одновременное присутствие нескольких «сознательных» состояний (узкое горлышко).

В терминах FOM GWT работает преимущественно на уровне организации регулятора B.

Она описывает:

- как информация распространяется между подсистемами;
- как возникает единый координационный узел;
- как обеспечивается синхронизация обработки.

Однако в FOM проводится дополнительное различие, которое в GWT обычно не фиксируется.

1. Уровень модели (W)

GWT описывает циркуляцию и доступность информации, но не формулирует явного метатеоретического тезиса о том, что активное содержание представляет собой модель реальности (W) в строгом смысле. То есть GWT объясняет как информация становится глобально доступной, но не отвечает на вопрос: что именно представляет собой эта информация по отношению к реальности R?

В FOM же активное состояние рассматривается как WA — активная модель реальности, участвующая в ставочной регуляции.

2. Уровень формата (Q)

GWT описывает функциональную интеграцию, но не вводит различие между содержанием и форматом его данности. В GWT «сознательное» — это то, что стало глобально доступным. В FOM этого недостаточно.

Глобальная доступность не равна квалиативному режиму. Q — это не просто широкое распространение информации, а режим, в котором активная модель организуется как единое ставочное состояние:

- индексированное через S («для кого»);
- структурированное через M (что важно);
- связанное с A (необратимость действия).

GWT объясняет механизм координации, но не фиксирует различие между интеграцией как вычислительной функцией и ставочной организацией модели относительно риска и необратимости.

3. Связь с необратимостью (A)

В FOM центральным элементом сознательной архитектуры является связь активной модели с якорем реальности A — с порогом необратимости. Сознательное состояние — это не просто доступное содержание, а состояние, относительно которого закрываются альтернативы.

GWT, как правило, не вводит различие между симулятивной обработкой и фиксацией действия в реальной каузальной истории. Она описывает распределение информации, но не делает необратимость (в терминах FOM — A) центральным элементом теории.

4. Связь с само-индексацией (S) и значимостью (M)

GWT не специфицирует:

- как именно формируется «для кого» данное состояние (S);
- как карта значимости (M) структурирует выбор и концентрирует цену ошибки.

Хотя в прикладных моделях GWT используются понятия внимания и ценности, они не выделяются как отдельные архитектурные компоненты сознания.

В FOM же сознательная архитектура — это связка: WASM в режиме Q.

GWT описывает глобальное распространение, но не формулирует эту связку как структурное условие.

Итог

- GWT корректна как теория глобального доступа и координации процессов внутри регулятора.
- Она убедительно объясняет ограниченность сознания, его роль в интеграции и управлении поведением.
- Однако она не проводит явного различия между моделью (W), форматом ее данности (Q), само-индексацией (S), значимостью (M) и якорем необратимости (A).
- Поэтому в терминах FOM GWT описывает транспортно-интеграционный слой архитектуры, но не охватывает целиком ставочную организацию сознательного состояния.

В этом смысле GWT может рассматриваться как частный механизм внутри более широкой архитектурной рамки, но не как исчерпывающая теория сознания.

4. Integrated Information Theor y (IIT)

Integrated Information Theory (Tononi, 2004; Oizumi et al., 2014)предлагает формальную теорию сознания, в которой центральным критерием является степень интеграции причинной структуры системы. Сознательное состояние определяется как структура с ненулевым значением Ф — то есть как система, в которой целое обладает причинной мощностью, не сводимой к сумме частей. IIT тем самым корректно фиксирует важный аспект: сознание связано не с простым наличием информации, а с определенной организацией причинных связей, где система образует интегрированное единство. Это сильная сторона IIT: она вводит формальный критерий структурной целостности. Однако в терминах FOM возникает дополнительное различие.

1. Интеграция и регуляторная ставка

IIT описывает степень интеграции причинных структур, но не вводит различие между интеграцией как структурным свойством системы и интеграцией, включенной в контур необратимой регуляции.

В FOM центральным элементом сознательной архитектуры является связь активной модели с:

- якорем необратимости (A),
- окном жизни (L),
- картой значимости и допустимых потерь (M),
- само-индексацией (S).

Сознательная ставка — это не просто интеграция, а интеграция, включенная в ситуацию риска и необратимости.

Система может быть:

- высоко интегрированной,
- каузально плотной,
- формально сложной,

но если ее состояние не связано с:

- ценой ошибки,
- закрытием альтернатив,
- удержанием регулятора в допустимой области,

то в терминах FOM она не образует ставочную архитектуру.

