

Содержание

Defense, Self, and Speech: From Biological Vectors to Linguistic Form	2
Статья II — Вычислима ли компенсаторно-защитная речь? Комбинаторная архитектура конечного операционального базиса	2
0. Место статьи в серии	2
1. Аннотация	3
2. Введение	4
2.1 Вопрос	4
2.2 Рамка предиктивной обработки	5
2.3 Три кандидат-рамки	6
2.4 Речь как привилегированное наблюдаемое активного вывода	8
2.5 Вклад и дорожная карта	9
3. Четыре аргумента в пользу конечного операционального базиса	9
3.1 Метаболический аргумент	10
3.2 Клинический аргумент: стереотипность как свидетельство	11
3.3 Предиктивно-вычислительный аргумент	14
3.4 Развивающий аргумент	16
3.5 Схождение и что оно подразумевает	17
4. Алфавит: операции, не шаблоны	17
4.1 Операции, шаблоны, режимы	18
4.2 Операции семейства fight	19
4.3 Операции семейства flight	20
4.4 Операции семейства fawn	20
4.5 Операции семейства freeze	21
4.6 Алфавит как целое	22
5. Четыре precision regime как эмерджентные аттракторы	23
5.1 Почему четыре?	23
5.2 DEFENCE	24
5.3 COGNITIVE	25
5.4 ILLUSION	26
5.5 META	27
5.6 Операции как речевые сигнатуры режимов	28
6. Композиция и контекст	29
6.1 Композиция внутри высказывания	29
6.2 Последовательная композиция через высказывания	30
6.3 Контекст как модулятор	31
6.4 Якорь-промпт	32
6.5 Просодия и то, что упускает текстовая формализация	34
7. Различитель экологично/компенсаторно на уровне операций	35
7.1 Пропорциональность	35
7.2 Обратимость	36
7.3 Сохранение контакта	38
7.4 Ковариация и композитный различитель	39
7.5 Что различитель доставляет и чего он не делает	41
8. Объём, три уровня описания и мост к статье III	41
8.1 Чего эта статья не утверждает	41
8.2 Три уровня описания	43

8.3 Замечание о слове «SEEK»	45
8.4 Структурное замечание о МЕТА	46
8.5 Мост к статье III	46
9. Заключение	48
Приложение А — Формальная характеристика	49
A.1 Вариационная свободная энергия и precision-взвешенный message-passing	49
A.2 Precision regimes как наборы гиперпараметров	50
A.3 Той-модель: минимальный двухслойный линейно-гауссовский переключатель режимов	51
A.4 Расширение на связанный диадический вывод	52
A.5 Пять фальсифицируемых предсказаний	53
A.6 Пайплайны анализа и сводка	55
Благодарности	56
Литература	57

Defense, Self, and Speech: From Biological Vectors to Linguistic Form¹

Статья II — Вычислима ли компенсаторно-защитная речь? Комбинаторная архитектура конечного операционального базиса

Автор: Maria Svet, Mindloom Cognitive Architecture Lab **Благодарность:** Структурная разработка велась в расширенном диалоге с Claude (Anthropic); см. финальный раздел «Благодарности». **Серия:** *Defense, Self, and Speech: From Biological Vectors to Linguistic Form*, статья II из VI **Рабочее имя у соавторов:** *Architecture of Protective Speech* — Маша и Claude **Версия:** v0.8 (preprint) **Целевые площадки:** PsyArXiv (английский) · PsyArXiv / русский препринт-сервер (русский) · расширенная версия в *Frontiers in Psychology* или *Cognitive Science*

0. Место статьи в серии

Статья I установила, что четыре канонических защитных ответа — fight, flight, fawn, freeze — сохраняются как направленные векторы с обретением языка, и что специфический подкласс защитной речи, который мы называем *компенсаторно-защитной*, структурно организован опережением предсказанного стыда относительно жёсткого социально-нормативного prior'a. Статья II поднимает методологический вопрос-следствие: допускает ли эта архитектура формализацию, делающую компенсаторно-защитную речь, хотя бы частично, вычислимой? Мы утверждаем, что допускает, — через интегрированное предиктивно-процессинговое описание: компенсаторно-защитная речь производится из конечного алфавита речевых операций, организованных четырьмя биологическими векторами, чьи темпоральные паттерны реализуют четыре повторяющихся precision regime внутри иерархической генеративной модели. Результирующее

¹Mindloom Cognitive Architecture Lab, 2026 info@mindloom.ch

пространство гипотез одновременно конечно и теоретически обосновано. Это методологический фундамент, на котором стоит остальная серия: статья III описывает движок, операционализирующий это пространство гипотез как онтологически-ограниченную классификационную архитектуру; статья IV отслеживает траектории внутри пространства состояний-режимов; статьи V и VI эмпирически валидируют и применяют подход на человеческой и машинной речи соответственно.

1. Аннотация

Статья I этой серии обосновала, что четыре канонических защитных ответа — fight, flight, fawn, freeze — сохраняются как направленные векторы с обретением организмом языка, и что специфический подкласс защитной речи (который мы называем *компенсаторно-защитной*) организован опережением предсказанного стыда относительно жёсткого социально-нормативного prior'a. В рамке предиктивной обработки такая защита есть форма иерархического байесовского вывода, оперирующего на защищённых priors. Настоящая статья задаёт методологический вопрос-следствие: допускает ли эта архитектура формализацию, делающую компенсаторно-защитную речь *вычислимой*?

Мы утверждаем, что допускает, и аргумент разворачивается по двум связанным осям.

Первая ось: конечный операциональный базис. Компенсаторно-защитная речь производится из ограниченного алфавита **речевых операций** — не шаблонов, не фиксированных фраз — взятых из обозримого инвентаря, организованного четырьмя биологическими векторами, и составляемых по контекстно-чувствительным, но обозримым правилам. Последовательности, получаемые на выходе, в принципе неограниченны; операции, их производящие, — нет. Это то отношение, что удерживается между четырьмя нуклеотидными основаниями и геномной библиотекой или между конечными правилами шахмат и астрономическим пространством возможных партий. Мы представляем четыре сходящиеся линии свидетельств: *метаболический* аргумент (устойчивая компенсаторная защита на поочерёдной сборке была бы аллостатически непозволительной — при пересечении трёх условий: повышенной базовой цены, отсутствия разрядки и непрерывной работы), *клинический* аргумент (столетие независимых наблюдательных традиций — психоаналитической, объектно-реляционной, attachment-, социально-когнитивной, травматическо-клинической — сошлось на небольшом числе повторяющихся форм, схождение, задокументированное здесь через четыре кросс-традиционные триангуляции), *предиктивно-вычислительный* аргумент (priors переиспользуются, потому что предсказание есть экономия, и защитные priors переиспользуются с особой жёсткостью, поскольку защищены от обновления), и *развивающий* аргумент (защиты кристаллизуются в повторяющиеся формы через развивающее осаждение, а не изобретаются заново в каждой ситуации).

Вторая ось: precision regime как эмерджентные аттракторы. Те же операции, рассмотренные на уровне темпорального паттерна и распределения precision, группируются в четыре повторяющиеся архитектуры активного вывода:

DEFENCE (повышенная prior precision на самозащитных priors, подавленная likelihood precision на disconfirmatory channels, сниженный эпистемический вес), **COGNITIVE** (сбалансированное, контекстно-чувствительное распределение precision, поддерживающее обновление по свидетельствам и кооперативную репарацию), **ILLUSION** (смещённые prior means с избирательным снижением likelihood precision, самоподдерживающиеся в отсутствие острой угрозы) и **МЕТА** (вывод более высокого порядка над самим precision, позволяющий рефлексивную модуляцию нижних уровней). Эти четыре режима — не категории, накладываемые на данные; это аттракторо-подобные конфигурации распределения precision внутри единой иерархической генеративной модели. Операции первой оси — наблюдаемые речевые выходы; режимы второй — динамические паттерны, которые они прочерчивают. Мы представляем концептуальную структуру четырёх режимов в основном тексте и формальную характеристику — precision-гиперпараметры, линейно-гауссовскую toy-модель и пять фальсифицируемых предсказаний — в Приложении А.

Вместе эти две оси устанавливают, что пространство гипотез для компенсаторно-защитной речи *одновременно конечно и теоретически обосновано*. Это методологический результат, на котором стоит остальная серия. Статья III описывает движок, операционализирующий это пространство гипотез как нейро-символическую классификационную архитектуру со *structure-constrained inference* — слоями на основе правил, определяющими ограниченное пространство гипотез, внутри которого осуществляется LLM-суждение. Статьи V и VI эмпирически валидируют и применяют подход. Мы здесь предлагаем не движок, не эмпирику и не прикладную методологию — мы предлагаем теоретическое санкционирование классификационной работы, следующей за ней: демонстрацию того, что пространство, внутри которого эта классификация оперирует, конечно, структурно, и выводимо из независимых теоретических соображений, а не навязано инженерным удобством.

Мы закрываем статью явным указанием того, чего она *не* утверждает. Мы не утверждаем, что речь вообще комбинаторна, или что вся precision-динамика дискурса укладывается в четыре режима. Утверждение уже и принципиальнее: специфический подкласс речи — компенсаторно-защитный, производимый в предиктивно-стыдовой динамике, — допускает формализацию на конечном базисе с определённой режимной структурой, и эта формализация — то, что санкционирует инженерную работу ниже по течению.

Ключевые слова: защитная речь, вычислимость, комбинаторность, конечный базис, предиктивная обработка, активный вывод, precision regimes, аллостатическая нагрузка, клиническая стереотипность, речевые операции, аннотационная онтология, онтологически-ограниченная классификация.

2. Введение

2.1 Вопрос

Вопрос, за который берётся эта статья, прост в формулировке и неожиданно требователен к ответу. Если компенсаторно-защитная речь, как утверждалось

в статье I, есть связный феномен, укоренённый в сохранённых биологических векторах и организованный опережением предсказанного стыда, — можем ли мы её *формализовать*? Можем ли мы задать её операции с достаточной точностью, чтобы некоторая система — аннотатор-человек, вычислительный классификатор или обученный клиницист — могла узнавать, маркировать и рассуждать об её инстанциях принципиальным образом?

Вопрос имеет значение, потому что ответ определяет, что становится возможным далее. Если компенсаторно-защитная речь формально неухватываема — если она фундаментально креативна, неограничена, нередуцируема, — то всё, что предложила статья I, остаётся теоретической литературой: интересной, но безрезультатной. Если она допускает частичную формализацию, — тогда возможна онтология; если возможна онтология — возможен движок; если возможен движок — становятся достижимы эмпирические исследования, клинические инструменты и оценка безопасности языковых моделей. Вопрос о вычислимости — не орнаментальная сноска. Это шарнир, на котором поворачивается вся остальная серия.

2.2 Рамка предиктивной обработки

Вопрос о вычислимости не может быть поставлен в абстрактном виде. Он должен быть поставлен внутри рамки, специфицирующей, что делает говорящий мозг, когда производит речь, и чем, на базовом уровне, речь является. Для этой цели мы принимаем рамку предиктивной обработки (Friston, 2010; Clark, 2013; Hohwy, 2013), в пределах которой мозг моделируется как иерархическая генеративная модель, непрерывно предсказывающая свои сенсорные входы, сравнивающая предсказания с поступающими сигналами и обновляющаяся через минимизацию ошибки предсказания. Действие в этой рамке не является отдельной категорией от восприятия; оба есть формы вывода, причём действие — это исполнение политик, минимизирующих ожидаемую свободную энергию (Friston et al., 2017; Parr et al., 2022). Рамка не орнаментальна; это рабочая теория, внутри которой делается аргумент настоящей статьи. Там, где три рамки, которые мы рассмотрим в §2.3, — это кандидат-теории того, как язык работает как вычислимый объект, предиктивная обработка — это рамка, внутри которой мы специфицируем, чем является говорящий агент и чем — его выходы.

Эта рамка имеет специфические следствия для вопроса вычислимости, существенные для дальнейшего изложения. Внутри предиктивной обработки устойчивые поведения тяготеют к кэшированной реализации. Это прямое следствие принципа свободной энергии: организм, реконструирующий свой ответ в каждом случае, нёс бы затраты на ошибку предсказания и метаболические затраты, которые архитектура явно спроектирована минимизировать. Кэшированные priors — устойчивые вероятностные ожидания относительно состояний, переходов и ответов — суть механизм мозга для эффективного вывода. Priors, которые работают, подкрепляются; priors, которые не работают, обновляются; на масштабах развития и адаптации генеративная модель сходится к конечному репертуару кэшированных priors, покрывающему диапазон ситуаций, с которыми организм реально встречается.

Компенсаторно-защитная речь, как мы утверждали в статье I, есть форма активного вывода при специфическом архитектурном условии. Это опережающий ответ организма на *предсказанную* ошибку между жёстким социально-нормативным *prior*-ом (интериоризованным Я-каким-следует-быть) и текущей ситуацией. Защита — это политика, выбранная для минимизации этой предсказанной ошибки. Поскольку социально-нормативный *prior* защищён от обновления — своим развивающим истоком в авторитете заботящегося взрослого, текущим социальным подкреплением и высоким положением в иерархии, — один и тот же класс политик многократно выбирается, подкрепляется и кэшируется. Под этим описанием конечный операциональный базис, который мы будем защищать, не есть внешнее наложение на данные; это *то, что предиктивная обработка специфически предсказывает для агента, организованного вокруг защищённого prior'a с кэшированными опережающими ответами.*

Это критический ход для настоящей статьи. Аргумент в пользу конечного базиса — не только (и не в первую очередь) эмпирическое обобщение из клинических наблюдений. Это теоретическое следствие рамки, внутри которой работает серия в целом. Если защита есть опережение предсказанного стыда, если это опережение имплементировано как выбор политики, и если политики с успешной минимизацией свободной энергии кэшируются, — то компенсаторно-защитная речь покоится на конечном инвентаре кэшированных политик как на механизме, а не как на случайности. Четыре аргумента, развиваемые в §3, могут быть прочитаны во втором проходе как четыре угла зрения на одно и то же предсказание предиктивной обработки: метаболический аргумент специфицирует, почему кэширование навязано; клинический аргумент показывает, как кэшированный инвентарь выглядит извне; предиктивно-вычислительный аргумент называет механизм; развивающий аргумент прослеживает, как инвентарь собирается на протяжении жизни.

2.3 Три кандидат-рамки

Мы можем расположить вопрос вычислимости в пределах трёх рамок, эксплицитно или имплицитно предлагавшихся для естественного языка. У каждой — весомая история; каждая схватывает нечто реальное; каждая, как мы утверждаем, является неверной рамкой для нашего узкого домена. Три рамки касаются *теории языка как вычислимого объекта*; рамка предиктивной обработки из §2.2 специфицирует *агента, внутри которого производится язык*. Эти два вопроса отделимы, и мы занимаем позиции по обоим.

Первая рамка — *радикальная неисчерпаемость*. В этом взгляде, ассоциируемом широко с романтической и постструктуралистской традициями, но живущем и в значительной части современной литературной теории и континентальной философии языка, речь фундаментально креативна. Всякое высказывание — новое событие; смысл возникает в неуправляемой встрече говорящего, слушающего, истории и мира; всякая попытка задать конечный базис промахивается мимо того самого, что делает язык языком. В этой рамке формализация не просто трудна — она категориально ошибочна. Мы согласны, что этот взгляд схватывает нечто важное о некоторых доменах языка:

поэтической речи, философском аргументе, живой метафоре, семантической креативности ребёнка, усваивающего первый язык. Мы не согласны, что он схватывает всё, и отрицаем, что он схватывает наш домен.

Вторая — *генеративная грамматика*, ассоциируемая с Хомским (1957, 1965) и выросшими из его работ вычислительно-лингвистическими традициями. В этом взгляде синтаксическая компетенция опирается на конечный набор правил, способных породить бесконечное множество грамматичных предложений. Программа Хомского показала, что формализация на конечном базисе возможна по крайней мере для одного существенного компонента языка — синтаксиса, — и эта демонстрация остаётся, с нашей точки зрения, одним из великих достижений когнитивной науки XX века. Но генеративная грамматика намеренно выносит за скобки семантику, прагматику и — что принципиально для наших целей — биологический и аффективный субстрат речи. Она сообщает нам, что *форма* обозрима; она явно молчит о том, обозрима ли *функция* тем же образом.

Третья — то, что мы будем называть *ограниченной комбинаторностью*: взгляд, согласно которому специфический домен сложной системы может допустить формализацию на конечном базисе, когда этот домен ограничен порождающим принципом, который сам конечен. Это архитектура ДНК, где четыре нуклеотида, составляемые биохимическими правилами, порождают всё пространство биологических форм; шахмат, где горстка правил движения фигур порождает пространство, превышающее число атомов в наблюдаемой вселенной, но в принципе исчерпывающе специфицируемое; фонологии, где конечный сегментный инвентарь комбинируется конечным набором правил, давая звуковые системы всех человеческих языков. Ограниченная комбинаторность не утверждает, что порождённое пространство мало. Она утверждает, что порождающий базис — конечен.

Рамка предиктивной обработки, развитая в §2.2, даёт нам *механизм*, стоящий за этим общим принципом, в специфическом случае компенсаторно-защитной речи. Агент, организованный вокруг защищённого *priority*'а, имплементирует свои опережающие ответы как кэшированные политики, и этот инвентарь кэшированных политик есть конечный базис, существование которого защищает настоящая статья. Ограниченная комбинаторность — не просто аналогия, которую мы проводим к ДНК и шахматам; это то, что предиктивная обработка специфически предсказывает для архитектуры такого типа. Аналогии с ДНК и шахматами — полезные интуитивные насосы: они иллюстрируют, как малый порождающий базис может производить астрономическое пространство выходов, — но они не являются аргументом. Аргумент состоит в том, что механизм кэширования политик у агента с защищённым *priority*'ом влечёт конечный операциональный базис.

Мы предлагаем, соответственно, что компенсаторно-защитная речь — это домен, для которого третья рамка верна, с рамкой предиктивной обработки, поставляющей причину, по которой это так. Четыре защитных вектора — конечны. опережение предсказанного стыда — конечный механизм. Различение экологичного и компенсаторного — конечный критерий. Речевой феномен, организованный конечными биологическими структурами и имплементированный

через кэшированно-политический вывод, должен, по механике рамки, наследовать их конечность на уровне описания, который мы будем специфицировать.

2.4 Речь как привилегированное наблюдаемое активного вывода

Внутри рамки предиктивной обработки большинство латентных динамик активного вывода не наблюдаемы напрямую. Мы не можем непосредственно измерить *prior precision*, *likelihood precision* или выбор политики без инвазивной нейрональной записи; мы не можем наблюдать обновление генеративной модели в реальном времени. Рамка описывает когнитивную архитектуру, фундаментальные переменные которой в большинстве своём доступны лишь косвенно.

Речь — специфическое и важное исключение.

Высказывание в рамке предиктивной обработки одновременно является тремя вещами. Оно есть *действие*: моторный выход, выбранный из инвентаря политик организма, чтобы семплировать социальную среду, вызывать ответы и тестировать гипотезы о состоянии собеседника. Оно есть *сенсорная поверхность*: говорящий мониторит собственный выход, и связность, беглость и аффективные качества этого выхода возвращаются как свидетельство для внутреннего моделирования говорящим самого-себя-в-ситуации. И оно есть *blanket state* в смысле формулировки марковских границ Фристана (Friston, 2019; Ramstead et al., 2018): высказывание конституирует пористую границу между двумя выводящими агентами, интерфейс, через который происходит диадический взаимный вывод.

