

# Метатеория функционального наблюдателя

## Часть 3 – Возражения

Версия 1.0

АРТЕМОВ ДМИТРИЙ

## Оглавление

<b>Аннотация</b> .....	2
<b>1. Функциональная избыточность сознания</b> .....	3
<b>2. Почему именно необратимость?</b> .....	5
<b>3. Проблема границы субъекта</b> .....	7
<b>4. Проблема минимального субъекта</b> .....	9
<b>5. Отсутствие количественных критериев</b> .....	11
<b>6. Проблема FOM-зомби</b> .....	12
<b>7. Проблема объяснения квалиа</b> .....	14
<b>Библиография</b> .....	16
Основная работа .....	16
Философия сознания .....	16
Теории сознания и когнитивная архитектура .....	16
Философия субъекта и идентичности .....	17

## Аннотация

Данная работа представляет третью часть исследования метатеории функционального наблюдателя (Functional Observer Metatheory, FOM) и посвящена анализу ключевых возражений, возникающих в отношении предложенной архитектурной модели сознания. Рассматриваются основные философские и методологические проблемы, включая функциональную избыточность феноменального опыта, роль необратимости действий в формировании субъекта, проблему границы и минимального субъекта, отсутствие универсальных количественных критериев, аргумент философских зомби, а также классическую «трудную проблему сознания».

В работе показано, что многие из этих возражений возникают из смешения различных уровней анализа: уровня архитектуры системы, уровня модели мира, а также уровня формата представления. В рамках FOM феноменальный режим Q интерпретируется не как дополнительная сущность или особый объект среди элементов модели, а как режим, в котором внутренняя модель мира становится доступной системе для регуляции действий и удержания собственной устойчивости.

Предлагается рассматривать субъект не как фиксированный анатомический или вычислительный объект, а как динамическую регуляторную конфигурацию, определяемую связкой компонентов модели мира ( $W$ ), оценки значимости ( $M$ ), якоря необратимых действий ( $A$ ) и окна устойчивости системы ( $L$ ). Показано, что многие традиционные философские проблемы сознания, включая аргумент зомби и «трудную проблему», могут быть интерпретированы как следствия категориального смешения между уровнем содержания модели и форматом её данности системе.

Работа носит метатеоретический характер и направлена на уточнение архитектурных условий, при которых внутренняя модель системы может функционировать в квалиативном режиме.

# 1. Функциональная избыточность сознания

Возражение

Все функции FOM могут быть реализованы **без субъективного опыта**:

- планирование
- прогнозирование
- управление
- выбор действий

Зачем тогда вообще нужен **феноменальный опыт**?

Ответ

FOM не утверждает, что перечисленные функции принципиально невозможны без субъективного опыта. Напротив, многие искусственные системы способны выполнять задачи планирования, прогнозирования и управления без феноменального режима.

Однако данное возражение игнорирует **архитектурную стоимость реализации этих функций**. В рамках FOM режим Q рассматривается не как дополнительная вычислительная способность, а как **регуляторный режим**, возникающий в системах, которые одновременно:

- должны принимать решения в условиях ограниченных ресурсов,
- сталкиваются с **необратимыми последствиями действий**,
- и не могут позволить себе полное аналитическое моделирование всех возможных состояний среды.

В таких системах возникает необходимость в механизме, который позволяет **интегрировать аналитическую обработку информации с непосредственным удержанием системы в пределах её окна устойчивости L**.

Иными словами, в биологических агентах нельзя полностью разделить:

- вычислительную работу по анализу среды,
- и задачу **самосохранения системы**.

Ошибки в вычислении или планировании непосредственно приводят к риску необратимого выхода системы из L. Поэтому архитектура агента должна обеспечивать быструю и непрерывную оценку значимости состояний и действий относительно собственной уязвимости.

FOM интерпретирует феноменальный режим Q как **эффективный способ организации такой регуляции** в условиях, где альтернативные решения (например, полное переборное моделирование среды или массивное параллельное масштабирование вычислений) оказываются либо слишком дорогими, либо недоступными.