2. Интеграция без нормативного измерения

IIT не вводит отдельный слой значимости (M). В ней интеграция оценивается по структуре причинных связей, а не по тому, что «важно» для системы в регуляторном смысле.

В FOM же:

- карта значимости определяет распределение цены ошибки;
- само-индексация определяет, чьи последствия учитываются;
- якорь A определяет, где закрываются альтернативы.

Интеграция без нормативной структуры может быть структурно богатой, но не обязательно включенной в ставочный режим.

3. Интеграция и формат (Q)

IIT предлагает количественный критерий (Φ), но не проводит различие между содержанием, структурной организацией и форматом данности. В FOM квалиативный режим (Q) не выводится из одной лишь степени интеграции. Он связан с архитектурной организацией связки: WASM. Иначе говоря: интеграция — необходимое условие для целостности, но недостаточное условие для ставочной организации.

Итог

- IIT корректно описывает структурную интеграцию причинных связей.
- Она убедительно фиксирует, что сознание связано с единством системы.
- Однако она не различает интеграцию как формальную целостность и интеграцию как включенность в ставочный регуляторный контур.
- Она не выделяет явно A (необратимость), M (нормативную значимость) и S (само-индексацию) как архитектурно самостоятельные компоненты.

В терминах FOM IIT описывает необходимое условие единства, но не охватывает целиком ставочную архитектуру сознания.

5. Predictive Processing / Free Energy Principle (PP / FEP)

Predictive Processing и Free Energy Principle (Friston, 2010; Hohwy, 2013; Clark, 2013) описывают систему как иерархическую модель, минимизирующую рассогласование между предсказаниями и сенсорными входами. Поведение (out) интерпретируется как активная индукция таких состояний среды, которые уменьшают предсказательную ошибку. Тем самым система удерживается в допустимой области состояний — в терминах FOM, в окне жизни L.

PP/FEP корректно фиксируют:

- центральность внутренней модели (W);
- динамическую регуляцию через цикл in → W → out;
- принцип стабилизации через минимизацию вариационной свободной энергии;
- связь поведения с поддержанием жизнеспособности.

Это сильная теория регуляции и адаптивности. Однако в терминах FOM возникает несколько дополнительных различий.

1. Регуляция и необратимость (A)

PP/FEP описывают непрерывную динамику коррекции предсказаний и действий. Но они обычно не проводят принципиального различия между симулятивной обработкой и моментом необратимой фиксации действия. В FOM якорь A — это структурное различие между возможными альтернативами и уже закрытыми возможностями. Сознательная ставка определяется не просто как снижение ошибки, а как состояние, относительно которого альтернативы закрываются и возникает цена ошибки. PP/FEP описывают динамическую стабилизацию, но не выделяют архитектурно различимый порог необратимости.

2. Стабилизация и нормативная значимость (M)

В PP/FEP минимизация свободной энергии задает общий регулятивный принцип. Однако сама функция минимизации не различает структурную ошибку и нормативную значимость.

В FOM карта значимости (M):

- распределяет цену ошибки;
- определяет допустимые и недопустимые потери;
- задает приоритеты между альтернативами.

Не вся предсказательная ошибка эквивалентна по значимости. Сознательная архитектура возникает там, где ошибка связана со ставкой. PP/FEP описывают общий принцип регуляции, но не выделяют нормативный слой как самостоятельный компонент.

3. Само-индексация (S)

В PP/FEP агент рассматривается как модель среды, включающая модель собственного тела. Однако явная архитектурная фиксация «для кого» дано состояние (S) обычно не формализуется отдельно.

В FOM само-индексация — это структурный компонент:

- граница «своего»,
- локализация последствий,
- распределение потерь и выгод.

Без S невозможно определить, чьи состояния минимизируются.

4. Формат данности (Q)

PP/FEP объясняют:

- как формируется модель,
- как она корректируется,
- как она стабилизирует поведение.

Но они не вводят различие между динамической регуляцией и форматом, в котором активная модель дана системе. В FOM квалиативный режим (Q) — это не просто активность модели, а режим ее ставочной организации: WASM. PP/FEP описывают вычислительный принцип, но не специфицируют, в каком формате система имеет доступ к своей активной модели как единому текущему состоянию.

Итог

- PP/FEP корректны как теории регуляции, предсказания и стабилизации.
- Они глубоко описывают динамику W и связь поведения с удержанием в окне жизни L.
- Однако они не проводят явного различия между симуляцией и необратимой фиксацией (A), не выделяют нормативную карту значимости (M) как отдельный архитектурный компонент, и не формулируют квалиативный режим (Q) как структурное свойство организации.