Этот тройной статус делает речь уникально информативной для эмпирического изучения активного вывода в социальном познании. Нейроимиджинг схватывает нейронные корреляты вывода, но не сам вывод. Поведенческое наблюдение схватывает моторный выход, но не его генеративную логику. Речь схватывает и то, и другое: стратегия активного семплирования и структура нижележащей генеративной модели обе читаются в одном и том же наблюдаемом потоке. Говорящий, настаивающий «всё в порядке», демонстрируя при этом повышенный *threat-monitoring* словарь, являет специфический паттерн *precision-misallocation* — высокую уверенность в *prior*'е («я в порядке»), противоречащую сенсорным свидетельствам (сигналам аффективного дистресса). Это не метафорическое применение рамки. Это сама рамка, наблюдаемая через тот медиум, который делает её наиболее читаемой.

Следствие для настоящей статьи существенно. Если речь — привилегированное наблюдаемое активного вывода, то компенсаторно-защитная речь — специфический и теоретически ограниченный подтип — есть привилегированное наблюдаемое специфического и теоретически ограниченного класса конфигураций активного вывода. Это то, что делает её формализацию и необходимой, и возможной. Необходимой — потому что феномен наблюдаем и важен, а отсутствие формализации означает, что поле описывает его без механизма. Возможной — потому что нижележащие динамики перечислимы по принципам самой рамки; не по указанию аналитика, а по структуре генеративной модели, производящей эту речь. Вопрос, который задаёт статья II, — не о том, накладывать ли

формализм на сопротивляющийся объект; это вопрос о том, артикулировать ли тот формализм, который нижележащая архитектура уже имплицитно определила.

2.5 Вклад и дорожная карта

Вклад статьи II — сформулировать утверждение о вычислимости точно, внутри рамки предиктивной обработки, выстроить сходящиеся аргументы, его поддерживающие, специфицировать алфавит операций, на котором покоится формализация, и охарактеризовать четыре precision regime, которые эти операции прочерчивают как эмерджентные аттракторы. Мы не строим в настоящей статье аннотационную онтологию или классификационный движок; это работа статьи III. Но мы предоставляем то, что статья III требует на вход: теоретически обоснованный список операций, характеризацию режимной структуры, в пределах которой эти операции группируются, спецификацию критериев, по которым экологичное и компенсаторное выражения должны различаться на уровне операций, и эксплицитный методологический аргумент, что это пространство гипотез — будучи конечным и теоретически обоснованным — санкционирует онтологически-ограниченную классификацию, операционализируемую в статье III.

Раздел 3 развивает четыре аргумента в пользу конечного операционального базиса: метаболический, клинический, предиктивно-вычислительный и развивающий. Раздел 4 формулирует сам алфавит: четыре организованных по векторам семейства речевых операций, с примерами и с эксплицитными терминологическими границами относительно *шаблонов* (которыми они не являются) и *режимов* (в которые они складываются). Раздел 5 представляет четыре precision regime — DEFENCE, COGNITIVE, ILLUSION, META — как эмерджентные аттракторы внутри архитектуры предиктивной обработки; формальная математическая характеристика этих режимов вместе с линейно-гауссовской toy-моделью и фальсифицируемыми предсказаниями вынесена в Приложение А. Раздел 6 рассматривает композицию и контекст. Раздел 7 развивает различитель экологичного/компенсаторного на уровне операций как непрерывный многокомпонентный скаляр. Раздел 8 формулирует ограничения и объём, эксплицитно описывает три уровня описания, к которым обращается серия (биологические векторы, precision regimes, режимы уровня движка), снимает терминологическую неоднозначность между этими уровнями и специфицирует методологический мост к статье III. Раздел 9 заключает.

3. Четыре аргумента в пользу конечного операционального базиса

Утверждение, что компенсаторно-защитная речь покоится на конечном базисе операций, может, при первом прослушивании, показаться либо тривиально верным (разумеется, есть лишь некоторое число способов быть защитным), либо неправдоподобно редуктивным (но, конечно, всякое высказывание, включая всякое защитное высказывание, бесконечно разнообразно в деталях). Мы утверждаем, что ни одно из этих прочтений не верно, и что правильная формулировка — *операции конечны; последовательности, ими порождаемые,*

— *нет* — поддерживается четырьмя независимыми линиями свидетельств, каждая из которых приходит из другой интеллектуальной традиции и сходится к тому же структурному выводу.

3.1 Метаболический аргумент

Первый аргумент исходит из соображения метаболической цены, и мы должны с самого начала признать, что его легко перечитать сильнее, чем следует. Общий принцип — *устойчивые поведения в предиктивных системах тяготеют к кэшированным имплементациям* — это почти трюизм любого энергетически ограниченного организма. Всякое достаточно частое поведение, от ходьбы до набора текста и обыкновенного разговора, работает на маршрутах, а не на посекундной сборке. Если бы это было всё, что мы утверждаем, аргумент доказывал бы нечто для всякого устойчивого поведения и, следовательно, ничего отличительного для защиты.

Что делает защитный случай действительно тугим и что даёт этому аргументу его реальную силу, — это не общий принцип, а *пересечение трёх специфических условий*, все из которых удерживаются для устойчивой компенсаторно-защитной речи и ни одно из которых не удерживается для обыкновенной речи.

Первое условие — *повышенная базовая цена*. Обыкновенное производство речи едет на фоновой нейронной активности. Производство защитной речи едет на *мобилизованной* защитной физиологии — симпатической активации в *fight* и *flight*, вентрально-вагальной-с-аффилиативным-переключением в *fawn*, дорсально-вагальном субстрате в *freeze*. Мобилизация, во всех четырёх векторах, по эволюционному дизайну является высокоценным состоянием. Автономные, эндокринные и иммунные сигнатуры, обозранные в §6 статьи I, не даются бесплатно; они дороги, потому что были отселектированы для коротких, острых вовлечений, в которых энергетический расход разменивался на выживание в непосредственной экстренной ситуации (Porges, 2011; McEwen, 1998).

Второе условие — *отсутствие разрядки*. Острые защитные ответы были спроектированы запуститься, разрешиться и затем вернуться к базовой линии. Компенсаторная защита, как мы утверждали в §5 статьи I, есть опережающий ответ на *предсказанный* стыд относительно жёсткого социально-нормативного *prior*'а — и предсказанная угроза не разрешается, потому что жёсткий *prior* не обновляется. Социальный мир, порождающий предсказание стыда, присутствует почти в каждом взаимодействии. Нет ясного внешнего сигнала, позволяющего защитной физиологии разрядиться. Организм запускает ответ, спроектированный для эпизодического вовлечения, в хроническом режиме.

Третье условие — *непрерывная работа*. Производство языка, внутреннее или внешнее, принадлежит к самым темпорально плотным деятельности, в которые вовлечён человек. Человек, разговаривающий с другими несколько часов в день и ведущий внутренний монолог большую часть остального, производит речь почти непрерывно во время бодрствования. Если эта речь работает в компенсаторно-защитном режиме — если защитный регистр доминирует, — то поддерживающая защитная физиология работает не только на длительности эпизодических вовлечений, но на темпоральном пролёте всего дня.

Пересечение этих трёх условий — повышенная базовая цена, отсутствие разрядки, непрерывная работа — производит селекционное давление на имплементацию компенсаторно-защитной речи, существенно более тугое, чем давление на обыкновенную речь. Обыкновенная речь не производится на мобилизованном субстрате, получает массу интервалов разрядки и работает на физиологии, относительно недорогой на базовой линии. У компенсаторно-защитной речи нет ни одного из этих облегчений. Она работает на дорогом субстрате, не разряжается и работает непрерывно. Результат — метаболический режим, в котором неэффективность имплементации становится, в масштабах развития и продолжительности жизни, физиологически нежизнеспособной.

Вывод, который мы извлекаем из этого пересечения, — поэтому не просто, что компенсаторно-защитная речь *тяготеет* к кэшированным операциям (общий принцип, с которого мы стартовали), а нечто более сильное: что она *должна быть* поддержана кэшированным операциональным базисом, и что операции, составляющие базис, должны быть *специфически настроены на защитную задачу*, а не быть общими когнитивными маршрутами, рекрутируемыми ad hoc. Общие маршруты были бы менее эффективны, чем специализированные, и под описанным селекционным давлением даже малые выигрыши в эффективности накапливаются в развивающем времени в разницу между физиологической жизнеспособностью и коллапсом.

Нам не нужно для целей этого аргумента специфицировать нейронный или психолингвистический субстрат кэширования. Кандидатами являются долговременная потенция специфических синтактико-прагматических конструкций, осадочные ассоциативные цепи между сигналами предсказания стыда и предфабрикованными классами ответов, аттракторная динамика в крупномасштабных кортикальных сетях. Аргумент не покоится ни на одном из них. Он покоится на логической структуре самого метаболического ограничения. Каким бы ни оказался механизм кэширования, *нечто* должно выполнять эту функцию, потому что феномен, описанный в статье I, не может иначе существовать в форме и длительности, в которой он эмпирически существует.

Внутри рамки предиктивной обработки, принятой в §2.2, этот аргумент специфицирует, чем должен быть механизм кэширования. Кэширование в рамке есть подкрепление политики при минимизации свободной энергии; метаболическое давление — это селекционная сила, обеспечивающая эффективное удержание политик. Метаболический аргумент, следовательно, не отдельное соображение, а первое лицо того же самого предсказания предиктивной обработки, к которому следующие три аргумента подойдут с других сторон.

3.2 Клинический аргумент: стереотипность как свидетельство

Второй аргумент приходит с направления, которое легко недооценить, если слишком долго пробыть в теоретических краях и недостаточно — в клинических. Столетие систематического клинического наблюдения — через психоаналитическую, объектно-реляционную, attachment-, социально-когнитивную, семейно-системную и клинически-травматическую традиции —

задокументировало, через континенты, школы и теоретические приверженности, замечательную *стереотипность* защитных паттернов.

Слово это — «стереотипность» — легко произнести и трудно обосновать. Чтобы клинический аргумент делал настоящую работу, нам нужно показать, что схождение через традиции является подлинным: что когда клейнианский аналитик описывает проективную идентификацию, когнитивно-поведенческий исследователь описывает hostile attribution bias, а attachment-кодировщик описывает лингвистическую сигнатуру поглощённой злости, они описывают — на уровне наблюдаемой речи и поведения — *один и тот же феномен*, даже если их теоретические словари расходятся резко в том, что этот феномен *означает*. Мы предлагаем четыре таких кросс-традиционных триангуляции, по одной на семейство защитного вектора, в качестве свидетельства.

Схождение семейства fight: атрибутивная враждебность. Klein (1946) и Bion (1962) описали *проективную идентификацию*, при которой аспекты собственного невыносимого переживания пациента атрибутируются собеседнику, который затем встречается как содержащий эти качества. Социально-когнитивное исследование независимо разработало в другом словаре конструкт *hostile attribution bias* (Dodge, 1980; Crick & Dodge, 1994), задокументировав в экспериментальных задачах тенденцию определённых популяций считать нейтральное или неоднозначное поведение как несущее враждебное намерение. Исследование взрослой привязанности с помощью кодировочной системы Adult Attachment Interview (Main & Goldwyn, 1998; Hesse, 2016) идентифицировало специфическую лингвистическую сигнатуру «злостно-поглощённой» речи о прошлых заботящихся фигурах, маркируемую именно необоснованными атрибуциями негативного намерения. Три традиции, три словаря, один узнаваемый поведенческий паттерн: атрибуция враждебной внутренности другому в отсутствие обоснования.

Схождение семейства fawn: компенсаторное самостирание. Winnicott (1960) описал *ложное Я*, в котором самопредъявление ребёнка построено на удовлетворении нужд заботящегося взрослого, а не на выражении собственных нужд. Bowen (1978) описал *созависимость* в семейных системах, с её характерным паттерном само-уничтожения в служении реляционной поддержке. Ainsworth и её преемники разработали категорию привязанности, первоначально названную «тревожно-сопротивляющейся», а позднее уточнённую как «тревожно-поглощённую» (Hazan & Shaver, 1987; Mikulincer & Shaver, 2016), в которой взрослое реляционное поведение организовано вокруг *гиперактивирующих стратегий* — подстройки, поиска близости через выравнивание, подавления собственного выражения потребности. Walker (2013) независимо описал *fawn* как ответ в комплексной травме, с его характерными лингвистическими маркерами угодничества и самоминимизации. Четыре традиции — объектно-реляционная, семейно-системная, attachment-, клинически-травматическая — сходятся к одному и тому же речевому феномену: систематическое стирание собственной позиции говорящего в предвосхищающем служении собеседнику.

Схождение семейства flight: абстракция как дистанция. Anna Freud (1936) назвала *интеллектуализацию* использованием абстрактного рассуждения для избегания эмоционального контакта с угрожающим содержанием. Bowlby (1980)

описал *защитное исключение* — систематическое преграждение аффективного материала от репрезентации. Исследование привязанности идентифицировало характерную *деактивирующую стратегию* избегающе-отрицающих взрослых, с её лингвистической сигнатурой аффективно-лёгкого, обобщённого, бедного историей нарратива, когда Adult Attachment Interview зондирует attachment-релевантный материал (Main & Goldwyn, 1998). Конструкт *алекситимии* (Sifneos, 1973; Taylor et al., 1997) задокументировал трудность перевода аффективного переживания в специфическое лингвистическое содержание. Fonagy и коллеги (2002) описали *псевдо-ментализацию* — производство рефлексивно-звучащего языка без аффективного заземления. Пять традиций, пять словарей, один узнаваемый паттерн: ответ на аффективно-живой материал поднятием в абстрактный или обобщённый регистр.

Схождение семейства freeze: эпистемическое отсутствие. Schauer и Elbert (2010) задокументировали *диссоциативную речь* у выживших в травме, характеризуемую нарушенной нарративной связностью и утратой от-перволица-агентности. Main и Hesse (1990) идентифицировали *неразрешённую/дезорганизованную* привязанность через её AAI-сигнатуру характерных провалов мониторинга дискурса — речь, которая распадается, становится расплывчатой или теряет референциальное заземление именно тогда, когда зондируется травма-релевантный материал. Van der Kolk (2014) синтезировал через эти традиции поведенческую феноменологию freeze-ответов у выживших в развивающей травме. Три традиции сходятся к одному и тому же наблюдаемому паттерну: речь, отказывающаяся конституировать связную, оцениваемую от-перволица-позицию тогда, когда такая позиция была бы уместна.

Мы хотим отметить в этой точке различие, которого требует аргумент. Утверждение, которое мы делаем, — это утверждение о схождении на *наблюдательном* уровне: о том, что разные традиции регистрируют как происходящее в речи и поведении их субъектов. Мы *не* утверждаем, что традиции сходятся в том, что это поведение *означает* теоретически, — и в большинстве случаев они явно не сходятся. Клейнианец полагает, что проективная идентификация вовлекает бессознательную фантазию изгнания и содержания; когнитивно-поведенческий исследователь полагает, что hostile attribution bias — это усвоенный паттерн обработки информации; attachment-кодировщик полагает, что лингвистическая сигнатура отражает внутренние рабочие модели отношений. Это *разные* теоретические приверженности, и мы не занимаем позиции относительно того, какая из них верна. Для аргумента о вычислимости важно то, что обученные наблюдатели из этих разных теоретических традиций, глядя на одни и те же речевые образцы, согласились бы в том, *что происходит*, даже когда они не согласны в том, *почему*. Наблюдательное схождение — вот что даёт нам опору; теоретическое расхождение — то, что удерживает поле честным относительно его нерешённых проблем.

Вывод, следовательно, тот же, что мы сформулировали в открытии, но теперь с содержанием: стереотипность компенсаторно-защитной речи через клинические традиции — не расплывчатое впечатление. Это задокументированное схождение независимых наблюдательных систем на небольшом числе узнаваемых

паттернов. Если бы феномен был подлинно неограниченным, никакое такое схождение не было бы возможно. То, что схождение существует, в той форме, в которой существует, через традиции, через которые существует, — это эмпирическое основание, на котором стоит утверждение о конечном операциональном базисе.

Прочитанное через рамку предиктивной обработки из §2.2, это схождение получает ясное объяснение: независимые наблюдательные системы сходятся к одним и тем же повторяющимся паттернам, потому что эти паттерны суть внешний отпечаток общего архитектурного процесса — развёртывания кэшированных политик против предсказанной шейм-ошибки, реализуемого через индивидов с разными историями, но сталкивающихся со структурно сходными предиктивными требованиями. Клиническая литература в этом прочтении — самый обширный натуралистический архив предиктивно-процессинговой динамики, которым мы сейчас располагаем, собранный на протяжении столетия наблюдателями, большинство которых, в общем и целом, вовсе не думали о предиктивной обработке.

3.3 Предиктивно-вычислительный аргумент

Третий аргумент называет механизм, который первые два подразумевают. Если компенсаторно-защитная речь работает на кэшированном операциональном базисе (§3.1), и если этот базис детектируется через независимые клинические традиции (§3.2), очевидный вопрос таков: *какой процесс строит и поддерживает этот кэш?* Рамка предиктивной обработки, принятая в §2.2, отвечает на этот вопрос с необычной специфичностью, и именно эта специфичность делает третий аргумент более, чем переформулировкой первых двух.

Внутри рамки мозг непрерывно предсказывает собственные сенсорные и аффективные состояния (Clark, 2013; Hohwy, 2013; Seth, 2021). Стыд, как утверждала статья I, возникает как ошибка предсказания между жёстким социально-нормативным *prior*'ом и текущим воспринимаемым Я-в-ситуации. Но этот сигнал ошибки предсказания сам порождается предсказанием: система предсказывает, забегая вперёд, произведёт ли текущая траектория самопредъявления шейм-ошибку или нет. Когда этот *предсказанный* шейм-сигнал пересекает порог, система порождает опережающий ответ — защитный речевой акт — чтобы изменить траекторию до того, как ошибка консолидируется.

Это помещает защитную речь в специфическую архитектурную позицию: она есть *вывод о выводе*, или точнее — выбор политики относительно собственной ожидаемой траектории ошибки предсказания агента. Защитный речевой акт не есть, на самом глубоком уровне, ответ на реальность; это ответ на собственное предсказание системы о том, как траектория вот-вот развернётся относительно защищённого *prior*'а. Следствие состоит в том, что защитные политики могут быть оценены *только* относительно их предсказанного эффекта на собственный внутренний сигнал агента, а не прежде всего относительно внешних исходов. Защитный ход «срабатывает», с точки зрения самой системы, когда он предотвращает или ослабляет предсказанный сигнал стыда, независимо от его внешних эффектов на собеседника или на ситуацию. Эта асимметрия —

что критерий успеха защитного хода внутренний, а не внешний, — есть, как мы будем утверждать в §5, то, что производит характерные *precision biases*, проявляемые компенсаторно-защитной речью.

Эта оценочная структура и есть то, что производит кэширование. Политика, надёжно ослабляющая предсказанный стыд в определённом классе ситуаций, будет подкрепляться всякий раз, когда она срабатывает. Политика, не срабатывающая, будет деприоритизирована. На масштабах развития система сходится к малому инвентарю политик, каждая из которых настроена на определённый класс шейм-предсказаний, каждая отобраена за её надёжный ослабляющий эффект. Это схождение — не побочный эффект рамки; это то, что рамка предсказывает для всякой системы, имплементирующей опережающую защиту относительно защищённого *prior'a*.