Это объясняет, почему биологические системы эволюционно используют режим Q, тогда как многие искусственные системы могут обходиться без него. Например, централизованные вычислительные системы или базы данных не находятся в состоянии постоянного риска необратимого разрушения и могут разделять аналитическую обработку данных и физическое существование системы.

Таким образом, феноменальный режим в FOM не является избыточным. Он представляет собой **архитектурное решение**, позволяющее системам с ограниченными ресурсами и высокой ставкой ошибок эффективно управлять собственным существованием в динамической среде.

## 2. Почему именно необратимость?

Возражение

FOM утверждает, что ключевым элементом является **якорь необратимых действий А**.

Но многие системы принимают **необратимые решения**:

- торговые алгоритмы
- автоматические регуляторы
- системы управления ракетами

Почему эти системы не обладают сознанием?

Ответ

FOM не утверждает, что **любая система, совершающая необратимые действия**, автоматически обладает субъективностью. Ключевым условием является не просто необратимость действия, а **необратимость последствий для самой системы**.

В архитектуре FOM якорь А обозначает действия, последствия которых могут вывести систему за пределы её **окна устойчивости L**, то есть разрушить её способность продолжать функционирование.

В таких условиях система должна непрерывно оценивать значение происходящего относительно собственной устойчивости. Именно эта регуляторная динамика связывает:

- модель мира **W**,
- оценку значимости **M**,
- и выбор действий относительно **A**.

Именно эта связка образует режим Q.

Во многих искусственных системах необратимость существует лишь **на уровне внешней задачи**, а не на уровне существования самой системы.

Например:

- торговый алгоритм может потерять деньги, но сам алгоритм не уничтожается;
- регулятор может принять неудачное решение, но система управления остаётся функционирующей;
- система управления ракетой выполняет миссию, но её собственное существование не является объектом регуляции.

Такие системы могут быть перезапущены, перепрограммированы или заменены без угрозы их собственной архитектурной целостности.

Биологические агенты находятся в принципиально иной ситуации. Их решения непосредственно связаны с риском **необратимого выхода из L** — травмой, разрушением организма или смертью. Ошибка не может быть «переиграна» или отменена.

Именно поэтому в биологических системах возникает необходимость в особом регуляторном режиме, который интегрирует:

- восприятие,
- оценку значимости,
- и выбор действий

в единую динамику, ориентированную на сохранение системы.

FOM интерпретирует этот режим как **феноменальный режим Q**.

### 3. Проблема границы субъекта

Возражение

Где проходит граница субъекта в FOM?

Например:

- мозг
- организм
- мозг + тело
- мозг + инструменты

Почему именно на одном уровне возникает субъект?

Ответ

В FOM проблема границы субъекта решается через различие **режима первого лица** и **режима третьего лица**.

**В режиме первого лица** граница субъекта не задаётся анатомически и не выводится из простой пространственной локализации. Она задаётся через компонент **S** (самоиндексация), а на уровне внутренней модели выражается как различие между **Wb** и **Wr**. Здесь **Wb** обозначает ту подмодель состояний, которые системой интерпретируются как относящиеся к ней самой: как собственные ресурсы, уязвимости, повреждения, возможности действия и условия сохранения. Соответственно, в режиме первого лица граница субъекта — это не граница «вещи в пространстве», а граница **самоиндексируемой области модели**, относительно которой оцениваются риски, потери и допустимые действия.

Это означает, что субъект в FOM не совпадает автоматически ни с мозгом, ни с телом как анатомическим объектом. Например, инструмент, протез, транспорт или внешний интерфейс могут временно входить в область **Wb**, если их состояние становится функционально эквивалентным состоянию самого агента в контуре удержания системы в окне устойчивости **L**. И наоборот, некоторые части тела могут в конкретном режиме регуляции не входить в актуальную границу **Wb**, если они не представлены системой как значимый элемент текущего самосохранения и действия.