В терминах FOM PP/FEP можно рассматривать как мощную теорию регуляторной динамики, но не как завершенную теорию сознательной ставочной архитектуры, если не дополнить ее спецификацией связки ASM и режима Q.

6. Higher-Order Theories (HOT)

Higher-Order Theories (Rosenthal, 2005; Lau & Rosenthal, 2011) утверждают, что состояние становится сознательным тогда, когда существует представление второго порядка — мета-состояние, которое делает первое состояние «отнесенным к себе». Сознательность тем самым объясняется через структуру мета-репрезентации: состояние не просто возникает, а представляется как «имеющееся у субъекта».

HOT корректно фиксируют важный момент:

- сознательность связана с само-отнесением;
- состояние должно быть «для меня»;
- требуется определенная форма индексации на субъекта.

В терминах FOM это напрямую связано с компонентом S — само-индексацией.

1. HOT и само-индексация (S)

HOT в значительной мере описывают механизм, благодаря которому:

- состояние внутри W становится «моим»;
- появляется различие между просто обработкой и состоянием, приписанным субъекту;
- формируется слой мета-репрезентации.

В этом смысле HOT можно интерпретировать как теории S-подсистемы. Они хорошо отвечают на вопрос: как состояние становится «про меня»?

2. Ограничения HOT в терминах FOM

Однако в FOM само-индексация — лишь один компонент сознательной архитектуры.

HOT обычно:

- не специфицируют различие между симуляцией и необратимой фиксацией (A);
- не выделяют нормативную карту значимости (M) как самостоятельный архитектурный слой;
- не вводят различие между содержанием и форматом данности (Q).

В HOT сознательное состояние — это состояние, представленное на более высоком уровне. В FOM сознательное состояние — это ставочная конфигурация: WASM в режиме Q.

То есть:

- S — необходимо, но недостаточно;
- само-отнесение не равнозначно ставочной организации.

3. Само-отнесение и ставка

HOT объясняют, почему состояние становится «моим». Но они не делают центральным вопрос:

- где закрываются альтернативы,
- где возникает цена ошибки,
- где «мое» и «важное» сходятся в моменте необратимости.

В FOM сознательная архитектура возникает не из одного лишь мета-представления, а из связки:

- S (для кого),
- M (что важно/недопустимо),
- A (где закрываются альтернативы),
- W (модель реальности),
- организованных в едином режиме Q.

Итог

- HOT корректны как теории само-отнесения и мета-репрезентации.
- Они глубоко прорабатывают компонент S.
- Однако они не специфицируют связь само-индексации с необратимостью (A), нормативной структурой значимости (M) и форматом данности (Q).
- Поэтому в терминах FOM HOT описывают необходимый компонент сознательной архитектуры, но не исчерпывают ее целиком.

7. Итоговое сопоставление

Проведенный анализ показывает, что различные теории сознания фиксируют разные архитектурные аспекты когнитивной системы:

- GWT — механизм глобального доступа и координации;
- IIT — структурную интеграцию причинных связей;
- PP/FEP — динамику регуляции и стабилизации;
- HOT — само-индексацию и мета-репрезентацию;
- панпсихизм — интуицию неустранимости формата;
- иллюзионизм — конструктивность содержания модели.

Каждая из этих теорий описывает реальный компонент архитектуры, однако ни одна из них в явном виде не охватывает целиком связку:

- **W** (модель реальности) —
- **A** (необратимость действия) —
- **S** (само-индексация) —
- **M** (структура значимости) —
- в режиме **Q** (квалиативной организации),

то есть ставочную организацию сознательного состояния как единой конфигурации.

Уровни, на которых возникают систематические смешения

Метатеория функционального наблюдателя позволяет развести несколько типов категориальных переносов:

- **формат и содержание** (Q ≠ W);
- **регуляцию и данность** (динамика B ≠ режим Q);
- **архитектуру агента и онтологию реальности** (свойства B ≠ свойства R);
- **симуляцию и необратимую фиксацию** (возможности в W ≠ закрытие альтернатив через A).

Именно при смешении этих уровней возникают псевдопроблемы — в том числе в дискуссии о «трудной проблеме».

Статус FOM

В этом смысле FOM не предлагается как еще одна частная теория сознания, конкурирующая с существующими моделями.