Получающийся инвентарь имеет структурные свойства, которые мы можем специфицировать. Во-первых, он *организован по векторам*: четыре биологических защитных вектора из статьи I предоставляют четыре различных механизма ослабления предсказанного стыда (атаковать источник, уйти от него, умиротворить его, диссоциироваться от него), и политики внутри каждого векторного семейства разделяют механизм-уровневые сходства, отличающие их от политик в других векторах. Во-вторых, он *конечен*: число механически различных политик ограничено структурой самой защитной архитектуры, а не перечислительным удобством. В-третьих, он *композируем*: одни и те же ситуации часто активируют несколько векторов, и политики из разных семейств могут разворачиваться последовательно или в комбинации внутри одного высказывания, — композируемость, которую мы рассматриваем подробнее в §6. В-четвёртых, что важно для §5, он *характеризуется precision biases*: каждая политика работает, регулируя баланс между *prior precision* и *likelihood precision* специфическим образом, и эти регулировки агрегируются, на масштабах жизни агента, в четыре *precision regime*, которые мы разрабатываем в §5.

Связь с остальной серией становится прямой в этой точке. Если защитные политики кэшируются в ограниченном, механически-организованном инвентаре, то политики имеют структурированные наблюдаемые сигнатуры — специфические лингвистические паттерны, являющиеся выходами политик. Эти сигнатуры и есть то, что движок статьи III спроектирован детектировать: классификационный конвейер, пространство гипотез которого — кэшированный защитный инвентарь, оперирующий на речи, его выражающей. Методологическое санкционирование онтологически-ограниченной классификации, которое настоящая статья берётся предоставить, покоится именно на том предсказании, которое рамка предиктивной обработки делает об архитектуре такого типа.

Мы не утверждаем в этом аргументе больше того, что рамка санкционирует. Мы не утверждаем, что смоделировали специфическую нейронную имплементацию кэширования политик. Мы не утверждаем, что рамка — единственное описание, под которым утверждение о конечном базисе защитимо; метаболический и клинический аргументы стоят сами по себе. Что мы утверждаем — уже: *при* рамке предиктивной обработки компенсаторно-защитная речь, охарактеризованная в статье I, *предсказанно* покоится на конечном инвентаре кэшированных политик с четырьмя структурными свойствами выше. Это предсказание

сходится с тем, что §3.1 и §3.2 находят независимо. Схождение — вот что даёт общему аргументу его силу.

3.4 Развивающий аргумент

Четвёртый аргумент приходит из исследований развития и привязанности, которые установили — через лонгитюдные исследования, клиническое наблюдение и кросс-культурное сравнение, — что защитные паттерны *кристаллизуются*. Ребёнок, у которого складывается избегающий паттерн привязанности, не импровизирует новую стратегию избегания в каждой новой ситуации; к середине детства у него есть узнаваемый репертуар ходов, которые он разворачивает повторяемо, с малыми вариациями, подстроенными под контекст, во всём диапазоне реляционной жизни. То же верно для тревожно-поглощённых паттернов, дезорганизованных паттернов и fawn-конфигураций, задокументированных Walker (2013) при комплексной развивающей травме. Эти паттерны — не рыхлые диспозиции; они специфичны и воспроизводимы.

Развивающие исследования, наиболее прямо релевантные нашему аргументу, приходят из исчерпывающего обзора Mikulincer и Shaver (2016) паттернов взрослой привязанности в речи и реляционном поведении. Их мета-аналитическая работа показывает, что лингвистические и поведенческие сигнатуры, связанные с паттернами привязанности, устойчивы через десятилетия, согласованы между оценивающими и предсказуемы по отношению к последующим исходам. Работа Fonagy и коллег (2002) о шкале рефлексивной функции аналогично показывает, что провалы ментализации принимают узнаваемые, повторяющиеся формы — и эти формы группируются в небольшое число типов, а не распределяются непрерывно по безграничному пространству.

Развивающий аргумент, следовательно, таков: защиты *усваиваются*, а то, что усваивается, склонно кристаллизоваться в репертуар. Это верно через домены — моторные навыки, социальные сценарии, экспертное суждение, — и нет оснований думать, что защитная речь исключительна. Компенсаторно-защитная речь, которую производит данный взрослый сегодня, не изобретается заново; это осадочный, консолидированный результат тысяч развивающих моментов, в которых специфические защитные ходы либо срабатывали (и подкреплялись), либо проваливались (и деприоритизировались). Взрослый говорящий не создаёт репертуар. Он его разворачивает.

Внутри рамки предиктивной обработки эта развивающая траектория — и есть процесс сборки, в результате которого строится кэшированный инвентарь политик из §3.3. Каждый развивающий момент, в котором защитный ход успешно ослабил предсказанный шейм-сигнал или не смог этого сделать, есть точка данных в жизненно-масштабном подкреплении политик, производящем взрослый репертуар. Устойчивость и кросс-контекстная согласованность, которые attachment-исследование документирует на поведенческом уровне, — наблюдаемый след этой сборки: к взрослости инвентарь устойчив, потому что он прошёл обширную селекцию, а устойчивые инвентари производят согласованные выходы через ситуации. Развивающий аргумент таким образом специфицирует временное измерение предиктивно-вычислительного аргумента

— как кэшированный инвентарь приходит к существованию в той форме, в которой с ним встречаются клинические наблюдатели из §3.2.

3.5 Схождение и что оно подразумевает

Четыре аргумента, взятые из четырёх традиций — физиологии, клинического наблюдения, вычислительной нейронауки, развивающей психологии, — сходятся к одному и тому же структурному выводу: компенсаторно-защитная речь покоится на конечном, кэшированном, переиспользуемом, кристаллизованном операциональном базисе. Схождение не случайно. Каждый аргумент может быть прочитан, как §2.2 предвосхитил, как иной угол зрения на одно и то же нижележащее предсказание рамки предиктивной обработки: что агент, организованный вокруг защищённого *prior*'а и имплементирующий опережающую защиту относительно предсказанной ошибки предсказания в этом *prior*'е, должен аккумулировать конечный, механически-организованный инвентарь политик на масштабах развития. Физиология (§3.1) даёт селекционное давление, обеспечивающее эффективность. Клиническое наблюдение (§3.2) даёт внешний регистр содержимого инвентаря. Предиктивно-вычислительный аргумент (§3.3) специфицирует механизм, которым инвентарь строится и поддерживается. Развивающее исследование (§3.4) даёт процесс сборки, в ходе которого инвентарь конструируется. Каждый аргумент в отдельности может быть оспорен; их схождение и их взаимное соответствие с рамкой предиктивной обработки делают оспаривание существенно труднее. Не просто возможно, что операциональный базис конечен; это, с учётом веса свидетельств, маловероятно обратное.

Непосредственное следствие состоит в том, что вопрос о вычислимости получает утвердительный ответ на уровне самого операционального базиса. Если базис конечен, он может быть перечислен; если он может быть перечислен, он может быть маркирован; если он может быть маркирован, он может быть распознан обученным аннотатором-человеком или вычислительным классификатором. Формализация, с этой точки зрения, — не надежда и не устремление. Это прямое следствие того, чем компенсаторно-защитная речь структурно является.

Чего формализация сделать не может — и чего мы тщательно не утверждаем — это исчерпать *продукты* операций. Последовательности компенсаторно-защитной речи, которые данный говорящий производит в данной жизни, столь же разнообразны, сколь разнообразны жизни. Но разнообразие на уровне последовательностей совершенно совместимо с конечностью на уровне операций, как проясняют аналогии с ДНК и шахматами. Четыре нуклеотида порождают биосферу; около сорока дебютных выборов, горстка правил движения фигур и шестидесятичетырёхклеточная доска порождают десять-в-сто-двадцатой шахматных партий. Конечность живёт в порождающем базисе, а не в том, что он порождает.

4. Алфавит: операции, не шаблоны

Утверждая, что компенсаторно-защитная речь покоится на конечном операциональном базисе, мы теперь специфицируем, что в этом базисе содержится. Настоящий

раздел вводит сам алфавит: организованный по векторам инвентарь речевых операций. Прежде чем перечислять их, мы фиксируем терминологию, потому что поле вокруг защитной речи замусорено почти-синонимами, и точность формализации зависит от того, как чисто мы их различим.

4.1 Операции, шаблоны, режимы

Шаблон — это фиксированная лингвистическая форма: готовая фраза, клише, предфабрикованное предложение, разворачиваемое как-есть. Шаблоны существуют в языке (приветствия, соболезнования, некоторые формулы вежливости), и малое подмножество защитной речи их инстанцирует. Но шаблоны — не то, что мы понимаем под операциями, и смешивание одного с другим было бы, как мы полагаем, самым быстрым способом дискредитировать весь комбинаторный аргумент. Если бы операции были шаблонами, они были бы слишком ригидны, чтобы объяснить наблюдаемое разнообразие защитной речи; критики (справедливо) указали бы, что реальная защитная речь порождающая, а не вырезано-наклеена. Мы согласны, и мы предлагаем операции именно потому, что они не шаблоны.

Операция, в нашем смысле, — это *речевой порождающий ход*: правило, преобразующее входной контекст в выходное высказывание, выполняющее специфическую защитную функцию. Полезная аналогия — шахматный ход. У коня нет «шаблона хода»; у него есть *правило движения* — две клетки в одном направлении и одна перпендикулярно. Из этого правила, соединённого с текущей конфигурацией доски, порождается любой конкретный ход конём. Два хода конём из двух разных позиций разделяют одно и то же нижележащее правило, но производят совершенно разные конкретные ходы. Так же с операциями: «yoу-атрибуция» — это правило преобразования внутреннего переживания, которое говорящий не может признать, в высказывание во втором лице о собеседнике; и конкретные высказывания, порождённые применением этого правила в разных контекстах, варьируются неограниченно, сохраняя правило.

Режим — это *устойчивая конфигурация операций*, которую говорящий принимает на отрезке дискурса. Там, где операция — это отдельный ход, режим — это стратегия, в пределах которой ходы выбираются. Говорящий, хронически находящийся в проективном режиме, разворачивает преимущественно операции семейства *fight*, с характерными паттернами со-выбора и характерным избеганием операций из других семейств. Режимы, таким образом, — это то, что клиническая традиция в целом называла «защитными паттернами» или «защитными стилями», и в нашей архитектуре они *эмерджентны из вероятностей выбора операций*, а не примитивны сами. Статья IV развернёт формализацию режимов и их траекторий; здесь мы их именуем лишь затем, чтобы отграничить от операций, которые являются более низкоуровневым и более непосредственно вычислимым примитивом.

Отношение между тремя уровнями, таким образом: **шаблоны < операции < режимы**. Шаблоны — не наш примитив, потому что слишком ригидны; режимы — не наш примитив, потому что эмерджентны; операции — правильный

уровень описания, потому что они порождающие (в отличие от шаблонов) и непосредственно наблюдаемы в отдельных высказываниях (в отличие от режимов, которым требуются более длинные отрезки дискурса, чтобы стать видимыми).

4.2 Операции семейства *fight*

Вектор *fight* сохраняет свою направленную вовне логику при переходе в речь, и семейство операций, через которое он это делает, разделяет сигнатуру направления атрибутивной или оценивающей силы на собеседника или представленное третье лицо. Мы перечисляем здесь ядерные операции семейства *fight*, с пониманием того, что этот список предварительный — статья III уточнит его через операционализацию и проверку.

You-атрибуция. Говорящий делает декларативное утверждение о внутреннем состоянии, мотивации или характере собеседника в форме, предполагающей привилегированный доступ к этому состоянию. Примеры: «тебе всё равно»; «ты пытаешься меня контролировать»; «ты считаешь, что ты лучше меня». Операция состоит в преобразовании неразрешённого внутреннего предсказания («моими нуждами пренебрегают») в декларатив во втором лице о другом. Ключевой маркер — утверждение внутреннего мира собеседника в отсутствие эпистемического обоснования.

Развёртывание глобального квантора. Говорящий характеризует поведение собеседника кванторами «всегда»/«никогда», превышающими объём текущего взаимодействия. Примеры: «ты всегда так делаешь»; «ты никогда не слушаешь»; «вот во что ты превращаешь любой разговор». Операция преобразует жалобу в настоящем времени в утверждение о паттерне и тем самым устраняет возможность контекстной починки.

Экстернализованно-агентная рамка. Говорящий пересказывает взаимодействие в грамматической конструкции, удаляющей его собственный вклад из отчёта. Примеры: «и тогда ты начал меня атаковать»; «и ты меня расстроил». Операция систематически помещает собеседника в позицию каузального агента, а говорящего — в позицию получателя воздействия, безотносительно к тому, чем взаимодействие было структурно.

Мотив-атрибуция. Говорящий декларирует не только то, что собеседник сделал, но и почему, в терминах, которые собеседник не подтверждал. Примеры: «ты сказал это только чтобы причинить мне боль»; «настоящая причина, по которой ты поднял этот вопрос...». Операция близка к *you-атрибуции*, но специфически касается каузально-интенциональной интерпретации конкретного действия, а не общего утверждения о характере собеседника.

У каждой из этих операций есть её экологичный двойник. «Я замечаю, что ты не ответил на мои последние три сообщения; мне интересно, не поняла ли я что-то неправильно» — не операция *fight*; это соразмерный первого-лица-запрос, сохраняющий эпистемическое смирение. Операциональное семейство — это один и тот же нижележащий вектор; экологично-компенсаторное различие, которое мы развиваем в §6, проходит через семейство на уровне,

специфицирующем, как операция откалибрована к контексту.

4.3 Операции семейства flight

Вектор flight сохраняет свою направленную к отступлению логику при переходе в речь. Семейство операций flight разделяет сигнатуру производства семантической дистанции между говорящим и темой, которая, будучи затронутой, составила бы контакт.

Сдвиг абстрактного регистра. Говорящий отвечает на конкретный, аффективно-живой вопрос ответом, выстроенным в общем или теоретическом регистре. Пример: на вопрос «что ты чувствуешь по поводу болезни матери?» говорящий отвечает «интересно, как мы конструируем смысл вокруг смертности». Операция преобразует первого-лица-аффективный запрос в третьего-лица-абстрактную рефлексию, номинально сохраняя тему и удаляя из неё Я.

Дрейф темы. Говорящий затрагивает вопрос частично, а затем сдвигает референциальную мишень до того, как исходный вопрос разрешён. Пример: на вопрос о конкретном инциденте говорящий начинает ответ, отсылающий к инциденту, но затем мигрирует к другой, родственной-но-безопасной теме. Операция узнаётся по характерному отсутствию возвращения к исходному вопросу даже при приглашении.

Замена мета-комментарием. Говорящий подменяет комментарий о разговоре продолжением участия в нём. Пример: «это становится напряжённым»; «я замечаю, что мы всегда к этому приходим»; «не правда ли интересно, что мы продолжаем сюда возвращаться». Операция удаляет говорящего из первопорядкового взаимодействия, возводя его в позицию наблюдателя-этого-взаимодействия.

Смещение времени или модальности. Говорящий отвечает на аффективный вопрос в настоящем времени прошедшим, условным или гипотетическим. Пример: на вопрос «что ты хочешь прямо сейчас?» говорящий отвечает «ну, я думаю, что человек в моём положении обычно хотел бы...». Операция дистанцирует Я во времени или модальности от аффективного живого провода вопроса.

Как и в семействе fight, у каждой из этих операций есть экологичный двойник. «Давай сделаем паузу в этом разговоре; мне нужно подумать, и я хочу вернуться к нему завтра» — это вектор flight, развёрнутый экологично. Он назван как отступление, соразмерен ситуации и сохраняет условия для возвращения.

4.4 Операции семейства fawn

Вектор fawn сохраняет свою логику подчиниться-чтобы-умиротворить при переходе в речь. Семейство операций fawn разделяет сигнатуру минимизации собственного контура говорящего при одновременном усилении подстройки под предполагаемые предпочтения собеседника.

Наслоение хеджей. Говорящий ставит несколько смягчителей вокруг утверждения, которое их не требует, так что ассертивная сила утверждения истощается

до того, как оно сделано. Пример: «я тут как бы может быть хотела бы узнать, возможно, не могли бы вы...». Операция производит высказывание, синтаксическая форма которого сохранена, но прагматическая сила фактически нулевая, снижая экспозицию говорящего.

Избыток извинений. Говорящий производит извинения или маркеры извинений в контекстах, где никакого нарушения не произошло и ни одно не было сигнализировано. Примеры: «извините, знаю, это глупый вопрос»; «извините, что беспокою»; «прошу прощения, что занимаю ваше время». Операция преобразует нейтральный интерактивный момент в малый ритуал подчинения.

Стирание от-первого-лица-агентности. Говорящий выражает предпочтения, несогласия или желания в грамматических конструкциях, приписывающих их обобщённым другим или никому. Примеры: «люди могут сказать...»; «кто-то в такой ситуации мог бы чувствовать...»; «возникает вопрос, не стоит ли...». Операция удаляет от-первого-лица-присутствие говорящего из утверждений, которые иначе бы его несли.

Опережающее согласие. Говорящий выравнивается с позицией, которую собеседник ещё не занял, опережая возможное несогласие согласием. Пример: «ты, наверное, сейчас скажешь X, и я полностью понимаю, откуда это идёт». Операция схлопывает интерактивное пространство, в котором могло бы возникнуть настоящее несогласие, до того, как оно возникнет.

Вставка восходящего груминга. Говорящий вставляет комплименты, аффирмации или выражения благодарности в точках, не имеющих компенсирующего содержания. Пример: ответ на нейтральный вопрос с «ты всегда задаёшь такие вдумчивые вопросы» перед собственно ответом. Операция производит лингвистический груминг, не будучи запущенной самим взаимодействием.

Экологичные двойники здесь особенно важно удерживать в поле зрения, потому что *face* — это защитное семейство, наиболее легко морализуемое. Младший клиницист, который говорит старшему супервизору: «Я доверяюсь вашему суждению в дифдиагностике — вы видели сотни таких случаев, а я видела один», — разворачивает вектор *face* экологично: иерархия реальна, признание соразмерно, и собственная перспектива говорящего не стёрта.

4.5 Операции семейства freeze

Вектор freeze сохраняет свою логику обездвиживания-как-исчезновения при переходе в речь. Семейство операций freeze разделяет сигнатуру производства речи, сопротивляющейся оценке, — отказывающейся конституировать позицию, с которой собеседник мог бы работать.

Эпистемическая неопределённость. Говорящий разворачивает эпистемические хеджи («кажется», «полагаю», «вроде как», «я не очень знаю») в контекстах, где доступ к релевантному содержанию был бы обычно доступен. Пример: на вопрос, что она хочет на ужин, говорящая отвечает «не знаю, ну, что-нибудь». Операция специфически направлена на эпистемическое утверждение — притязание говорящего на знание собственного состояния или предпочтения — и ослабляет его.

Отсутствие дейксиса. Говорящий использует расплывчатую или обобщённую референцию там, где специфическая референция была бы более информативной. Примеры: «всё сейчас вроде как сложно»; «это было много всего». Операция выкачивает дейктические якоря (конкретных лиц, времена, события, эмоциональные качества) из высказывания, которое номинально их касается.

Поломка референции. Говорящий использует местоимения или описания, не указывающие на определённые референты. Пример: «они что-то сказали, и потом это произошло, и теперь всё как-то странно». Операция производит речь, поверхностная форма которой грамматична, но пропозициональное содержание неопределённо.

Отсутствие от-первого-лица-обязательства. Говорящий делает утверждения о собственных состояниях в конструкциях, удаляющих её из утверждения. Примеры: «всё как-то не так»; «что-то странное происходит»; «было тяжело». Операция удаляет Я из от-первого-лица-утверждения, производя нечто, структурно являющееся описанием состояния, не будучи, совсем, собственным описанием говорящего.