**В режиме третьего лица** ситуация иная. Наблюдатель не имеет прямого доступа к **Q** и не наблюдает «субъектность» как отдельный объект среди прочих объектов мира. Согласно FOM, наблюдателю даны только элементы **W**: физические процессы, телесные движения, сенсорные и моторные корреляции, паттерны регуляции, изменения поведения, а также следы организации системы. Поэтому субъект не может быть просто «увиден» как особая сущность внутри **W**. Он может быть лишь **реконструирован** по тем динамическим признакам, которые указывают на наличие агента, удерживающего некоторую конфигурацию состояний в пределах **L**.

Отсюда следует важный методологический вывод: граница субъекта в режиме третьего лица должна определяться **не статически, а динамически**. Иными словами, вопрос состоит не в том, какие элементы расположены рядом друг с другом или принадлежат одному телу, а в том, какие элементы реально включены в единый контур самосохранения, планирования и необратимого действия. Если некоторое внешнее устройство или инструмент систематически участвует в удержании агента в  $L$ , включается в само-индексацию  $S$  и получает высокую значимость в структуре  $M$ , то оно может рассматриваться как часть субъектной конфигурации. Если же элемент лишь каузально взаимодействует с агентом, но не входит в его ставочную архитектуру, он остаётся частью  $W_r$ , то есть внешней реальности для данного субъекта.

Таким образом, FOM отвергает фиксированную и универсальную границу субъекта. Вместо этого теория предлагает **операционально-динамический критерий**: граница субъекта проходит там, где заканчивается область состояний, индексируемых системой как собственные и функционально вовлечённых в удержание её устойчивости, планирование и необратимые действия.

В этом смысле субъект в FOM не является просто анатомическим объектом и не является чисто вычислительным центром. Он представляет собой **динамически собираемую регуляторную конфигурацию**, определяемую связкой  **$W_b-S-M-A$**  относительно окна устойчивости  $L$ .

## 4. Проблема минимального субъекта

Возражение

Какая **самая простая система**, реализующая FOM?

Например:

- муха
- осьминог
- бактерия
- клетка

Где проходит **порог сознания**?

Ответ

FOM не связывает существование субъекта с размером системы, уровнем интеллекта или биологической «сложностью». Теория также не вводит фиксированный универсальный порог, при котором сознание появляется или исчезает.

Вместо этого FOM формулирует **архитектурные условия**, необходимые для возникновения режима Q. Эти условия включают:

- наличие активной модели мира **W**,
- якоря необратимых действий **A**,
- механизма оценки значимости **M**,
- и окна устойчивости **L**, в пределах которого система должна удерживать своё существование.

Минимальный субъект в таком случае определяется не по морфологическим или биологическим критериям, а по **минимальной конфигурации системы**, способной реализовать эту регуляторную архитектуру.

При этом разные исследовательские задачи могут выделять различные предельные случаи минимальности. Например, можно рассматривать системы с:

- минимальным сенсорным входом **in**,
- минимальной способностью воздействовать на среду (**out**),
- минимальным окном устойчивости **L**,
- или минимальной набором состояний внутренней модели **W**.

Определение того, какие реальные биологические или искусственные системы удовлетворяют этим условиям, относится к области **эмпирических наук** — нейробиологии, этологии, когнитивной науки и робототехники.

Таким образом, FOM не фиксирует конкретный биологический уровень, на котором возникает субъектность. Вместо этого теория предлагает **операциональный критерий**: субъект возникает там, где система реализует архитектуру WASM и вынуждена

удерживать себя в пределах собственного окна устойчивости **L** посредством необратимых действий.

В этом смысле вопрос о минимальном субъекте не является чисто философским. Он становится **исследовательской задачей**, связанной с выявлением минимальной системы, способной реализовать соответствующую регуляторную динамику.

## 5. Отсутствие количественных критериев

Возражение

Теория вводит параметры **L, C, T**, но не определяет как их количественно измерять.

Без таких критериев теория может показаться слишком абстрактной и плохо проверяемой.

Ответ

FOM сознательно не вводит универсальные количественные метрики для параметров **L, C, T**. Это связано с тем, что теория формулируется на уровне **архитектурных принципов**, а не на уровне конкретных моделей отдельных наук.