Ее задача иная: выступить метатеоретической рамкой, в которой становится ясно:

- какую архитектурную функцию описывает каждая теория;
- почему эти теории оказываются частично корректными;
- и почему они остаются неполными без явной фиксации всей ставочной связки WASM в режиме Q.

Тем самым FOM предлагает не замену существующих подходов, а структурное разведение уровней анализа, позволяющее устранить систематические категориальные ошибки и уточнить область применимости частных теорий.

Библиография

1. Общий философский контекст сознания

Block, N. (1995). On a confusion about a function of consciousness. Behavioral and Brain Sciences, 18(2), 227–247.

Chalmers, D. J. (1996). The conscious mind: In search of a fundamental theory. Oxford University Press.

Chalmers, D. J. (2010). The character of consciousness. Oxford University Press.

Nagel, T. (1974). What is it like to be a bat? The Philosophical Review, 83(4), 435–450.

Metzinger, T. (2003). Being no one: The self-model theory of subjectivity. MIT Press.

2. Global Workspace Theory (GWT)

Baars, B. J. (1988). A cognitive theory of consciousness. Cambridge University Press.

Baars, B. J. (2005). Global workspace theory of consciousness: Toward a cognitive neuroscience of human experience. Progress in Brain Research, 150, 45–53.

Dehaene, S. (2014). Consciousness and the brain: Deciphering how the brain codes our thoughts. Viking.

Dehaene, S., & Changeux, J.-P. (2011). Experimental and theoretical approaches to conscious processing. Neuron, 70(2), 200–227.

3. Integrated Information Theory (IIT)

Tononi, G. (2004). An information integration theory of consciousness. BMC Neuroscience, 5, 42.

Tononi, G. (2008). Consciousness as integrated information: A provisional manifesto. The Biological Bulletin, 215(3), 216–242.

Oizumi, M., Albantakis, L., & Tononi, G. (2014). From the phenomenology to the mechanisms of consciousness: Integrated information theory 3.0. PLoS Computational Biology, 10(5), e1003588.

Tononi, G., Boly, M., Massimini, M., & Koch, C. (2016). Integrated information theory: From consciousness to its physical substrate. Nature Reviews Neuroscience, 17, 450–461.

4. Predictive Processing / Free Energy Principle (PP/FEP)

Friston, K. (2010). The free-energy principle: A unified brain theory? Nature Reviews Neuroscience, 11, 127–138.

Friston, K. (2013). Life as we know it. Journal of the Royal Society Interface, 10, 20130475.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behavioral and Brain Sciences, 36(3), 181–204.

Hohwy, J. (2013). The predictive mind. Oxford University Press.

5. Higher-Order Theories (HOT)

Rosenthal, D. M. (2005). Consciousness and mind. Oxford University Press.

Rosenthal, D. M. (2002). How many kinds of consciousness? Consciousness and Cognition, 11(4), 653–665.

Lau, H., & Rosenthal, D. (2011). Empirical support for higher-order theories of conscious awareness. Trends in Cognitive Sciences, 15(8), 365–373.

Carruthers, P. (2011). The opacity of mind. Oxford University Press.

6. Иллюзионизм

Dennett, D. C. (1991). Consciousness explained. Little, Brown.

Frankish, K. (2016). Illusionism as a theory of consciousness. Journal of Consciousness Studies, 23(11–12), 11–39.

Frankish, K. (Ed.). (2017). Illusionism as a theory of consciousness. Imprint Academic.

7. Панпсихизм

Chalmers, D. J. (2015). Panpsychism and panprotopsychism. In T. Alter & Y. Nagasawa (Eds.), Consciousness in the physical world (pp. 246–276). Oxford University Press.

Goff, P. (2017). Consciousness and fundamental reality. Oxford University Press.

Seager, W. (2010). Panpsychism, aggregation and combinatorial infusion. Mind & Matter, 8(2), 167–184.

Shani, I. (2015). Cosmopsychism: A holistic approach to the metaphysics of experience. Philosophical Papers, 44(3), 389–437.

8. Дополнительные современные обзоры

Seth, A. K. (2021). Being you: A new science of consciousness. Dutton.

Doerig, A., Schurger, A., Hess, K., & Herzog, M. H. (2021). The unfolding argument: Why IIT and other causal structure theories cannot explain consciousness. Consciousness and Cognition, 72, 102758.

Michel, M., & Lau, H. (2020). The measurement of consciousness: A framework for studying conscious processing. Philosophy and the Mind Sciences, 1, 1–27.