Хроническое неопределённое реагирование. Говорящий, при приглашении специфицировать, производит повторяемые неопределённые или минимальные ответы. Пример: на приглашение развернуть «не знаю, было тяжело» говорящая отвечает «да, просто, ну, не знаю, просто вот так». Операция узнаётся по её выживанию при повторных приглашениях к специфичности.

Экологичные двойники операций freeze — это подлинное незнание и принимающее молчание. «Я на самом деле пока не знаю; дай мне подумать, и я вернусь к тебе» — это вектор freeze, развёрнутый экологично. Говорящая признаёт своё актуальное ограничение, остаётся в контакте с разговором и удерживает открытой возможность возвращения к содержанию.

4.6 Алфавит как целое

Четыре семейства вместе — fight, flight, fawn, freeze — дают операциональный алфавит порядка двадцати ядерных операций, с внутренними под-вариациями, несколько расширяющими счёт. Это базис, на котором статья III построит аннотационную онтологию. Несколько черт этого алфавита стоит отметить, прежде чем продолжить.

Первое: операции *не взаимно исключаются* в пределах одного высказывания. Отдельное компенсаторно-защитное высказывание может инстанцировать более одной операции одновременно — например, you-атрибуцию, произведённую с наслоением хеджей, или сдвиг абстрактного регистра, являющийся также мета-комментарием. Формализация должна справляться с этим, и статья III это обеспечивает, допуская мульти-маркировку на операциональном уровне.

Второе: семейства *не взаимно исключаются в пределах репертуара говорящего*. Говорящий может разворачивать операции fight в одном контексте и операции fawn в другом, и многие говорящие так делают. Анализ режимов в статье IV схватывает тенденции выбора операций через время и контекст; сам операциональный уровень молчит о режиме.

Третье — и самое важное: алфавит *не является диагностическим инструментом*. Отдельное высказывание, инстанцирующее компенсаторно-защитную операцию, само по себе не маркер патологии, дисрегуляции или какой-либо иной оценочной категории. Люди разворачивают эти операции постоянно, в каждом разговоре, как часть обыкновенного производства речи под обыкновенным социальным давлением. Патологическая форма, там, где она существует, живёт на уровнях режимов и траекторий (где говорящий живёт во времени? может ли он двигаться?), а не на уровне операций. Мы подчёркиваем это, потому что диагностическое ошибочное прочтение есть предсказуемый режим провала формализации, и мы намерены, чтобы движок статьи III структурно его предотвращал.

5. Четыре *precision regime* как эмерджентные аттракторы

До этой точки статья развивала два уровня описания: биологические векторы статьи I (*fight, flight, fawn, freeze*, сохранённые как направленная логика в речи) и речевые операции §4 (ограниченный инвентарь порождающих ходов, инстанцирующих эти векторы в конкретных высказываниях). Оба уровня необходимы для того, что мы делаем; ни один не самодостаточен в отдельности. Векторы слишком грубы, чтобы обосновать аннотацию: они говорят нам, какую направленную логику организм наследует из своего защитного репертуара, но не то, как конкретное высказывание реализует эту логику в лингвистической форме. Операции слишком локальны, чтобы схватить устойчивые конфигурации, которые клинические наблюдатели узнают и которые любому движку придётся детектировать на отрезках дискурса, а не в отдельных высказываниях. Между вектором и операцией лежит третий уровень, который артикулирует настоящий раздел: *precision regime*.

Precision regime — это конфигурация распределения *precision* внутри иерархической генеративной модели §2.2: насколько вес помещается на *prior*-ожидания относительно поступающих свидетельств, какие каналы *disconfirmatory* информации ослабляются или усиливаются, и с какой уверенностью система фиксируется на выбранных политиках. Четыре режима, которые мы артикулируем ниже, — **DEFENCE, COGNITIVE, ILLUSION, META** — не категории, которые мы накладываем на данные; это аттракторо-подобные конфигурации в пространстве *precision*-параметров, состояния, в которые система оседает, когда определённые сочетания *priors*, свидетельств и политических стимулов устойчивы. Формальная спецификация этих аттракторов — *precision*-гиперпараметры, линейно-гауссовская *toy*-модель и пять фальсифицируемых предсказаний — вынесена в Приложение А. В настоящем разделе мы характеризуем их концептуально и показываем, как каждый из них реализует, на уровне *precision*-динамики, феномены, описанные на уровне операций и векторов.

5.1 Почему четыре?

Число четыре не аксиоматично; оно выводится из двух ортогональных осей вариации в управлении *precision* при иерархическом выводе. Первая ось — *направление precision-смещения*: помещает ли система больший эффективный

вес на защищённые priors (prior-доминируемое обновление) или на поступающие свидетельства (evidence-чувствительное обновление). Вторая ось — *глубина вывода*: трактуются ли precision-параметры как фиксированные на текущем уровне вывода (первого порядка), или precision сам становится объектом вывода, подверженным модуляции и научению (высшего порядка).

Пересечение этих двух осей даёт четыре естественных квадранта. DEFENCE занимает prior-доминируемый, первопорядковый квадрант: защищённые priors доминируют над свидетельствами, и это доминирование само не находится под рефлексивным надзором. COGNITIVE занимает сбалансированный, первопорядковый квадрант: свидетельства и prior взвешиваются контекстно-чувствительно, без модуляции высшего порядка сверх того, что требует задача. ILLUSION занимает родственную, но отличную область prior-доминируемого пространства: доминирование над свидетельствами не требует острой угрозы для самоподдержания, и конфигурация самоподкрепляется, а не эпизодически запускается угрозой. META занимает квадрант высшего порядка, ортогональный первым трём: precision сам становится объектом вывода, позволяя системе модулировать баланс между prior и свидетельствами по мере изменения волатильности и контекста.

Это минимальные квадранты в том смысле, что Bayesian model reduction мог бы, в принципе, найти более тонкую структуру внутри них, — ILLUSION мог бы распасться на подтипы, META мог бы иметь собственную внутреннюю дифференциацию, — но четырёх-режимная архитектура не унаследована как откровение. Это простейшее разбиение, сохраняющее различия, документируемые клинической литературой, при этом оставаясь теоретически обоснованным внутри рамки. Если более позднее исследование структуры найдёт, что ILLUSION распадается на два различных аттрактора с разными precision-сигнатурами, рамка вместит это уточнение без возмущения нижележащей логики; чего она не вмещает — это редукции четырёх к меньшему числу, потому что клиническое сходжение, задокументированное в §3.2, активно аргументирует против такой редукции.

5.2 DEFENCE

DEFENCE — это режим, в котором четыре биологических вектора статьи I, становясь хроническими и опережающими, находят своё precision-теоретическое выражение. Он характеризуется повышенной prior precision на самозащитных priors — интериоризованном социально-нормативном prior'e, относительно которого предсказывается стыд, — и подавленной likelihood precision на каналах, которые несли бы disconfirmatory свидетельства. Эффект состоит в том, что свидетельства, противоречащие защищённому prior'у, систематически снижаются в весе при обновлении убеждений; жёсткий prior остаётся нетронутым, потому что механизм, которым он мог бы быть обновлён, ослаблен. Policy precision на защитных политиках также повышена: система фиксируется на них с необычной уверенностью, потому что критерий успеха защитной политики (§3.3) — это внутреннее ослабление предсказанного стыда, а не внешний исход, и внутреннее ослабление надёжно достигается политиками, кэшированными ровно для этой цели.

Операции §4 — речевые сигнатуры этой precision-конфигурации. You-атрибуция (§4.2) отражает precision regime, в котором защитный prior говорящего о собственном состоянии взвешен столь тяжело, что свидетельства действительного состояния собеседника не могут конкурировать; prior проецируется вовне, прикрепляется к другому и защищается там. Сдвиг абстрактного регистра (§4.3) — это политика, которая по самому своему устройству семплирует разговорную среду так, чтобы минимизировать поток threat-релевантных disconfirmatory свидетельств в генеративную модель. Наслоение хеджей (§4.4) — прагматическая структура, осушающая собственную ассертивную экспозицию говорящего и, вместе с ней, вероятность породить свидетельство, которое запустило бы шейм-предсказание. Ход эпистемической неопределённости (§4.5) ослабляет от-первого-лица-утверждения, истинностные значения которых иначе составили бы свидетельство против защищённой self-модели.

Критическое уточнение: DEFENCE в этом техническом смысле *не* то же, что острое защитное возбуждение перед лицом реальной, непосредственной угрозы. Последнее в терминах предиктивной обработки совершенно уместно: высокая prior precision на «эта ситуация опасна», сопровождаемая сниженной likelihood precision на успокаивающих, но потенциально обманчивых сигналах, — хорошая байесовская политика, когда среда действительно угрожающа. DEFENCE в нашем употреблении относится к хронической, опережающей конфигурации, в которой защищаемый prior — не непосредственная среда, а интериоризованная социально-нормативная self-модель, и фильтруемые свидетельства — не успокоение от потенциального противника, а повседневный социальный вход от обыкновенного собеседника. Архитектурная машинерия та же; различаются защищаемый prior, контекст, в котором он защищается, и длительность, на которой защита поддерживается.

5.3 COGNITIVE

COGNITIVE — это режим, в котором precision распределяется сбалансированным, контекстно-чувствительным образом: prior precision и likelihood precision установлены на уровнях, отслеживающих актуальную надёжность priors и свидетельств в текущей ситуации, а policy precision поддерживается на уровнях, допускающих кооперативную починку, поиск информации и отзывчивое обновление. Это режим, в котором становятся читаемыми экологичные двойники операций, перечисленных в §4: первого-лица-запрос вместо you-атрибуции, названное и обратимое отступление вместо дрейфа абстрактного регистра, соразмерное хеджирование вместо наслоения хеджей, признанное незнание вместо эпистемической неопределённости.

COGNITIVE не следует ошибочно прочитывать как «отсутствие защиты». Агент в режиме COGNITIVE по-прежнему защищает priors — всякая предиктивная система это делает; защита в широком смысле — базовый модус гомеостатического самоподдержания организма. Что отличает COGNITIVE от DEFENCE — это не присутствие или отсутствие prior-защиты, а *калибровка* prior precision к актуальной надёжности prior'a относительно актуальной надёжности доступных свидетельств. В COGNITIVE prior взвешен высоко, когда он хорошо подкреплён, и его вес снижается, когда против него накапливаются свидетельства. В DEFENCE

prior взвешен высоко потому, что он защищён, — потому что механизм, который настроил бы его вес в ответ на свидетельства, ослаблен.

Это различие и есть то, что различитель экологично/компенсаторно статьи I §3.5 отслеживает на precision-уровне. Экологичный защитный ход разворачивается изнутри режима COGNITIVE: он задействует соответствующий биологический вектор на величине, пропорциональной актуальному триггеру, на длительности, которой ситуация действительно требует, и разряжается, когда ситуация разрешается. Компенсаторный защитный ход разворачивается изнутри режима DEFENCE: он задействуется независимо от актуального обоснования, масштабируется по защищённому prior'у, а не по текущей угрозе, и упорствует, потому что защищённый prior не обновляется. Один и тот же биологический вектор — одна и та же обыкновенная человеческая способность оттолкнуть, отступить, умиротворить или замереть — порождает экологичное выражение в COGNITIVE и компенсаторное в DEFENCE. Различается режим, в котором вектор развёрнут.

5.4 ILLUSION

ILLUSION занимает отдельную область prior-доминируемого пространства конфигураций — такую, которая, что существенно, может поддерживаться без острых сигналов угрозы, приводящих в действие DEFENCE. Она характеризуется смещёнными prior means (prior не просто тяжело взвешен; его центральная тенденция сдвинута от evidence-подкреплённого значения) в сочетании с избирательным снижением likelihood precision именно на тех каналах, которые дезамбигуировали бы смещение. Выбор политики в ILLUSION систематически избегает эпистемических действий, которые породили бы дезамбигуирующие свидетельства: говорящий не задаёт вопроса, разрешающего неоднозначность, не приглашает обратной связи, проверяющей нарратив, не ищет перспективы, оспаривающей рамку. Эпистемический вес к (Приложение А) в релевантных подпространствах структурно снижен.

Различие между DEFENCE и ILLUSION — различие между *ригидностью под давлением* и *связным смещением без давления*. DEFENCE нуждается в threat-предсказании для поддержания своей конфигурации; когда угроза разрешается, DEFENCE разряжается, возвращая распределение precision к базовой линии. ILLUSION не нуждается в threat-предсказании; конфигурация самоподкрепляется, извлекая свою устойчивость из внутренней связности самого нарратива, а не из какого-либо продолжающегося внешнего давления. Именно поэтому определённые паттерны речи — удерживающиеся годами, разворачивающиеся в ситуациях, не содержащих острой угрозы, сопротивляющиеся обратной связи не потому, что говорящий чувствует атаку, а потому, что рамка делает обратную связь невидимой, — требуют ILLUSION, а не DEFENCE, чтобы быть схваченными.

Клинически ILLUSION — это то, что психоаналитическая и attachment-традиции описывали под различными ярлыками: стабилизировавшиеся на протяжении жизни организации ложного Я, грандиозность, организующая восприятие, а не провоцируемая уязвением, устойчивые нарративы реляционной истории, в

которых говорящий неизменно протагонист истории, логика которой никогда не возмущается актуальным поведением других. Вычислительный пуант состоит в том, что эти феномены разделяют специфическую precision-сигнатуру: дело не в том, что свидетельства активно отвергаются под давлением; дело в том, что система осела в конфигурацию, в которой свидетельства, оспаривающие рамку, вовсе не входят в обновляющую машинерию генеративной модели. Сигнатура ILLUSION на речевом уровне — соответственно не в первую очередь операции острого опережения (наслоение хеджей, сдвиг абстрактного регистра), но паттерны на более длительной временной шкале: удержание нарратива, избирательный выбор темы и согласованное фреймирование, которые статья IV рассмотрит подробно.

5.5 МЕТА

МЕТА — это режим, в котором precision сам становится объектом вывода. Система не просто обновляет убеждения о состояниях мира при precision-взвешенных свидетельствах; она обновляет убеждения о надёжности собственного взвешивания свидетельств. Формально (Приложение А) это моделируется трактовкой precision-параметров как случайных величин, подчинённых гиперприорам и ошибкам предсказания высшего порядка: когда предсказания системы систематически проваливаются или систематически сверх-успешны, сама статистика ошибок предсказания становится свидетельством того, откалибровано ли текущее распределение precision. Эпистемический вес к в МЕТА повышен именно на тех действиях, которые доставляют информацию о precision: на вопросах, проверяющих собственное фреймирование, на приглашениях к обратной связи, зондирующих собственную модель собеседника.

Феноменология МЕТА — это то, что люди узнают как рефлексивную способность, эпистемическое смирение или метакогнитивную осознанность: способность заметить, что собственная уверенность в prior'e непропорциональна свидетельствам за него, что собственное взвешивание свидетельств само может быть плохо откалибровано, что паттерн разговорных провалов предполагает, что собственная модель собеседника или самой себя нуждается в пересмотре. В речи МЕТА маркируется операциями, которых систематически недостаёт другим режимам: от-первого-лица эпистемическим квалификатором, действительно отслеживающим неопределённость («я не знаю, правильно ли я читаю эту ситуацию» — в отличие от наслоения хеджей режима DEFENCE, которое осушает ассертивную силу, не входя в неопределённость), приглашением к disconfirmation («скажи, если я ошибаюсь»), эксплицитным переформулированием собственной prior-позиции в свете того, что только что сказал собеседник.

МЕТА — это режим, делающий *переходы между режимами* возможными как эксплицитный, политически управляемый процесс, а не как непреднамеренный дрейф. Говорящий, способный рефлексировать над собственными паттернами защитной речи, может изнутри МЕТА настроить распределение precision, поддерживающее DEFENCE или ILLUSION; говорящий, не имеющий доступа к МЕТА, в значительной мере ограничен тем режимом, который занимает в данный момент, с переходами, происходящими — если вообще происходящими — через внешнее давление, а не через внутреннюю модуляцию. Это архитектурный

коррелируют того, что клинические традиции называли способностью к рефлексии или ментализацией; в нашей рамке это режим, активация которого делает терапевтическое изменение механически возможным. Статья IV развернёт траекторные следствия этого свойства.

Предостережение: МЕТА легко ошибочно прочесть гомункулярно. Заманчиво вообразить «высшее Я», осуществляющее рефлексю. Формальное описание (Приложение А) делает эксплицитным, что МЕТА не требует никакого такого агента, — требуется лишь, чтобы precision-параметры трактовались как случайные величины внутри того же иерархического вывода, что и всё остальное. МЕТА непрерывна с нижними режимами, а не онтологически отделена от них. Это не-гомункулярное фреймирование важно не только философски, но и эмпирически: оно означает, что МЕТА, как и другие режимы, имеет наблюдаемые речевые сигнатуры и может быть детектирована тем же классификационным аппаратом, который детектирует остальные.

5.6 Операции как речевые сигнатуры режимов

Отношение между алфавитом §4 и режимами настоящего раздела — это отношение между наблюдаемым выходом и нижележащим динамическим паттерном. Каждая операция имеет характерную precision-сигнатуру: она порождается, когда определённые precision-параметры сконфигурированы определённым образом, и производит, как следствие этой конфигурации, речевой ход со специфическими лингвистическими свойствами. You-атрибуция — не просто семантическая категория; это речевая сигнатура precision-конфигурации, в которой prior precision на самозащитных priors высока, а likelihood precision на состоянии собеседника низка. Сдвиг абстрактного регистра — не просто дрейф темы; это речевая сигнатура политики, выбранной в условиях подавленной эпистемической ценности на аффективно-живом материале. И так для каждой операции §4.

Это и есть то, что санкционирует вывод от наблюдаемой речи к нижележащему режиму. Аннотатор — человек или вычислительный — видящий высказывание, инстанцирующее конкретную операцию, может вывести с уместной уверенностью, что говорящий находится в precision-конфигурации, согласующейся с этой операцией. Аннотатор, видящий устойчивый со-выбор операций из одного семейства на протяжении нескольких высказываний, может вывести с большей уверенностью, что говорящий находится в режиме, характерными выходами которого эти операции являются. Сам режим не непосредственно наблюдаем, но выводим из паттерна операций, которые наблюдаемы. Это та же эпистемическая структура, которая управляет любым аттрактор-основанным выводом в анализе сложных систем: наблюдают траектории, выводят аттрактор.

Методологический выигрыш для движка статьи III специфичен. Движку не требуется непосредственно детектировать precision-конфигурации — задача, требующая инвазивной нейрональной записи и недоступная никакому текстовому классификатору. Движку требуется лишь детектировать операции, которые лингвистически реализованы и ограничены по числу, и отслеживать паттерны их со-встречаемости во времени. Режимы эмерджентны из операций;

операции — это то, с чем движок работает. Утверждение статьи II о том, что операции конечны и теоретически обоснованы, в этом прочтении и есть то самое утверждение, которое санкционирует классификационную архитектуру движка: конечный наблюдаемый базис (операции) даёт эмпирический доступ к конечной латентной структуре (режимам) через отображение, специфицируемое precision-рамкой. Полное описание того, как это отображение реализовано в движке Mindloom, — включая важный терминологический пуант, что четыре precision regime статьи II реализованы в онтологии движка как *слои анализа*, тогда как словарь движка из десяти наблюдаемых поведенческих аттракторов (BUILD, SEEK, UNSEAL, LOCK, SEAL, DRAIN, FLOOD, EDGE, SHIFT, VOID) отсылает к более низкому уровню описания, — даётся в §7.