Количественная интерпретация этих параметров неизбежно зависит от **выбранной модели системы и исследовательской задачи**. В разных дисциплинах одна и та же система описывается различными моделями и уровнями абстракции.

Например, объект «человек» может рассматриваться:

- в психологии — как система когнитивных процессов и поведенческих стратегий;
- в биологии — как совокупность физиологических и нейронных механизмов;
- в организационном управлении — как агент, принимающий решения в институциональной среде.

Каждая из этих моделей использует собственные переменные, масштабы и методы измерения. Соответственно, параметры **L, C, T** получают количественную интерпретацию только **внутри конкретной прикладной модели**.

В этом смысле FOM выполняет роль **метатеоретической схемы**, которая задаёт структурные требования к системе, но не фиксирует единственный способ их реализации или измерения.

Поэтому вопрос о количественных критериях не имеет универсального ответа на уровне метатеории. Он должен решаться на уровне конкретных дисциплин — нейробиологии, когнитивной науки, этологии или исследований искусственных агентов.

## 6. Проблема FOM-зомби

Возражение

Можно ли представить систему, которая реализует архитектуру FOM:

- активную модель мира **W**
- якорь действий **A**
- оценку значимости **M**
- жизненное окно **L**

но при этом **не обладает субъективным опытом?**

Если такая система возможна, то FOM описывает лишь функциональную архитектуру агента, но не объясняет сознание.

Ответ

Это возражение основано на аргументе **философских зомби** (Chalmers), согласно которому возможно представить систему, функционально идентичную человеку, но лишённую субъективного опыта. Однако данный аргумент опирается на предположение, что мы можем сравнивать системы **вне набора наблюдаемых свойств**.

В действительности любое сравнение систем происходит относительно **некоторого набора характеристик**, которые можно назвать *контейнером сравнения*.

Если два объекта идентичны по набору свойств **P**, то любое различие вне **P не может быть установлено сравнением объектов относительно P**.

Например, два механических компонента могут быть идентичны по:

- геометрии
- допускам
- функции

но различаться по

- цвету
- маркировке
- материалу покрытия

Если цвет, маркировка и материал не входит в контейнер сравнения, то для рассматриваемой задачи эти детали являются **эквивалентными**.

Аналогичным образом утверждение о «зомби» предполагает существование различия, которое **по определению исключено из набора сравниваемых свойств**.

FOM принимает принцип **операциональной идентичности**:

*если две системы идентичны по всей наблюдаемой архитектуре WASM и демонстрируют одинаковую регуляторную динамику, то утверждение о наличии дополнительного различия (опыта) не имеет операционального содержания.*

Иначе говоря, различие между системой с опытом и системой без опыта становится **неопределяемым в рамках сравнения систем**.

Поэтому аргумент FOM-зомби либо:

1. утверждает различие, которое невозможно операционально выявить,
2. либо предполагает дополнительные свойства, не входящие в архитектуру системы.

В первом случае утверждение оказывается **метафизически пустым**, во втором — требует введения **дополнительной теории**, выходящей за пределы рассматриваемой модели.

Таким образом FOM не опровергается аргументом зомби, поскольку данный аргумент предполагает различие, которое не может быть определено в рамках операционального сравнения систем.

## 7. Проблема объяснения квалиа

Возражение

Даже если архитектура FOM описывает условия функционирования субъекта, остаётся классическая **«трудная проблема сознания»**, сформулированная David Chalmers:

почему физические процессы в мозге сопровождаются феноменальным переживанием? Иными словами, почему система не просто функционирует, а **«что-то чувствует»**?

Ответ

В рамках FOM исходная формулировка трудной проблемы требует уточнения. Ключевое различие вводится между **содержанием модели** и **форматом её данности системе**.

В архитектуре FOM:

- **W** — внутренняя модель мира, включающая представления о телесных состояниях, внешних объектах, причинных связях и динамике процессов;
- **Q** — режим, в котором эта модель может быть **дана системе как значимая для действия**.