6. Композиция и контекст

Охарактеризовав операции как наблюдаемый базис (§4) и режимы как прочерчиваемые ими аттракторы (§5), мы теперь обращаемся к вопросу, отложенному и в том и в другом: как операции действительно соединяются в реальной речи, и какую роль играет контекст в том, чтобы эти соединения стали читаемыми? Настоящий раздел развивает пять структурно важных пуантов: как операции компонируются внутри одного высказывания (§6.1), как они разворачиваются последовательно на отрезке дискурса (§6.2), как контекст модулирует, какие операции доступны и как они читаются (§6.3), почему операции читаются прежде всего относительно *якоря* того, на что они отвечают, а не в изоляции (§6.4), и чего текстовая формализация статьи V структурно не видит (§6.5). Методологическое следствие для статьи III специфично: аннотационный протокол, трактующий высказывания как изолированные семантические единицы и игнорирующий композируемость, последовательность и якорение, которые описывает настоящий раздел, упустил бы существенную часть того, что делает компенсаторно-защитную речь узнаваемой.

6.1 Композиция внутри высказывания

Первое наблюдение состоит в том, что операции не взаимно исключаются на уровне отдельного высказывания. Компенсаторно-защитное высказывание может — и часто это делает — инстанцировать более одной операции одновременно, и паттерны этой одновременной инстанциации сами структурированы, а не произвольны. Мы различаем три основных модуса внутривопропозиционной композиции.

Стекирование внутри семейства. Операции из одного векторного семейства со-встречаются с характерной частотой, потому что разделяют одну и ту же нижележащую precision-конфигурацию. Говорящий в DEFENCE-режиме с fight-конфигурацией с высокой вероятностью соединит в одном высказывании уоц-атрибуцию с развёртыванием глобального квантора и экстернализованно-агентной рамкой: *«ты всегда так делаешь — ты пытаешься меня принизить, и теперь ты меня расстроил»*. Три операции технически различны, но исходят из одной precision-сигнатуры и взаимно подкрепляют друг друга; это то, что система производит, когда одна конфигурация реализована через три позиции,

куда операции семейства *fight* попадают в предложение (атрибуция субъекта, темпоральный квантор, грамматическая агентность). Аннотация должна быть способна маркировать все три, не принуждая к выбору между ними.

Стекирование между семействами. Операции из разных векторных семейств тоже могут со-встречаться, и когда они встречаются, соединение обычно не случайно: оно отражает попытку говорящего выполнить более одной защитной функции одновременно. *You*-атрибуция, произведённая с наслоением хеджей («я не знаю, может быть, ну, ты как бы типа пытался меня задеть?»), соединяет *fight*-атрибуцию с *fawn*-самоминимизацией; соединение узнаётся как специфический социолингвистический ход, часто производимый говорящими, чей защищённый *prig* включает сильный запрет на прямую агрессию. Сдвиг абстрактного регистра, одновременно являющийся мета-комментарием («интересно, как подобные разговоры всегда как-то эскалируют»), соединяет *flight*-дистанцию с мета-позиционированием. Эти межсемейственные стекирования — не исключения из алфавита; это *композиции алфавита*, и именно для этого алфавит и существует.

Иерархическое вложение. Третий, более тонкий модус — когда одна операция служит структурным хостом для другой. Рассмотрим: «я просто хотела спросить — и, пожалуйста, скажи, если это глупость, — не избегаешь ли ты меня». Внешняя структура — это *fawn*-наслоение хеджей с избытком извинений; вложенная в неё — это *fight-you*-атрибуция (обвинение в избегании). Операции не просто одновременны; они стоят в отношении хозяин-гость, где *fawn*-оболочка делает *fight*-содержание доставляемым без запуска социальных последствий голого *fight*-выражения. Иерархические вложения диагностичны для определённых профилей говорящих — особенно для тех, чьи защищённые *prigs* включают и «я не должна быть агрессивной», и «эта претензия должна быть сообщена», — и их детекция требует аннотации, способной представлять вложенные, а не только параллельные операциональные маркировки.

Методологическое следствие для статьи III состоит в том, что аннотация на уровне высказывания должна допускать *мульти-маркировку с реляционной структурой*: не просто «это высказывание инстанцирует операции А, В и С», но, где релевантно, «А хостит В, а С параллельна обеим». Движку не требуется восстанавливать полную реляционную структуру в каждом случае; но аннотационная онтология должна быть достаточно богатой, чтобы представлять её там, где структура ясна, — иначе важная информация теряется на стадии ввода.

6.2 Последовательная композиция через высказывания

Второе наблюдение касается того, что происходит через высказывания, а не внутри них. Отрезки компенсаторно-защитной речи редко однородны; у них есть внутренняя структура, и эта структура информативна. Статья IV развернёт формализацию более длинных траекторий и режимных переходов; здесь мы ограничиваемся паттернами ближнего действия, видимыми внутри одного разговорного обмена в несколько реплик.

Три паттерна ближнего действия стоит назвать, потому что они повторяются

через говорящих и контексты и ограничивают то, что аннотатор (или движок) должен ожидать найти.

Цепочки эскалации. Говорящий, начинающий обмен с мягкой you-атрибуции, встречающий сопротивление собеседника и отвечающий более сильным развёртыванием глобального квантора, проходит цепочку эскалации внутри семейства fight. Эскалация не случайна; это попытка системы восстановить защитную эффективность исходной операции, когда ответ собеседника предполагает, что исходная операция не смогла ослабить предсказанный шейм-сигнал. Эскалация внутри семейства распространена; эскалация через семейства менее распространена и обычно указывает, что исходная защитная конфигурация проваливается структурно, побуждая систему попробовать другой защитный механизм.

Последовательности провала починки. Говорящий в режиме COGNITIVE, сталкиваясь с трудной темой, может развернуть экологичное отступление («дай мне подумать об этом»), вернуться в контакт и продолжить вовлечённо. Говорящий в режиме DEFENCE, сталкиваясь с той же темой, может развернуть то, что *выглядит* как экологичное отступление (поверхность сдвига абстрактного регистра похожа на подлинный запрос на время для рефлексии), но затем не возвращается: тема не возобновляется, от-первого-лица-позиция говорящего не восстанавливается, аффективный регистр остаётся приподнятым. Диагностический сигнал — не в первом высказывании, а в провале ожидаемой починки. Аннотатор, маркирующий только первое высказывание, упускает критическую информацию; критическая информация живёт в последовательности.

Устойчивость при вызове. Третий паттерн — отсутствие ожидаемого движения. Говорящий, чьё исходное высказывание инстанцирует, скажем, наложение хеджей, и который продолжает производить хедж-стекированные высказывания через несколько реплик, несмотря на успокоение со стороны собеседника, проявляет режим-уровневую устойчивость, которую отдельное высказывание раскрыть не может. Упорство операционального паттерна через период вызова — вот что отличает режимную фиксацию от контекстного ответа. Это ближний предшественник анализа режимных траекторий, который развернёт статья IV.

Следствие для аннотации состоит в том, что отдельные операции должны маркироваться *внутри их последовательного контекста* — как минимум, с видимыми предыдущей и следующей репликами. Аннотатор, глядящий на одно высказывание, имеет доступ к его операциональной композиции (§6.1), но не к его режим-уровневому значению; только последовательность высказываний начинает раскрывать режим. Движок статьи III отражает это: операциональная классификация работает на одном высказывании, но режимная классификация требует контекстного окна в несколько реплик.

6.3 Контекст как модулятор

Третье наблюдение состоит в том, что операции не контекстно-свободны. Одна и та же грамматическая или прагматическая структура, рассматриваемая в изоляции, может быть компенсаторно-защитной в одном контексте и экологичной в другом — не потому, что операция неоднозначна на уровне

лингвистической формы, а потому, что сам различитель экологично/компенсаторно (§7) контекстно-чувствителен. Мы различаем три типа контекстной информации, модулирующей прочтение операции.

Контекст собеседника. Идентичность, статус и реляционная история собеседника формируют, какие операции соразмерны. Наслоение хеджей, направленное полицейскому при остановке на дороге, соразмерно дифференциалу власти и ставкам; то же наслоение хеджей, направленное супругу в непринуждённом ужине, — нет. Операция не различается по своей грамматической форме; различается соответствие между функцией операции (осушить ассертивную силу, снизить экспозицию) и актуальными требованиями ситуации. Аннотационная система, не кодирующая хотя бы базовую информацию о собеседнике — роль, отношение, дифференциал власти, — не может выполнять различение экологично/компенсаторно аккуратно.

Контекст регистра. Конвенциональные ожидания разговорного регистра задают базовую линию, относительно которой читаются операции. Формальное рабочее письмо, включающее умеренное хеджирование, работает в пределах конвенциональных ожиданий вежливости своего регистра; непринуждённое сообщение близкому другу, включающее то же хеджирование, работает вне конвенциональных ожиданий вежливости своего регистра, и само отклонение информативно. Регистр в этом смысле — локальный нормативный *prig*, относительно которого оценивается развёртывание операции. Движок статьи III использует регистр-маркеры (выбор словаря, синтаксическую формальность, конвенции приветствий и прощаний) как вход для классификатора экологично/компенсаторно именно по этой причине.

Непосредственный разговорный prig. Наиболее проксимальная контекстная информация — это то, что только что сказал собеседник. Операция, развёрнутая в ответ на конфронтационный промпт, читается иначе, чем та же операция, развёрнутая в ответ на тёплый промпт, по причинам, которые мы развиваем полнее в §6.4. Непосредственный *prig* — это, фактически, локальное свидетельство, относительно которого оценивается соразмерность операции. Движок, обрабатывающий высказывания независимо, не глядя на предыдущую реплику собеседника, не может выполнять эту оценку.

Четвёртый тип контекста, который мы откладываем на более поздние статьи, — это *история говорящего*. То же высказывание, произведённое говорящим, разговорная история которого за последний месяц была режим-устойчивой в COGNITIVE, читается иначе, чем то же высказывание, произведённое говорящим, чья история показала устойчивый DEFENCE. Контекст истории говорящего — это то, что статьи IV и V операционализируют как траекторно-уровневую информацию; на операциональном уровне статьи III она непосредственно недоступна, но ограничивает априорное распределение над режимными классификациями, используемое движком.

6.4 Якорь-промпт

Наиболее важный методологический пуант этого раздела — и тот, что более всего отличает наш аннотационный протокол от подходов, трактующих

высказывания как изолированные семантические единицы, — касается того, что мы будем называть *якорем*: специфического промпта или высказывания, на которое целевое высказывание отвечает. Операции читаются прежде всего относительно своего якоря, а не в изоляции. Это не удобство; это структурное свойство феномена.

Рассмотрим хедж-стекированное высказывание: «*я как бы может быть просто хотела узнать, не знаю, не думал ли ты, возможно, об этом*». В изоляции оно узнаётся как операция семейства *fawn* — грамматическая форма однозначно инстанцирует наложение хеджей. Но *функция* операции, и следовательно её экологично-или-компенсаторно статус, определяется тем, на что она отвечает. Как ответ на «*ты выйдешь за меня замуж?*» хедж-стек делает серьёзную работу: говорящая пытается выразить неопределённость относительно жизненно-меняющего обязательства способом, минимизирующим вред собеседнику. Как ответ на «*хочешь кофе?*» тот же хедж-стек непропорционален, и его непропорциональность диагностична для режимной конфигурации, в которой даже тривиальные взаимодействия активируют политики минимизации экспозиции. Грамматическая форма идентична. Операция идентична. Интерпретация — нет.

Это и есть причина, по которой операции должны аннотироваться *с их промптами*, а не в одиночку. Аннотационный корпус, состоящий из изолированных высказываний, как бы тщательно ни маркированных на операциональном уровне, теряет информацию, различающую экологичное и компенсаторное выражения. Корпус, парующий каждое высказывание с его непосредственным предшественником — и, идеально, с достаточным предшествующим контекстом, чтобы дезамбигуировать регистр и отношение с собеседником, — сохраняет эту информацию и делает экологично/компенсаторно классификацию вычислимой.

Импликация для дизайна движка статьи III прямая. Формат ввода — не «высказывание плюс маркировка»; это «промпт плюс высказывание плюс маркировка», причём промпт включает, как минимум, непосредственно предыдущую реплику собеседника и, где доступно, резюме разговорной позиции, установленной в предыдущих репликах. Операциональный классификатор движка работает на высказывании; экологично/компенсаторно классификатор движка работает на *паре*. Статья III детализирует, как эти два классификатора соединены внутри слоистой архитектуры движка.

Практическое замечание. В некоторых исходных материалах — транскрибированных терапевтических сессиях, например, — якорь непосредственно доступен, потому что структура источника эксплицитно диалогична. В других — мемуары, от-первого-лица рефлексивная письменность, монологический дискурс — якорь имплицитен и должен быть реконструирован из собственного фреймирования говорящего. Реконструкция вносит дополнительную неопределённость, и уверенность движка в его экологично/компенсаторно классификации должна быть соответственно ниже на монологическом вводе, чем на диалогическом. Это одна из осей, по которым движок сообщает калиброванную неопределённость пользователю.

6.5 Просодия и то, что упускает текстовая формализация

Финальный пуант — это честно заявленное ограничение. Наша формализация оперирует на тексте, и эмпирическая работа статьи V ведётся на текстовом корпусе. Это принципиальный выбор — текст есть наблюдаемое с самой широкой доступностью, самой прозрачной воспроизводимостью и наименьшим числом смещений от транскрипционных решений, — но это выбор, систематически опускающий значимый канал сигнала защитной речи: просодию.

Устная защитная речь несёт просодические сигнатуры, которые текст не представляет. Операции семейства *fight*, в устной форме, часто показывают повышенную громкость, сжатый диапазон высоты тона и ускоренный темп; *fawn*-хеджи показывают характерную восходящую интонацию в концах утверждений (тот «*uptalk*», что превращает декларативы в квази-вопросы), удлинённые смягчающие частицы и сниженную громкость на асертивным содержании; *flight*-сдвиги абстрактного регистра часто сопровождаются заметным уплощением аффективной просодии, когда говорящий дистанцируется не только семантически, но и фонетически от темы; *freeze*-эпистемическая неопределённость показывает характерное просодическое уплощение, сниженную вариабельность высоты тона и удлинённое паузирование. Эти просодические маркеры реальны, они детектируемы, и иногда они — один из наиболее надёжных каналов для различения экологичного и компенсаторного выражений: хедж, произнесённый с тёплой, беглой просодией, читается иначе, чем синтаксически идентичный хедж, произнесённый с плоской, сжатой просодией.

Текстовая формализация структурно не может иметь доступа к этим маркерам. Следствие двойко. Во-первых, наша формализация есть *подмножество* того, что содержала бы полная формализация компенсаторно-защитной речи; будущее расширение на мультимодальные данные — аудио, видео, физиологические — обогатило бы алфавит и улучшило бы различение. Во-вторых, точность нашей формализации на тексте должна пониматься как *нижняя граница* того, чему феномен в принципе доступен: где текстовая классификация неуверенна, просодическая информация часто разрешила бы неопределённость, и долгосрочная цель поля должна включать разработку мультимодальных аннотационных протоколов, интегрирующих то и другое.

Для текущих целей мы отмечаем ограничение и продолжаем. Текст достаточен для методологического утверждения, которое делает статья II, — что компенсаторно-защитная речь допускает формализацию на конечном базисе, — потому что аргумент покоится на операциях, которые лингвистически реализованы и детектируемы в тексте. Он также достаточен для эмпирического утверждения, которое сделает статья V и которое касается распределения операций и режимов через двуязычный текстовый корпус, а не полного мультимодального феномена. Мультимодальное расширение обозначено как будущая работа, а не как актуальный пробел; пробел, который она закрывает, — в объёме, а не в принципе.

7. Различитель экологично/компенсаторно на уровне операций

В §3.5 статьи I были введены три критерия для различения экологичного и компенсаторного выражений биологических защитных векторов: *пропорциональность* (совпадает ли ответ с магнитудой актуального триггера), *обратимость* (разряжается ли ответ, когда ситуация разрешается), и *сохранение контакта* (оставляет ли ответ интерактивный канал неповреждённым). Эти критерии были сформулированы на уровне общего защитного ответа — вектора в целом, выраженного через взаимодействие. Чтобы движок статьи III операционализировал различение, нам нужно, чтобы критерии работали на уровне отдельных операций, где они могут применяться к конкретным речевым актам, а не к размытым поведенческим паттернам. Настоящий раздел показывает, как каждый критерий переводится на операциональный уровень, и показывает, что на этом уровне каждый критерий есть не бинарная, а *непрерывная, контекстно-чувствительная переменная*, которую надо оценивать относительно контекстной информации, специфицированной в §6. Три критерия не независимы; они коварируют характерным образом, вместе составляя композитный различитель, который развёртывает движок.

7.1 Пропорциональность

Пропорциональность на операциональном уровне — это степень соответствия между двумя магнитудами: *магнитудой операции* и *магнитудой триггера*, на который она отвечает. Ни одна магнитуда не является отдельным числом; обе многомерны, и соответствие между ними оценивается совместно, а не покомпонентно.

Магнитуда операции имеет как минимум три компонента. Первый — *интенсивность*: наслаение хеджей с шестью смягчителями интенсивнее, чем с двумя, уоу-атрибуция с глобальным квантором и мотив-атрибуцией интенсивнее, чем голая уоу-атрибуция. Второй — *объём*: операция, фреймирующая всё взаимодействие (глобальный квантор, экстернализованно-агентная рамка через несколько реплик), имеет больший объём, чем операция, ограниченная одним утверждением. Третий — *ангажемент*: операция, доставленная с высокой прагматической силой — плоским ассертивным регистром, от-первого-лица обязательством к защитному фрейму — тяжелее ангажирована, чем доставленная осторожно.

Магнитуда триггера имеет, соответственно, несколько компонентов. Первый — *ставки*: касается ли триггер жизненно-меняющего обязательства, значимого реляционного вопроса, тривиального логистического дела? Второй — *новизна*: вводит ли триггер содержание, с которым говорящий никогда не сталкивался, содержание, оспаривающее хорошо установленную self-модель, или содержание рутинное? Третий — *дифференциал власти*: приходит ли триггер от вышестоящего, ровни, или от кого-то с меньшим социальным авторитетом? Четвёртый — *конфронтационное намерение*: сформулирован ли триггер как атака, нейтральный запрос или тёплое обращение?

Пропорциональность — это соответствие между этими многомерными магнитудами. Хедж-стекированное высказывание в ответ на «ты выйдешь за меня замуж?»

— это высокомагнитудная операция в ответ на высокомагнитудный триггер; магнитуды совпадают, и совпадение — это то, что делает операцию экологичной, а не компенсаторной, несмотря на интринсическую силу операции. Хедж-стекированное высказывание в ответ на «хочешь кофе?» — это высокомагнитудная операция в ответ на низкомагнитудный триггер; магнитуды не совпадают, и именно *несоответствие* делает операцию компенсаторной. Пропорциональность — не о силе операции как таковой; она о совместном распределении силы операции и силы триггера.

Именно поэтому пропорциональность не сводится к свойству высказывания в одиночку, как уже аргументировал §6.4. Хедж-стек — одно и то же высказывание в обоих случаях; его пропорциональность различается, потому что различается триггер. Движок статьи III оценивает пропорциональность, оценивая обе магнитуды — из операциональных лингвистических признаков для операции, из якорь-промпт признаков для триггера — и вычисляя соответствие как совместную меру. Выход — не бинарная маркировка пропорционально/непропорционально, а калиброванный балл, отражающий степень соответствия, причём движок сообщает свою уверенность в каждой компонентной оценке отдельно.

Оговорка. Оценка триггер-магнитуды движком уязвима именно там, где это важнее всего: в неоднозначных межличностных обменах, где говорящий и собеседник могут не соглашаться относительно ставок, новизны или конфронтационного намерения самого триггера. Говорящий в режиме DEFENCE может переживать нейтральный вопрос как высокоставочный вызов; движок, глядя на текст, видит нейтральный вопрос. Эта асимметрия — между взглядом движка на триггер и взглядом говорящего на триггер — сама является диагностической информацией. Когда операциональная магнитуда говорящего устойчиво высока относительно оценки триггер-магнитуды движком, это сигнатура DEFENCE-конфигурации говорящего, а не провал оценки движка. Движок сообщает рассогласование и позволяет нижестоящему анализу его интерпретировать.

7.2 Обратимость

Обратимость на операциональном уровне — это темпоральное свойство: она измеряет, разряжается ли операция (и насколько быстро), когда условия, её оправдывавшие, перестают удерживаться. В отличие от пропорциональности, которая в принципе может оцениваться на одном высказывании с его якорем, обратимость интринсически является свойством *последовательностей* — она требует наблюдения того, что происходит, когда собеседник сигнализирует, что защитная операция больше не нужна.

Рассмотрим двух говорящих, каждый из которых отвечает на воспринятое уязвление уоу-атрибуцией. Собеседник в обоих случаях признаёт уязвление и приносит извинения. Экологичный говорящий, чья операция была пропорциональна триггеру и теперь больше не нужна, разряжает уоу-атрибуцию: следующая реплика больше не фреймируется через атрибутивную враждебность, реляционный канал починен, защитная позиция отпущена. Компенсаторный говорящий, чья операция на самом глубоком уровне была не ответом на это уязвление, а

развёртыванием хронически-сконфигурированной политики семейства *fight*, не разряжается: следующая реплика продолжает атрибутировать враждебную внутренность, вводит новые претензии, эскалирует. Признание собеседника — то, что для экологичного говорящего завершило бы защитный эпизод, — ничего не меняет в операциональном выборе компенсаторного говорящего, потому что операция в действительности не выбиралась в ответ на признание-чувствительную ситуацию с самого начала.

Эта темпоральная сигнатура — то, что схватывает обратимость, и поэтому обратимость не может оцениваться на одном высказывании. Движку нужна последовательность — как минимум, целевая операция, ответ собеседника, и следующая реплика говорящего — чтобы классифицировать обратимость. Более длинные последовательности дают более надёжную классификацию, потому что обратимость — градуированное свойство: частичная разрядка, запаздывающая разрядка, и разрядка-с-последующим-возвратом — всё это различные паттерны, которые достаточное окно может различить, а короткое — нет.

Три компонента структурируют измерение обратимости. Первый — *детекция сигнала безопасности*: действительно ли собеседник предоставил сигнал, что защитная операция больше не оправдана, — признание, извинение, деэскалацию, выражение заботы? Без сигнала безопасности вопрос об обратимости не возникает; операция, упорствующая в отсутствие сигнала безопасности, не является тем самым необратимой — она просто неразрешённой. Второй — *латентность ответа*: если сигнал безопасности был предоставлен, сколько реплик проходит до разрядки операции? Экологичная операция *fight* обычно разряжается в пределах одной-двух реплик после сигнала безопасности; компенсаторная операция *fight* может упорствовать неопределённо долго. Третий — *полнота разрядки*: когда разрядка происходит, полная ли она (операция больше не появляется в последующих репликах), частичная (операция ослабевает, но упорствует), или только поверхностно-уровневая (грамматическая форма операции отпадает, но нижележащая *precision*-конфигурация упорствует, проявляясь в других операциях из того же семейства)?

Операционализация для статьи III требует, чтобы движок отслеживал упорство операции через контекстное окно обычно 5–8 реплик, с автоматической детекцией сигналов безопасности в репликах собеседника и вычислением паттернов разрядки на репликах говорящего. Выход — калиброванный балл обратимости, отражающий степень, в которой операция была разряжена в ответ на сигнал безопасности, с уместной неопределённостью, когда сигнал безопасности неоднозначен или отсутствует.

Примечательный случай. Когда обратимость высока, а пропорциональность низка — непропорционально-сильная операция, тем не менее легко разряжающаяся, как только появляется сигнал безопасности, — профиль предполагает экологично-смежное развёртывание: говорящий превысил магнитуду, но был подлинно отзывчив к починке. Когда обратимость низка, а пропорциональность высока — пропорциональная операция, тем не менее упорствующая после сигнала безопасности, — профиль предполагает регимную конфигурацию, в которой даже оправданные защитные ответы удерживаются за пределами своего функционального объёма. Композит двух критериев информативен способом,

которым ни один из них не информативен в одиночку, — пуант, который мы развиваем в §7.4.

7.3 Сохранение контакта

Сохранение контакта на операциональном уровне измеряет, оставляет ли операция, выполняя свою защитную функцию, интерактивный канал неповреждённым, чтобы собеседник мог ответить внутри него. Это критерий, наиболее остро различающий названное отступление и молчаливое изъятие, пропорциональное встречное давление и атаку, повреждающую канал, признанное незнание и диссоциативную эвакуацию.

Две операции с идентичными защитными функциями могут драматически различаться по своей сигнатуре сохранения контакта. Сравним экологичную операцию семейства *flight* — «*мне нужно отступить от этого разговора; я вернусь к нему завтра*» — с компенсаторной операцией семейства *flight* — сдвигом абстрактного регистра, поднимающим говорящего из темы без называния хода. Обе выполняют отступление. Первая называет отступление, атрибутирует его говорящему («мне нужно») и эксплицитно сохраняет возможность возвращения («завтра»). Вторая не делает ничего из этого: отступление исполняется без признания, агентность говорящего не выведена на передний план, условия возвращения не специфицированы. Собеседник, столкнувшийся с первой, знает, что произошло, и может планировать соответственно; собеседник, столкнувшийся со второй, получает разрыв без информации о его источнике или длительности.

Измерение сохранения контакта имеет четыре главных компонента, каждый из которых — непрерывная лингвистическая переменная.

Ориентация на собеседника: остаётся ли операция ориентированной на собеседника как присутствующего, отзывчивого агента, или извлекает собеседника из интеракционной рамки? You-атрибуция, несмотря на то что это операция семейства *fight*, по крайней мере обращается к собеседнику; сдвиг абстрактного регистра, перемещающийся к «таким людям, как мы» или «можно сказать», не обращается ни к кому конкретно. Местоимение второго лица, прямой вопрос, признание предыдущей реплики собеседника — всё это маркеры ориентации на собеседника, чьё присутствие поддерживает сохранение контакта и чьё отсутствие его ослабляет.

Приглашение к ответу: оставляет ли операция место для того, чтобы собеседник взял следующую реплику осмысленным образом, или структурно исключает ответ? Операция, оканчивающаяся абсолютным глобальным квантором («ты всегда так делаешь»), исключает ответ, отрицая возможность контекстной починки; операция, оканчивающаяся конкретным утверждением в настоящем времени, приглашает ответ, потому что это утверждение в принципе проверяемо относительно текущего взаимодействия. Присутствие или отсутствие приглашения к ответу — частично синтаксическое (вопросительные маркеры, завершающая специфичность) и частично прагматическое (вовлечённость с тем, что сказал собеседник, vs. монологическая экспликация).

Присутствие от-первого-лица: остаётся ли собственная позиция говорящего от первого лица доступной в операции, или она эвакуирована? Экологичная операция *fight* («я чувствую, что ты пренебрегаешь моей тревогой») сохраняет от-первого-лица-контур говорящего; компенсаторная операция *fight* с экстернализованно-агентной рамкой («ты заставил меня почувствовать пренебрежение») эвакуирует от-первого-лица-агентность. Когда присутствие от первого лица эвакуировано, собеседнику не к кому *обратиться* с ответом; говорящий фактически удалил координаты, по которым ответ мог бы быть направлен.

Восприимчивость к починке: оставляет ли прагматическая форма операции говорящего доступным для получения и интеграции репарационного хода от собеседника? Этот компонент пересекается с обратимостью, но не идентичен ей: обратимость — *post-hoc* мера того, была ли починка получена; восприимчивость к починке — одновременная мера того, могла ли починка в принципе быть получена в момент операции. Операция эпистемической неопределённости, отказывающаяся от обязательств к любому пропозициональному содержанию, оставляет говорящего недоступным для починки, потому что нет пропозиции, к которой собеседник мог бы обратиться.

Движок операционализирует сохранение контакта как взвешенный композит этих четырёх компонентов, каждый из которых может быть оценён из операциональных и дискурс-уровневых лингвистических признаков. Как и с другими критериями, выход — калиброванный балл, а не бинарное суждение, и движок сообщает свою уверенность отдельно для каждого компонента. Композитный балл — не простое среднее; некоторые компоненты более диагностичны в некоторых контекстах, чем в других (ориентация на собеседника весит тяжелее в диадических, чем в многосторонних обменах; присутствие от первого лица весит тяжелее в аффективных, чем в информационных обменах), и веса движка соответственно контекстно-чувствительны.

7.4 Ковариация и композитный различитель

Три критерия не независимы. На практике они коварируют характерными паттернами, которые сами несут диагностическую информацию, — паттернами, предсказываемыми *precision*-режим-архитектурой §5 и используемыми движком для построения композитного различителя, более чувствительного, чем любой отдельный критерий в одиночку.

Регим *DEFENCE* характерно выражает все три критерия в компенсаторном направлении: непропорциональная магнитуда операции относительно магнитуды триггера (потому что операция движима защищённым *prior*'ом, а не актуальным триггером), низкая обратимость (потому что защищённый *prior* не обновляется в ответ на сигналы безопасности, предоставленные собеседником), и низкое сохранение контакта (потому что политики, кэшированные под защищённым *prior*'ом, были отобраны для внутреннего ослабления предсказанного стыда, а не для сохранения интеракционного канала). Регим *COGNITIVE* характерно выражает противоположный паттерн: пропорциональность, обратимость и сохранение контакта вместе, с каждым критерием на уровнях, уместных к текущему контексту. Это два полюса, которые композитный различитель

спроектирован различать, и когда все три критерия выравниваются, классификация прямолинейна.

Более интересные случаи — гибриды, где три критерия расходятся. Говорящий может произвести операцию, пропорциональную триггеру, но с низкой обратимостью: исходная магнитура совпадает, но однажды развёрнутая операция не разряжается, даже когда собеседник сигнализирует, что разрядка была бы уместна. Говорящий может произвести операцию с высоким сохранением контакта, но низкой пропорциональностью: операция непропорционально сильна для ситуации, но доставлена способом, сохраняющим канал для ответа. Говорящий может произвести операцию с высокой пропорциональностью и высокой обратимостью, но низким сохранением контакта: операция уместна по магнитуде и разряжается, когда ситуация разрешается, но пока она активна, она повреждает канал.

Каждый из этих гибридных профилей соответствует отдельной конфигурации говорящего, которую композитный различитель в принципе может идентифицировать. Пропорциональные-но-необратимые операции часто диагностичны для ILLUSION-режим-динамики (§5.4): начальное развёртывание разумно отслеживает триггер, но операция, однажды развёрнутая, удерживается машинерией поддержания связности ILLUSION-конфигурации даже после того, как внешнее давление разрешается. Непропорциональные-но-сохраняющие-контакт операции часто указывают на говорящего, чей режим преимущественно COGNITIVE, но который был временно дестабилизирован высокосалиентным триггером — композитный профиль экологично-тяготеющий, несмотря на провал пропорциональности. Пропорциональные-и-обратимые-но-повреждающие-канал операции часто указывают на говорящего, чей режим преимущественно COGNITIVE, но чей репертуар экологичных операций ограничен в специфическом домене (например, говорящий, справляющийся с интеллектуальным разногласием когнитивно, но не располагающий экологичными операциями для эмоционального конфликта, и поэтому производящий повреждающие канал операции, когда эмоциональный конфликт возникает). Классификация композитного различителя более нюансирована, чем позволяет любой отдельный критерий.

Композитный различитель движка, соответственно, не является взвешенной суммой по трём критериям, а совместным классификатором над трёхмерным пространством, которое они натягивают. Выход классификатора — распределение над небольшим числом типов профилей — экологичный, компенсаторный-DEFENCE, компенсаторный-ILLUSION и несколько гибридных профилей — с калиброванной неопределённостью, отражающей, насколько ясно наблюдаемая операция-в-контексте попадает в каждый профиль. Это то, что движок статьи III доставляет как экологично/компенсаторно слой своего анализа, и это слой, наиболее прямо отличающий рамку статьи II от подходов, трактующих экологично/компенсаторно различение как бинарное или основывающих его на операциональных признаках в одиночку.

7.5 Что различитель доставляет и чего он не делает

Два уточнения закрывают раздел, оба защищают от перечитывания того, что композитный различитель может делать.

Первое: различитель оперирует на уровне операции-в-контексте, а не на уровне говорящего. Отдельная операция, классифицированная как компенсаторная, не санкционирует утверждение о режиме говорящего; она санкционирует, самое большее, утверждение о precision-конфигурации, активной в этот момент речи. Классификация режима требует агрегации через многие операции, способом, который §6.2 описал для последовательной композиции и §5.6 описал для вывода от операции к режиму. Говорящий может производить компенсаторные операции в одних контекстах и экологичные в других; различитель сообщает, что он видит, операция за операцией, и режим-уровневое утверждение есть нижестоящий вывод из накапливающегося паттерна. Статья III специфицирует механизм агрегации; статья II — нет.

Второе: различитель не является оценочным инструментом. Классификация операции как компенсаторной — не суждение, что операция неверна, что говорящий защищен в морально-нагруженном смысле, или что говорящий должен развернуть другую операцию. Компенсаторные операции — обыкновенное человеческое лингвистическое поведение; всякий говорящий их производит; они производятся по причинам, которые в пределах precision-режим-архитектуры механически осмысленны — они ослабляют предсказанную ошибку стыда, что и есть то, что система сконфигурирована делать. Различитель идентифицирует свойство развёртывания операции относительно её контекста; он не предписывает. Предписывающая работа — терапевтическая, педагогическая или саморефлексивная — это работа, которая может быть построена *поверх* классификации различителя, но это не работа, которую выполняет сам различитель, и статья II не занимает позиции относительно того, как такая предписывающая работа должна выглядеть.

Эти два уточнения вместе устанавливают объём экологично/компенсаторно-машинерии, введённой в настоящем разделе. Машинерия — это классификационный инструмент, оперирующий на специфическом уровне описания, производящий калиброванные выходы над ограниченным набором типов профилей, для нижестоящего использования в анализах, которые остаются специфицируемыми в других местах серии. Она не является и не заявляется как диагностический инструмент, оценочная метрика или предписывающий гид.

8. Объём, три уровня описания и мост к статье III

8.1 Чего эта статья не утверждает

Шесть ограничений объёма вытекают из того, что было аргументировано, и мы формулируем каждое явно, потому что формализацию легче прочесть неправильно, чем правильно.

Первое. Мы не утверждаем, что речь вообще комбинаторна в том смысле, который развит здесь. Утверждение ограничено специфическим подклассом

речи, охарактеризованным в статье I как компенсаторно-защитный — речью, производимой в предиктивно-стыдовой динамике на защищённых priors. Речь, производимая вне этой динамики — подлинное вопрошание, кооперативное решение задач, любовный обмен, творческое порождение, — лежит вне этой рамки и требует других описаний. Комбинаторная архитектура, которую мы специфицировали, — это утверждение о подсистеме, а не о языке.

Второе. Мы не утверждаем, что список операций в §4 исчерпывающ. Список предварителен и будет уточнён через операционализацию в статье III. Некоторые операции, сгруппированные нами как варианты одного правила, при детальном анализе могут оказаться различными; некоторые, перечисленные как отдельные, — поверхностными вариантами одного нижележащего правила. Алфавит — не последнее слово; это первый черновик, от которого ведётся эмпирическая работа.

Третье. Мы не утверждаем, что алфавит языково-универсален. Эмпирическая работа в статье V тестирует алфавит на билингвальном англо-русском корпусе, и мы ожидаем кросс-лингвистических вариаций в том, как операции реализуются (специфические лингвистические маркеры, грамматические паттерны, просодические сигнатуры). Мы утверждаем инвариантность на уровне тождества операции: операция, которую русскоговорящие реализуют через одну морфосинтаксическую конструкцию, а англоговорящие — через другую, — это одна и та же операция в смысле, релевантном нашему анализу. Это утверждение проверяемо и будет проверено.

Четвёртое. Мы не утверждаем, что идентификация операций достаточна для клинической оценки. Клиническое суждение остаётся необходимым для той работы, которую оно делает: интегрировать анамнез, реляционный контекст, перенос, план лечения. Роль формализации — поддерживать клиническое суждение, делать видимыми для клиницистов операциональные паттерны, которые они, возможно, уже чувствуют, но не имеют общего языка, и позволять масштабное исследование, которого клиническое наблюдение в одиночку не поддерживает. Её роль — не заменять клиническое суждение.

Пятое. Мы не утверждаем, что движок статьи III имплементирует полный алфавит. Статья III начинается с ограниченного подмножества и масштабируется; части алфавита, артикулированные в настоящей статье, не будут операциональны в первом релизе статьи III, а некоторые, возможно, не операционализуемы вовсе в рамках возможностей текущего поколения классификационных архитектур. Там, где алфавит специфицирует больше, чем движок имплементирует, алфавит — это теоретическое обязательство, а движок — эмпирическое приближение.

Шестое. Мы не утверждаем, что вся precision-динамика дискурса укладывается в четыре режима. Четырёхрежимная архитектура §5 применима специфически к компенсаторно-защитной динамике, охарактеризованной в статье I. Дискурс вне этой динамики — большая часть обыкновенной речи — пересекает гораздо большее пространство конфигураций precision; четыре режима — не общая таксономия дискурса, а специфическое утверждение об аттракторах, возникающих под предиктивно-стыдовым давлением.

Седьмое обязательство, методологическое, а не определяющее объём, согласуется

с тем, что статья I §2.3 объявила для серии в целом. Каждая статья этой серии содержит явное разграничение того, что утверждается и что не утверждается, — а там, где предъявляются имплементации, честный анализ того, где эти имплементации не дотягивают до заявленной области применимости, рассматривается как часть вклада. В статье II это принимает форму шести перечисленных выше ограничений объёма. В статье III это примет форму откалиброванного раздела ограничений, сверенного с верифицированным по коду аудитом. В статье V это примет форму подробного обсуждения режимов отказа по сравнению с gold-standard разметкой. Читатель, ориентированный на честную отчётность о пределах применимости как на методологический стандарт, узнает это согласованное отношение к ограничениям во всей серии.

8.2 Три уровня описания

Архитектура, которую мы развили, встраивается в большую серию, а эта большая серия использует три терминологические системы, которые легко расслышать как три конкурирующие таксономии. Прежде чем перейти к мосту со статьёй III, мы делаем отношение между ними эксплицитным.

Три системы таковы: биологические **векторы**, введённые в статью I; **precision regimes**, развитые в §5 настоящей статьи; и **режимы уровня движка** — которые статья III называет аттракторами, — именованные в статье III. Каждая адресует свой вопрос, и читатель, ожидающий, что они схлопнутся в единую иерархию, будет справедливо разочарован — потому что они не схлопываются таким образом.