Важно, что **Q не является дополнительной сущностью или отдельным объектом**, существующим наряду с элементами модели. Q представляет собой **формат представления**, то есть способ, в котором состояния модели становятся доступными системе (агенту).

Формат представления принципиально отличается от содержимого представления. Элементы вроде нейронов, сигналов, физических процессов и причинных взаимодействий относятся к **содержанию модели W**. Они описывают структуру и динамику представляемых состояний.

Q, напротив, относится к **уровню формата**, то есть к режиму, в котором эти состояния вообще могут быть даны системе.

Поэтому попытка обнаружить Q «в мозге» как один из элементов физической структуры смешивает два уровня анализа: **уровень формата и уровень содержания**. Она ищет способ представления среди представляемых элементов.

Именно такое смешение лежит в основе классической формулировки трудной проблемы. В ней предполагается разрыв между физическим описанием процессов и субъективным переживанием. Однако это противопоставление формируется уже **внутри модели W**, поскольку и «физические процессы», и «объекты», и «причинность» принадлежат уровню описания модели.

Мы не имеем прямого доступа к реальности **R** вне формата представления; нам даны лишь структурированные состояния модели W. Это не отрицает существование R, но

фиксирует эпистемическую ситуацию: любые описания, включая физические, относятся к уровню модели.

В результате трудная проблема принимает форму требования **вывести формат представления из содержимого представления**. Однако формат не является элементом множества содержаний; он задаёт способ их организации и доступности системе.

Поэтому требование объяснить феноменальность через физические процессы предполагает вывод формата из его собственных элементов, что делает исходную постановку логически некорректной.

После разведения уровней вопрос изменяется. Он перестаёт звучать как:

**«Как физические процессы порождают субъективный опыт?»**

и формулируется иначе:

**«При какой архитектуре система может иметь внутреннюю модель, данную ей в режиме Q?»**

ГОМ отвечает именно на этот вопрос. Она анализирует **архитектурные условия**, при которых внутренняя модель может организовываться в качественный режим.

Метатеория функционального наблюдателя не предлагает метафизического объяснения самой «данности» как таковой. Вместо этого она реконструирует **структуру системы**, при которой внутренняя модель может быть дана агенту и использоваться для регуляции его действий.

# Библиография

## Основная работа

Артемов Д. А. (2026).

**Метатеория функционального наблюдателя. Часть I: Архитектура и формальная структура квалиативного режима.**

PREPRINTS.RU.

<https://doi.org/10.24108/preprints-3114622>

## Философия сознания

Chalmers, D. J. (1996).

**The Conscious Mind: In Search of a Fundamental Theory.**

Oxford University Press.

Chalmers, D. J. (1995).

**Facing Up to the Problem of Consciousness.**

Journal of Consciousness Studies, 2(3), 200–219.

Block, N. (1995).

**On a Confusion about a Function of Consciousness.**

Behavioral and Brain Sciences, 18(2), 227–247.

Nagel, T. (1974).

**What Is It Like to Be a Bat?**

Philosophical Review, 83(4), 435–450.

Jackson, F. (1982).

**Epiphenomenal Qualia.**

Philosophical Quarterly, 32(127), 127–136.

Dennett, D. (1991).

**Consciousness Explained.**

Little, Brown and Company.

## Теории сознания и когнитивная архитектура

Vaars, B. (1988).

**A Cognitive Theory of Consciousness.**

Cambridge University Press.

Dehaene, S. (2014).

**Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts.**  
Viking.

Tononi, G. (2004).

**An Information Integration Theory of Consciousness.**  
BMC Neuroscience.

Tononi, G., Koch, C. (2015).

**Consciousness: Here, There and Everywhere?**  
Philosophical Transactions of the Royal Society B.

Философия субъекта и идентичности

Parfit, D. (1984).

**Reasons and Persons.**  
Oxford University Press.

Metzinger, T. (2003).

**Being No One: The Self-Model Theory of Subjectivity.**  
MIT Press.