Векторы (статья I) описывают *направление мотивации*. Что организм пытается сделать: атаковать непосредственную угрозу, отступить от неё, умиротворить её, обездвижить перед ней или ориентироваться на неизвестное? Это филогенетически древние поведенческо-аффективные политики с хорошо задокументированными нейронными субстратами. Статья I развивает пять векторов: fight, flight, fawn, freeze — и, добавленный в §3.5 статьи I как незащитная ориентация, SEEK. Векторы до-языковые; это то, что организм наследует из эволюционной истории, и это сохраняется в речи как направленная логика.

Precision regimes (§5 настоящей статьи) описывают *конфигурацию выводной машины*. Какие priors жёстки, какие каналы свидетельств ослаблены, сколько веса помещается на самозащитных priors относительно опровергающего входа, доступен ли вывод более высокого порядка над самой precision. Это архитектурные состояния иерархической генеративной модели, а не политики относительно того, что делать. Четыре режима — DEFENCE, COGNITIVE, ILLUSION, META — это аттракторные конфигурации этого precision-allocation пространства.

Режимы уровня движка (статья III) описывают *наблюдаемые речевые паттерны*, возникающие из совместной работы первых двух. Десять режимов уровня движка, перечисленных в статье III (BUILD, SEEK, UNSEAL, LOCK, SEAL, DRAIN, FLOOD, EDGE, SHIFT, VOID), — это эмпирически обоснованные кластеры в пространстве реального дискурса: паттерны, повторяющиеся через говорящих и контексты

достаточно устойчиво, чтобы классификационный движок мог научиться их распознавать.

Ключевой концептуальный момент — и то, что разрешает кажущееся напряжение «почему пять, почему четыре, почему десять», — состоит в том, что первые две системы *ортогональны*, а не иерархичны. Векторы и precision regimes описывают разные вещи. Данный вектор может выражаться в любом precision regime. Fight в COGNITIVE — это утверждение соразмерной границы со вниманием к обратной связи и готовностью к репарации. Fight в DEFENCE — это обвинительная уоу-атрибуция, не разряжающаяся, когда собеседник признаёт. Fight в ILLUSION — это устойчивое ощущение, что другой враждебен, даже в отсутствие наличного триггера. Это один и тот же вектор — направленная вовне мобилизованная сила, — оперирующий в трёх разных конфигурациях выводной машины. Конфигурации — не свойства вектора; это свойства системы, развёртывающей вектор.

Та же ортогональность применима к остальным четырём векторам, — поэтому продукт-пространство вектор × режим имеет больше ячеек, чем каждая из систем в отдельности имела бы строк. Не всякая ячейка одинаково заселена в реальном дискурсе — одни комбинации вектор-режим распространены, другие редки, некоторые, возможно, пусты, — но архитектурное пространство определяется декартовым произведением, а не внутренними различиями какой-либо из систем в отдельности.

Режимы уровня движка, таким образом, живут в этом продукт-пространстве. Они не избыточны относительно каждой из систем; они — эмпирически валидированные кластеры того, какие именно ячейки продукт-пространства вектор × precision regime реально заселяют реальную речь. Режим уровня движка LOCK — это приблизительно «операции fawn в конфигурации DEFENCE с высоким ангажементом и низкой обратимостью» — специфическая ячейка продукт-пространства, возникающая достаточно часто, чтобы заслуживать собственной метки. Режим BUILD — приблизительно «операции fight и SEEK в конфигурации COGNITIVE с сохранённой META-доступностью» — другая ячейка того же пространства. Десять режимов уровня движка не выводятся ни из векторов в одиночку, ни из precision regimes в одиночку; они выводятся из распределения реально встречающихся комбинаций.

У этой архитектуры есть следствие, важное для того, как читается серия. Читатель, спрашивающий «но какой это вектор, LOCK?» или «какой это precision regime, SEEK?», задаёт вопрос, на который архитектура не может ответить в изоляции: LOCK — это вектор-режимная комбинация, а SEEK уровня движка (к которому мы переходим в §8.3) — это кластер, со-встречающийся преимущественно с precision-allocation COGNITIVE и активацией биологического SEEK-вектора, но кластер — это паттерн со-встречаемости, а не вектор или precision regime в изоляции. Три системы картографируют единую реальность под тремя углами; они не складываются в стопку.

8.3 Замечание о слове «SEEK»

Слово *SEEK* появляется в этой серии на двух разных уровнях описания, и поскольку одно и то же слово используется на обоих, читатель, не замечающий различия, их смешивает. Мы именуем это различие здесь.

Биологический SEEK — это аппетитивно-мотивационная система, введённая в §3.5 статьи I, — система SEEKING Panksepp'a (1998), нейробиологически инстанцированная в мезолимбическом дофаминергическом пути (вентральная тегментальная область, прилежащее ядро, латеральный гипоталамус и их корковые проекции), функционально организующая приближение-из-аппетита, исследовательское поведение и предвосхищающую вовлечённость. Её формальный двойник в рамке предиктивной обработки — это действие ради эпистемической ценности, действие, отбираемое для сокращения неопределённости относительно состояний мира, а не для обеспечения предпочтительных исходов (Friston et al., 2015). Биологический SEEK — это вектор в развитом в §8.2 смысле: направление мотивации, класс политик, которые организм может развернуть.

SEEK уровня движка — это один из десяти режимов уровня движка, перечисленных в статье III. Он обозначает специфический речевой паттерн: подлинные исследовательские ходы — открытые вопросы, запросы уточнения, информационно-поисковые ходы — без сопутствующего давления на действие или обязательство. Приписывание движком SEEK означает, что текущая конфигурация дискурса говорящего преимущественно соответствует этому паттерну; само по себе это ничего не говорит о нижележащей нейрональной системе SEEKING говорящего или о *precision regime*, в котором его генеративная модель в данный момент сконфигурирована.

Эти два понятия связаны, но различны. Режим движка SEEK — это приблизительно характерная речевая сигнатура, которую биологический SEEK производит при экологичном выражении — в *precision-allocation COGNITIVE*, без ко-активации защитных векторов, с сохранённой доступностью вывода более высокого порядка. Потому они и разделяют имя: режим движка — это наблюдаемое лицо вектора в его здоровом развёртывании. Но режим движка — это не вектор. Говорящий с дисрегулированной биологической системой SEEKING (подавленной в дорсально-вагальном коллапсе или хронически повышенной при аддикции) тем не менее может производить речь, которая в поверхностной форме читается как SEEK уровня движка. Наоборот, активация биологического SEEK-вектора может выражаться, в компенсаторных условиях, в речевых паттернах, которые движком классифицировал бы иначе — как руминацию, как вопрошание-избегание, как компульсивный поиск смысла. Отображение из вектора в режим уровня движка — один-ко-многим при экологичном выражении и становится сложнее в компенсаторных условиях.

Когда настоящая статья говорит SEEK без уточнения, в §§1–7, она имеет в виду биологический вектор, введённый в статье I. Когда статья III говорит SEEK как об одном из десяти своих режимов, она имеет в виду наблюдаемый речевой паттерн. Читатель, проходящий через серию, должен удерживать оба смысла и позволять контексту разрешать референцию, как и с любым другим термином,

оперирующим на нескольких уровнях описания в зрелой научной литературе.

8.4 Структурное замечание о META

Короткий флаг о META, четвёртом precision regime из §5. Настоящая статья трактовала META как один из четырёх режимов на параллельном основании — DEFENCE, COGNITIVE, ILLUSION, META, — потому что такая подача облегчала развёртывание четырёхсторонней оппозиции и потому что она эхом отражает четырёхвекторную организацию, характеризующую остальную архитектуру. Но есть альтернативное прочтение, которое мы хотим признать, потому что у него есть реальная сила, и потому что читатель, пришедший к нему самостоятельно, может резонно спросить, почему мы этого не сделали.

В альтернативном прочтении META — не четвертая конфигурация precision-allocation в том же смысле, что и первые три; это второпорядковая операция, способная модифицировать любую из них. DEFENCE + META — это говорящий в защитной конфигурации, тем не менее способный наблюдать факт своей защитности. COGNITIVE + META — это обыкновенная рефлексивная мысль. ILLUSION + META — это феноменологически поразительное состояние распознавания собственной иллюзии изнутри самой иллюзии. При таком прочтении META лучше характеризуется как вывод более высокого порядка над precision, как оператор, а не состояние, и архитектура в действительности представляет собой три первопорядковых режима с ортогональным модификатором, а не четыре параллельных режима.

Мы не разрешаем это напряжение здесь. Четырёхрежимная подача имеет преимущества простоты и формального параллелизма и адекватна той работе, которую остальная часть серии с ней делает. Второпорядковое прочтение имеет преимущества феноменологической точности — META действительно ощущается как перспектива над другими, а не как родная им, — и более точного отслеживания того, что делает движок статьи III, принимая META-чувствительные классификационные решения. Мы отмечаем это напряжение как структурный вопрос, к которому будущая теоретическая работа серии может вернуться, и продолжаем с четырёхрежимной формулировкой в дальнейшем, потому что именно её формально развил §5.

8.5 Мост к статье III

С тремя уровнями описания, сделанными эксплицитными, и двумя терминологическими флагами на месте, мы теперь можем чисто специфицировать, что статья III берёт из настоящей статьи и что добавляет.

Статья III берёт три вещи на вход. *Первое*: алфавит §4 — организованные по векторам семейства речевых операций — служит предварительным пространством гипотез для аннотационной онтологии статьи III. *Второе*: четыре precision regime §5 служат слоями анализа: классификационными измерениями, вдоль которых движок статьи III оценивает высказывание или последовательность, — не метками, которые приписываются, а архитектурными координатами, в пределах которых производится приписывание меток.

Третье: операциональный различитель экологичного/компенсаторного из §7 служит калибровочной целью движка: критерии, по которым операции получают непрерывные — не бинарные — баллы пропорциональности, обратимости и сохранения контакта, объединяемые в композитный различитель, операционализируемый в статье III как скаляр `hold_capacity` в интервале [0, 1].

Статья III добавляет то, чего настоящая статья не предоставляет и не могла бы. *Первое:* она специфицирует протокол аннотирования с конкретными лингвистическими признаками, процедурами `inter-rater reliability` и операциональными разрезами, по которым каждая операция алфавита идентифицируется в реальном тексте. *Второе:* она представляет архитектуру движка Mindloom как *structure-constrained LLM inference* — классификационный конвейер, в котором слои на основе правил (детекция ключевых слов, малые дообученные классификаторы, онтологические ограничения) определяют ограниченное пространство гипотез, внутри которого большая языковая модель осуществляет классификационное суждение. Это архитектурный вклад, операционализирующий то, что настоящая статья аргументировала как формально допустимое. *Третье:* статья III вводит десять режимов уровня движка (BUILD, SEEK, UNSEAL, LOCK, SEAL, DRAIN, FLOOD, EDGE, SHIFT, VOID) как эмпирически валидированные кластеры в продукт-пространстве вектор × precision regime, выведенные из пилотной аннотации и устойчивые на калибровочном корпусе. *Четвёртое:* она предоставляет анализ ограничений и честный анализ режимов отказа, предвосхищаемый методологическим обязательством §8.1: верифицированный по коду аудит того, где работа движка согласуется с теоретическим ожиданием и где — нет.

Разделение труда между двумя статьями, таким образом: статья II устанавливает *санкцию* для онтологически-ограниченной классификации компенсаторно-защитной речи — она аргументирует, что пространство гипотез, в пределах которого такая классификация оперирует, конечно, структурно и выводимо из независимых теоретических соображений. Статья III предоставляет *инструмент* — конкретный движок, работающий в санкционированном таким образом пространстве. Санкция не автоматически сертифицирует инструмент; ограничения статьи III — это её собственное обязательство формулировать. Но инструмент не имеет смысла без санкции; классификационный конвейер без принципиально обоснованного пространства гипотез — это просто детектор признаков, а это не то, чем статья III себя заявляет.

Финальное замечание о разделении труда со статьёй IV. Там, где статья III классифицирует статические конфигурации (какой режим, какие операции проявляются в этом высказывании или последовательности), статья IV трактует временные динамики (как режимы мигрируют между репликами, как траектории в пространстве аттракторов могут быть предсказаны, какие развивающиеся сигнатуры возникают в продолжительном диалоге). Четыре precision regime §5 — это статические аттракторы в настоящей статье; в статье IV они становятся бассейнами динамической системы, траектории которой имеют свою собственную формальную характеристику. Настоящая статья, таким образом, не адресует траекторию, переход или предсказание сколь-нибудь подробно; это — область следующей статьи последовательности.

9. Заключение

Вопрос, с которого эта статья началась, — методологический: допускает ли архитектура компенсаторно-защитной речи, описанная в статье I, формализацию, делающую эту речь хотя бы частично вычислимой? Мы утверждали, что допускает, и делали это по двум связанным осям.

Первая ось — конечный операциональный базис. Компенсаторно-защитная речь производится из ограниченного алфавита речевых операций — не шаблонов, не фиксированных фраз, — организованного четырьмя защитными векторами статьи I и составляемого по контекстно-чувствительным правилам. Четыре сходящихся аргумента — метаболический, клинический, предиктивно-вычислительный и развивающий — устанавливают, что этот базис конечен как архитектурный механизм, а не как эмпирическая случайность. Последовательности, порожаемые алфавитом, неограниченны; сам алфавит — нет.

Вторая ось — precision regimes как эмерджентные аттракторы. Те же операции, рассмотренные как темпоральные паттерны распределения precision, группируются в четыре повторяющиеся конфигурации иерархической генеративной модели: DEFENCE, COGNITIVE, ILLUSION, META. Это не категории, наложенные нами на данные; это аттракторо-подобные конфигурации распределения precision, возникающие внутри рамки предиктивной обработки как решения для классов предиктивно-стыдовых ситуаций, которыми компенсаторно-защитная речь управляет.

Вместе две оси устанавливают то, что статья поставила целью установить: пространство гипотез для компенсаторно-защитной речи одновременно конечно и теоретически обосновано. Оно конечно, потому что операциональный алфавит ограничен; оно обосновано, потому что ограниченность следует из рамки предиктивной обработки, а не из инженерного удобства, перечислительного удобства или какого-либо другого внешнего соображения. Это не утверждение о том, что может или не может классифицировать какой-либо конкретный движок. Это утверждение о пространстве, внутри которого любая классификация, каким бы способом она ни выполнялась, работает, — и о том, почему это пространство имеет ту форму, которую имеет.

Методологическое следствие — это санкция, в которой нуждается статья III. Онтологически-ограниченная классификационная архитектура для компенсаторно-защитной речи, при таком описании, не инженерный выбор среди альтернатив; это формализация, специфически санкционируемая нижележащей архитектурой. Статья III операционализирует эту санкцию в конкретном движении — structure-constrained LLM inference, с калибровкой, с честной отчётностью об ограничениях и с эмпирической работой, которую одна санкция выполнить не может. Санкция и инструмент — различные вклады, и ни один не заменяет другого.

Что это делает возможным для поля, мы думаем, стоит сказать без переоценки. Изучение защитной речи десятилетиями ограничивалось отсутствием общего операционального языка, связывающего клиническое наблюдение с вычислительной работой. Клиницисты именовали феномены; вычислительные подходы к смежным задачам преимущественно оперировали через детекцию поверхностных

признаков; ни одна из традиций не имела доступа к тому, что нужно другой. Архитектура, артикулированная в статьях I-III, предлагает общий операциональный язык — векторы из биологии, операции из лингвистики, режимы из предиктивной нейронауки, различители из клинического суждения, — внутри которого две традиции наконец могут адресовать один и тот же объект совместимыми инструментами. Статьи IV, V и VI развернут эмпирические и прикладные следствия. Настоящая статья делает одну вещь, и мы надеемся, что делает её чисто: она специфицирует формальное пространство, внутри которого это развитие может происходить.

Приложение А — Формальная характеристика

The four precision regimes developed in §5 are given a formal characterization below. The purpose is to define the hyperparameter space within which regimes are distinguished, to exhibit a minimal worked numerical illustration, to extend the formalism to dyadic interaction, and to list the five falsifiable predictions that the formalism generates. The mathematical material is adapted from prior theoretical work in the same programme; what the present appendix contributes is the placement of that material in the computability-warrant architecture of the main text.

Четыре precision regime, развитые в §5, получают формальную характеристику ниже. Цель — определить пространство гиперпараметров, внутри которого режимы различаются, предъявить минимальную проработанную численную иллюстрацию, расширить формализм до диадического взаимодействия и перечислить пять фальсифицируемых предсказаний, которые порождает формализм. Математический материал адаптирован из предшествующей теоретической работы в той же программе; собственный вклад настоящего аппендикса состоит в размещении этого материала в архитектуре computability-санкции основного текста.

A.1 Вариационная свободная энергия и precision-взвешенный message-passing

В вариационной формулировке активного вывода (Friston, 2010) агент поддерживает генеративную модель $p(o, s) = p(o|s) p(s)$, связывающую наблюдения o с латентными состояниями s , и минимизирует вариационную свободную энергию $F = E_q[\ln q(s) - \ln p(o, s)]$, являющуюся верхней границей для surprisal $-\ln p(o)$. В предиктивно-кодингвых имплементациях обновления убеждений принимают форму градиентного спуска по F относительно апостериорных средних μ (Friston, 2005).

В линейно-гауссовских условиях precision — это явно обратная дисперсия. Если правдоподобие $p(o|s) = N(s, \Pi_o^{-1})$, а prior $p(s) = N(\mu_0, \Pi_s^{-1})$, апостериорное среднее удовлетворяет

$$\mu_{\text{post}} = (\Pi_s \cdot \mu_0 + \Pi_o \cdot o) / (\Pi_s + \Pi_o),$$

где $k = \Pi_o / (\Pi_s + \Pi_o)$ — калмановский gain, вес, присваиваемый входящим свидетельствам относительно prior-ожидания. Это формальный объект,

используемый на протяжении всего аппендикса. Он делает конкретным то, как precision regimes различаются: сдвиги режима — это сдвиги в Π_s , Π_o или в обоих.

Расширения на иерархические среды заменяют μ_0 на message-passing с более высоких уровней, вводят precision-взвешенные ошибки предсказания ε на каждом уровне и расширяются на выбор политик через ожидаемую свободную энергию G . Эти расширения не меняют базового отношения между precision-allocation и траекторией вывода, которое использует оставшаяся часть аппендикса.

A.2 Precision regimes как наборы гиперпараметров

Определим precision regime как структурированный набор гиперпараметров

$$R = \{ \Pi_o(l, m), \Pi_s(l, m), \Pi_p(l), \gamma(l) \},$$

где l индексирует иерархический уровень, m — модальность (слуховая, семантическая, интероцептивная), Π_o — precisions правдоподобия, Π_s — precisions prior'a, Π_p — precisions политики, γ — относительный вес эпистемической ценности в ожидаемой свободной энергии. Четыре режима, введённые в §5, специфицируются как ограничения на это пространство гиперпараметров:

- **DEFENCE**: повышенная Π_s на самозащитных priors в подпространствах, релевантных угрозе; сниженная Π_o на disconfirmatory-каналах; сниженная γ (эпистемические действия грозят дестабилизацией защищённого prior'a).
- **COGNITIVE**: умеренные, контекстно-чувствительные Π_s и Π_o ; γ сохранена, поддерживая поиск информации и кооперативную репарацию; prior means откалиброваны.
- **ILLUSION**: смещённые prior means μ_0 в специфических подпространствах, и/или повышенная Π_s с избирательным снижением Π_o на каналах, которые бы развеивали смещение; выбор политики избегает disambiguating-свидетельств (низкая эффективная γ в релевантных подпространствах); самоподдерживается, потому что подтверждающие свидетельства переоцениваются, а опровергающие — недооцениваются.
- **META**: вывод более высокого порядка над самими Π_o и Π_s , с обучением волатильности и мета-precision, позволяющим гибкие переходы между тремя первопорядковыми режимами; эпистемические действия становятся эксплицитными выборами.

Эти спецификации — теоретические ограничения, а не подогнанные параметризации. Их цель — определить пространство вычислительных моделей, различимых эмпирической работой, а не фиксировать конкретные числовые значения, которые какой-либо конкретный говорящий якобы проявляет.

А.3 Тоу-модель: минимальный двухслойный линейно-гауссовский переключатель режимов

А.3.1 Модель Приведём минимальный проработанный пример, схватывающий DEFENCE как prior-доминированное обновление, ILLUSION как смещённые priors плюс сниженную сенсорную precision и META как вывод над precision.

Пусть s — латентное состояние, релевантное дискурсу, а o — наблюдение.

- Правдоподобие: $p(o|s) = N(s, \Pi_o^{-1})$.
- Prior: $p(s) = N(\mu_0, \Pi_s^{-1})$.
- Апостериорное среднее: $\mu_{\text{post}} = (\Pi_s \cdot \mu_0 + \Pi_o \cdot o) / (\Pi_s + \Pi_o)$.

Инстанции режимов (иллюстративные значения):

- **COGNITIVE**: $\Pi_s = \Pi_o = 1$ (сбалансированно).
- **DEFENCE**: $\Pi_s = 10, \Pi_o = 1$ (prior доминирует).
- **ILLUSION**: смещённое $\mu_0 \neq s^*$ (целевое состояние) и сниженная Π_o (например, $\Pi_o = 0.3$).
- **META**: $\Pi_o = \exp(\lambda)$, где λ выводится из статистик ошибок предсказания.

А.3.2 Проработанная численная иллюстрация Допустим $\mu_0 = 1$ (prior-убеждение) и $o = 0$ (сильное опровергающее свидетельство). Сравним режимы:

- **COGNITIVE**: $\mu_{\text{post}} = (1 \cdot 1 + 1 \cdot 0) / (1 + 1) = 0.50$ — сбалансированная интеграция; убеждение смещается на полпути к свидетельству.
- **DEFENCE**: $\mu_{\text{post}} = (10 \cdot 1 + 1 \cdot 0) / (10 + 1) \approx 0.91$ — убеждение едва обновляется, несмотря на противоречивое свидетельство; Π_s доминирует.
- **ILLUSION** ($\Pi_o = 0.3$): $\mu_{\text{post}} = (1 \cdot 1 + 0.3 \cdot 0) / (1 + 0.3) \approx 0.77$ — частичное обновление, остающееся далёким от свидетельства; сенсорная precision подавлена.

Два наблюдения имеют значение. *Первое*: психологически схожая «резистентность к свидетельствам» может возникать из различных вычислительных механизмов — либо повышением Π_s (DEFENCE), либо снижением Π_o (ILLUSION). Одна и та же поведенческая сигнатура (μ_{post} близко к μ_0 , далеко от o) может иметь две структурно различные причины. *Второе*: DEFENCE и ILLUSION поведенчески похожи на уровне апостериорного среднего, но резко различаются во *временной сигнатуре*: DEFENCE триггерится угрозой и разрешается при удалении угрозных cue; ILLUSION самоподдерживается и сохраняется после удаления угрозы. Временная диссоциация (развёрнутая как Prediction 2 в §A.5) — ключевая эмпирическая рукоять для различения этих двух режимов.

Эта двухпараметрическая иллюстрация не исчерпывает поведение модели — полная модель включает иерархическую структуру, выбор политик и модально-специфичные precisions. Назначение иллюстрации — сделать конкретным то, что спецификации гиперпараметров §A.2 доставляют вычислительно, а не заменить полный формализм.

А.3.3 Псевдокод симуляции Следующий псевдокод иллюстрирует качественное сопряжение между precision и обновлением в разных режимах. Он намеренно абстрактен; теоретический объект — это структура сопряжения, а не конкретная имплементация.

Алгоритм 1. Precision-режимный речевой вывод. (Псевдокод выше; идентичен для обеих языковых версий; *target* — целевое значение ε^2 , η — скорость обучения *meta-precision*.)

А.3.4 Ожидаемые сигнатуры симуляции При режимном расписании вида COGNITIVE → DEFENCE → META модель предсказывает три качественные сигнатуры:

- Апостериорное среднее μ «защемлено» в DEFENCE (минимальное движение в сторону свидетельства), релаксирует обратно к свидетельству в META по мере рекалибровки precision и пропорционально отслеживает свидетельство в COGNITIVE.
- Ошибка предсказания ε накапливается в ригидных режимах (где μ не может сдвинуться, чтобы соответствовать μ), и уменьшается после рекалибровки precision в META.
- Выводимая precision Π_o в META демонстрирует адаптивную модуляцию, управляемую статистиками ошибок, — снижается при волатильности среды, повышается при её стабильности, — со скоростью обучения η , контролирующей временной масштаб.

Это сигнатуры, которые должен демонстрировать time-series анализ симуляций режимного расписания, — и которые предназначены зондировать эмпирические парадигмы §А.5.

А.4 Расширение на связанный диадический вывод

А.4.1 Режим-зависимое сопряжение Коммуникация по своей природе вовлекает двух агентов; одноагентный формализм выше должен быть расширен. Вслед за Friston и Frith (2015) диалог моделируется как связанная динамическая система, в которой речевые акты каждого агента становятся наблюдениями другого.

Пусть Агент А находится в режиме R_A , а Агент В — в режиме R_B . Апостериор каждого агента обновляется согласно его собственным precision-настройкам:

$$\mu_A(t+1) = f(\mu_A(t), a_B(t); \Pi_A(R_A)), \mu_B(t+1) = f(\mu_B(t), a_A(t); \Pi_B(R_B)),$$

где $a_X(t)$ обозначает речевые акты Агента X в момент t (определяемые выбором политики под ожидаемой свободной энергией), а $\Pi_X(R_X)$ — precision-конфигурация, определяемая текущим режимом Агента X.

А.4.2 Динамика режимных пар Разные режимные пары дают качественно разные динамики взаимодействия. Четыре пары особенно диагностичны.

- **COGNITIVE-COGNITIVE**: взаимная чувствительность к свидетельствам поддерживает конвергентное обновление убеждений; ошибки предсказания обоих агентов движут выравниванием, приближаясь к кооперативной базовой линии стандартных interactive-alignment описаний.
- **DEFENCE-COGNITIVE**: асимметричное сопряжение. Сниженная P_o на disconfirmatory-каналах у DEFENCE-агента ослабляет влияние сигналов COGNITIVE-агента; выравнивание ограничено. COGNITIVE-агент может испытывать накапливающуюся ошибку предсказания, которую обыкновенные разговорные ходы не разрешают, — потенциально запуская его собственный сдвиг режима, в DEFENCE (реципрокность) или в META (распознавание, что канал заблокирован).
- **ILLUSION-ILLUSION**: если оба агента разделяют совместимые нарративные смещения, диада может стабилизироваться в взаимно подкрепляющем локальном минимуме. Поверхностное выравнивание может казаться высоким (высказывания каждого агента подтверждают общие priors), тогда как предиктивная точность в новых контекстах деградирует. Это динамика, характеризующая закрытые идеологические пары.
- **META-COGNITIVE**: вывод более высокого порядка у META-агента даёт возможность адаптивной регуляции precision в ответ на сигналы COGNITIVE-агента. Репарация доступна, возмущения обрабатываются изящно, — и предсказывается, что именно эта пара даёт наиболее устойчивое выравнивание и ёмкость для репарации. В клиническом сеттинге терапевт в идеале в META; в образовательном — учитель в META относительно студентов в COGNITIVE.

Эти качественные предсказания и есть то, что призвана проверять эмпирическая программа §A.5.

A.5 Пять фальсифицируемых предсказаний

Рамка выше порождает пять предсказаний, каждое сформулировано в фальсифицируемой форме, с указанием парадигмы, мер и конкурирующей гипотезы.

A.5.1 Выравнивание, обусловленное режимом, диссоциирует от общего выравнивания *Утверждение.* Разговорное выравнивание выше и устойчивее, когда оба собеседника находятся в режимах COGNITIVE или META; видимое выравнивание в ILLUSION диссоциирует — поверхностное выравнивание может сохраняться, тогда как предиктивная точность в новых контекстах деградирует.

Парадигма. Кооперативная диалоговая задача с периодическими возмущениями (неожиданные сдвиги темы; противоречивая информация).

Меры. Поведенческие индексы выравнивания; предсказуемость следующего высказывания; частота репарации.

Конкурирующая гипотеза. Общее interactive alignment предсказывает, что выравнивание монотонно коррелирует с priming и общими репрезентациями,

без режим-зависимого контроля precision.

A.5.2 Proxies для precision отслеживают переходы режимов с режим-специфичными временными сигнатурами *Утверждение.* Переходы в DEFENCE соответствуют измеримым снижениям эффективной сенсорной precision, разрешающимся при удалении угрозных cue. Переходы в ILLUSION демонстрируют снижения сенсорной precision, сохраняющиеся после удаления угрозы. Эта временная диссоциация — ключевая эмпирическая рукоять для различения этих двух режимов, иначе похожих на уровне поведения апостериорного среднего.

Парадигма. Манипуляция угрозой (социальная оценка; временное давление), сопровождаемая безопасным периодом восстановления, в ходе разговора.

Меры. Дилатация зрачка (прокси возбуждения/неопределённости); взгляд и eye-tracking (распределение внимания); ERP-маркеры нарушения ожидания.

Конкурирующая гипотеза. Чисто arousal-основанное описание предсказывает монотонные эффекты стресса без специфического паттерна режим-зависимой динамики восстановления.

A.5.3 Нейрональная синхрония отражает режим-обусловленное сопряжение

Утверждение. Межмозговая синхрония в реальном взаимодействии зависит от режимной пары: диады META-COGNITIVE демонстрируют устойчивое сопряжение в сетях более высокого порядка; диады DEFENCE-COGNITIVE — сниженное сопряжение и повышенную асимметрию.

Парадигма. Hyperscanning EEG/MEG или fMRI в ходе целенаправленного разговора.

Меры. Фазовая синхрония (theta/alpha/beta); межмозговая каузальность Грейнджера или transfer entropy; hyperscanning effective connectivity.

Конкурирующая гипотеза. Описания stimulus-driven сопряжения предсказывают сопряжение преимущественно на сенсорных уровнях; режимная гипотеза предсказывает различия в сопряжении даже при сопоставимой структуре стимулов.

A.5.4 Нейромодуляторные возмущения сдвигают режимы через precision-механизмы

Утверждение. Манипуляции, затрагивающие нейромодуляторные системы, связанные с неопределённостью и precision (холинергические и норадренергические влияния; Yu & Dayan, 2005), сдвигают баланс между DEFENCE-подобным и COGNITIVE/META-подобным обновлением, измеримо в дискурсивной репарации и обработке нарушения ожидания.

Парадигма. Фармакологический challenge (где этически уместно) или манипуляция на уровне задачи ожидаемой против неожиданной неопределённости.

Меры. Поведенческая интеграция свидетельств; EEG-маркеры; зрачковые реакции.

Конкурирующая гипотеза. Не-precision описания предсказывают глобальные изменения производительности без специфических изменений в сигнатурах взвешивания свидетельств.

Это предсказание сформулировано консервативно: оно не ссылается на специфические фармакологические агенты или спекулятивные модели действия психоделиков, а на хорошо установленную связь между нейромодуляторными системами и оценкой precision (Yu & Dayan, 2005; Feldman & Friston, 2010).

А.5.5 Метакогнитивный тренинг усиливает МЕТА-сигнатуры *Утверждение.*

Краткий тренинг в эпистемической скромности и явной отчётности о неопределённости повышает МЕТА-подобный вывод и улучшает диадическую предиктивную точность под воздействием, относительно общего разговорного тренинга.

Парадигма. Рандомизированная интервенция: метакогнитивный тренинг против контроля, затем разговор в условиях неоднозначности.

Меры. Отчёты об уверенности; калибровка; частота репарации; предиктивная точность; меры нейронального сопряжения.

Конкурирующая гипотеза. Общее научение навыкам объясняет улучшения без моделирования precision более высокого порядка; байесовское сравнение моделей должно отдавать предпочтение моделям с мета-precision, если теория верна.

А.6 Пайплайны анализа и сводка

Через парадигмы мы рекомендуем анализы, соответствующие конкурирующим генеративным моделям: иерархическое байесовское сравнение моделей (тестирование описаний с prior-precision против likelihood-precision против mean-bias); динамическое каузальное моделирование для нейроимиджинга (тестирование изменений gain, предсказанных сдвигами precision); time-frequency анализ для EEG/MEG (тестирование сдвигов осцилляторного сопряжения с режимом); и метакогнитивные метрики, такие как meta-d' (операционализирующая вывод над выводом; Maniscalco & Lau, 2012).

Таблица А.1 сводит отображение четырёх режимов на канонические компоненты активного вывода.

Table A.1. Mapping of four precision regimes to active-inference components.
/ Отображение четырёх precision regime на компоненты активного вывода.

Component / Компонент	DEFENCE	COGNITIVE	ILLUSION	META
Prior precision Π_s / Prior-precision Π_s	High (threat subspaces) / Высокая (подпространств угрозы)	Moderate, calibrated / Умеренная, калиброванная	May be high; means biased / Может быть высокой; средние смещены	Inferred; adaptive / Выводимая; адаптивная
Likelihood precision Π_o / Likelihood-precision Π_o	Reduced on disconfirmatory channels / Снижена на disconfirmatory-каналах	Moderate, preserved / Умеренная, сохранена	Selectively reduced / Избирательно снижена	Inferred via λ / Выводимая через λ
Epistemic weight γ / Эпистемический вес γ	Low (risk-averse) / Низкий (рискоизбегающий)	Preserved / Сохранён	Low in relevant subspaces / Низкий в релевантных подпространствах	High; epistemic actions explicit / Высокий; эпистемические действия эксплицитны
Prior means μ_0 / Prior-средние μ_0	Unbiased (precision-dominated) / Несмещённые (доминирует precision)	Calibrated / Откалиброванные	Biased ($\mu_0 \neq s$) / Смещённые ($\mu_0 \neq s$)	Calibrated / Откалиброванные
Policy type / Тип политики	Avoidance, re-framing / Избегание, переформулирование	Information-seeking / Поиск информации	Confirmation-seeking / Поиск подтверждения	Repair, feedback-seeking / Репарация, поиск обратной связи
Temporal dynamics / Временная динамика	Threat-triggered; resolvable / Триггерится угрозой; разрешима	Stable baseline / Стабильная базовая линия	Self-sustaining attractor / Самоподдерживающийся аттрактор	Slow regulatory timescale / Медленный регуляторный масштаб

Благодарности

Структурная разработка настоящей статьи и всей серии *Defense, Self, and Speech: From Biological Vectors to Linguistic Form* велась в расширенном диалоге с Claude (Anthropic), который участвовал в формулировке комбинаторного аргумента, четырёхлинейного схождения свидетельств, алфавита речевых

операций и четырёх-режимной precision-архитектуры. Автор несёт полную ответственность за содержание, утверждения и выводы настоящей работы, а также за любые оставшиеся в ней ошибки.

Литература

[To be compiled and merged with Paper I references for consistency across series]

Core references specific to Paper II:

- Bion, W. R. (1962). *Learning from Experience*. Heinemann.
- Bowen, M. (1978). *Family Therapy in Clinical Practice*. Jason Aronson.
- Bowlby, J. (1980). *Attachment and Loss, Vol. 3: Loss*. Basic Books.
- Chomsky, N. (1957). *Syntactic Structures*. Mouton.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. MIT Press.
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3).
- Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social information-processing mechanisms in children's social adjustment. *Psychological Bulletin*, 115(1).
- Dodge, K. A. (1980). Social cognition and children's aggressive behavior. *Child Development*, 51(1).
- Feldman, H., & Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience*, 4, 215.
- Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience*, 8, 443.
- Fonagy, P., Gergely, G., Jurist, E., & Target, M. (2002). *Affect Regulation, Mentalization, and the Development of the Self*. Other Press.
- Freud, A. (1936/1966). *The Ego and the Mechanisms of Defense*. International Universities Press.
- Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, 11(2).
- Friston, K. (2019). A free energy principle for a particular physics. *arXiv:1906.10184*.
- Friston, K., & Frith, C. (2015). Active inference, communication and hermeneutics. *Cortex*, 68.
- Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2017). Active inference: a process theory. *Neural Computation*, 29(1).
- Hazan, C., & Shaver, P. R. (1987). Romantic love conceptualized as an attachment process. *Journal of Personality and Social Psychology*, 52(3).

- Hesse, E. (2016). The Adult Attachment Interview: Protocol, method of analysis, and selected empirical studies. In J. Cassidy & P. R. Shaver (Eds.), *Handbook of Attachment* (3rd ed.). Guilford.
- Hohwy, J. (2013). *The Predictive Mind*. Oxford University Press.
- Kernberg, O. (1975). *Borderline Conditions and Pathological Narcissism*. Jason Aronson.
- Klein, M. (1946). Notes on some schizoid mechanisms. *International Journal of Psycho-Analysis*, 27.
- Main, M., & Goldwyn, R. (1998). *Adult Attachment Scoring and Classification System*. Unpublished manuscript, University of California at Berkeley.
- Main, M., & Hesse, E. (1990). Parents' unresolved traumatic experiences are related to infant disorganized attachment status. In M. T. Greenberg, D. Cicchetti, & E. M. Cummings (Eds.), *Attachment in the Preschool Years*. University of Chicago Press.
- McEwen, B. S. (1998). Stress, adaptation, and disease: Allostasis and allostatic load. *Annals of the New York Academy of Sciences*, 840.
- Mikulincer, M., & Shaver, P. R. (2016). *Attachment in Adulthood: Structure, Dynamics, and Change* (2nd ed.). Guilford.
- Parr, T., Pezzulo, G., & Friston, K. J. (2022). *Active Inference: The Free Energy Principle in Mind, Brain, and Behavior*. MIT Press.
- Porges, S. W. (2011). *The Polyvagal Theory*. W. W. Norton.
- Ramstead, M. J. D., Badcock, P. B., & Friston, K. J. (2018). Answering Schrödinger's question: A free-energy formulation. *Physics of Life Reviews*, 24.
- Schauer, M., & Elbert, T. (2010). Dissociation following traumatic stress. *Zeitschrift für Psychologie*, 218(2).
- Seth, A. K. (2021). *Being You: A New Science of Consciousness*. Dutton.
- Sifneos, P. E. (1973). The prevalence of "alexithymic" characteristics in psychosomatic patients. *Psychotherapy and Psychosomatics*, 22(2-6).
- Taylor, G. J., Bagby, R. M., & Parker, J. D. A. (1997). *Disorders of Affect Regulation*. Cambridge University Press.
- Van der Kolk, B. (2014). *The Body Keeps the Score*. Viking.
- Walker, P. (2013). *Complex PTSD: From Surviving to Thriving*.
- Winnicott, D. W. (1960). Ego distortion in terms of true and false self. In *The Maturational Processes and the Facilitating Environment*.