

## Эмоции и диспозиционные предикаты

*Синицкий Д. А.,*  
ИАТЭ НИЯУ МИФИ, г. Обнинск  
[das\\_sid@mail.ru](mailto:das_sid@mail.ru)

**Аннотация:** Аффективные вычисления и исчисления представляют собой в настоящее время одну из интенсивно развивающихся областей междисциплинарных исследований. Целью статьи является рассмотрение подхода к построению формализованной теории эмоций, развитого голландскими логиками Б. Штейнбринком, М. Дагани и Дж.-Дж. Ч. Мейером, а также анализ этой теории на предмет использования в ней диспозиционных предикатов. Редукционные определения для предикатов наблюдения и диспозиционных предикатов в анализируемой теории эмоций конструируются как определения триггеров эмоций и предикатов переживания эмоций, которые затем редуцируются до выражений языка эпистемической и динамической логик.

**Ключевые слова:** аффективные вычисления, аффективные исчисления, эмоция, диспозиционный предикат.

Аффективные вычисления и исчисления представляют собой в настоящее время одну из интенсивно развивающихся областей междисциплинарных исследований. Термин «аффективные вычисления» (affective computing) был введен в оборот в 1997 году Розалиндой Пикард<sup>1</sup> и с тех пор охватывает совокупность теоретических и прикладных областей, так или иначе касающихся формализации и логического анализа психологических теорий эмоций и психоэмоциональных (аффективных) процессов, с последующими приложениями в технологиях моделирования эмоций и аффективных состояний и в технологиях моделирования способов выражения эмоциональных состояний и процессов в системах искусственного интеллекта и робототехнических устройствах, а также в виртуальных и реальных мультиагентных (многохарактерных) средах.

Наряду с термином «аффективные вычисления» мы предлагаем также использовать термин «аффективные исчисления» (affective calculus) для обозначения того класса прикладных логических исчислений, которые используются в качестве синтаксической и семантической формализации теорий эмоций и аффективных состояний и процессов.

Прикладные цели авторов подобных исследований весьма разнообразны — от автоматического распознавания психоэмоциональных состояний агентов в той или иной социальной среде до моделирования характеров и внешнего выражения психоэмоциональных реакций персонажей в компьютерных играх и в анимированных системах.

---

<sup>1</sup> Picard R. W. Affective Computing. — MIT Press, 1997.

По ряду причин, о которых будет сказано ниже, наше внимание привлёк подход, развитый группой голландских логиков и специалистов по ИТ из Утрехтского университета, работавших в рамках нескольких проектов, в том числе и в рамках проекта разработки эмоционального роботизированного человеко-компьютерного интерфейса (робот-компаньон iCat фирмы Philips).

Во-первых, авторы этого подхода — Б. Штейнбринк, М. Дагани и Дж.-Дж. Ч. Мейер — осуществили логико-синтаксическую и логико-семантическую формализацию широко известной теоретико-психологической модели эмоций, восходящей к работам А. Ортона, Дж. Л. Клора и А. Коллинза. Модель эмоций Ортона-Клора-Коллинза мы далее для краткости будем обозначать **OCC**, а формализацию, выполненную Б. Штейнбринком, М. Дагани и Дж.-Дж. Ч. Мейером, — **SDM**, по начальным буквам фамилий авторов. Здесь мы опираемся и ссылаемся прежде всего на известную работу А. Ортона, Дж. Л. Клора и А. Коллинза «Когнитивная структура эмоций»<sup>2</sup> и диссертационное исследование Бастиана Штейнбринка «Логическая структура эмоций»<sup>3</sup>, но см. также библиографию в этой диссертации, где автор указывает совместные с М. Дагани и Дж.-Дж. Ч. Мейером работы, из которых мы отметим такие, как «Логика эмоций для разумных агентов»<sup>4</sup>, «Формальная модель эмоций: объединённые качественные и количественные аспекты»<sup>5</sup>, «Пересмотр модели **OCC**»<sup>6</sup>.

Во-вторых, данный подход интересен тем, что он представляет собой не просто формализованное в языке комбинированной (доксатической + динамической) логики (язык **ZAPL** плюс язык **KARO**) описание 22 типов эмоций **OCC**, интерпретированное в терминах двухслойной крипковской семантики (интенциональной семантики моделей полагания-убеждения + трансформационной семантики действий), но, по сути, предлагает мощный инструментарий для логического анализа и сравнения различных теорий психоэмоциональных состояний, их взаимосвязи с процессами принятия решений и выполнения действий.

В-третьих (и это — главное для целей нашей статьи), наше внимание формализация **SDM** привлекла ещё и потому, что она является очень наглядным воплощением идеи так называемой эмпирической интерпретации диспозиционных

---

<sup>2</sup> Ortony A., Clore G. L., Collins A. The Cognitive Structure of Emotions. — Cambridge University Press. 1990. 224 p.

<sup>3</sup> Steunebrink Bas. R. The Logical Structure of Emotions. — SIKS Dissertation Series № 2010-23 Printed by GVO drukkers & vormgevers B. V. j Ponsen & Looijen, Ede. 2010. 228 p.

<sup>4</sup> Bas R. Steunebrink, Mehdi Dastani, and John-Jules Ch. Meyer. A logic of emotions for intelligent agents. In Proceedings of the Twenty-Second Conference on Artificial Intelligence (AAAI'07). AAAI Press, 2007.

<sup>5</sup> Bas R. Steunebrink, Mehdi Dastani, and John-Jules Ch. Meyer. A formal model of emotions: Integrating qualitative and quantitative aspects. In Ghallab Mali, Constantine D. Spyropoulos, Nikos Fakotakis, and Nikos Avouris, editors, Proceedings of the 18th European Conference on Artificial Intelligence (ECAI'08), pages 256–260. IOS Press, 2008.

<sup>6</sup> Bas R. Steunebrink, Mehdi Dastani, and John-Jules Ch. Meyer. The OCC model revisited. In Dirk Reichardt, editor, 4th Workshop on Emotion and Computing, Paderborn, Germany, 2009.

предикатов, сформулированной выдающимся отечественным логиком В. А. Смирновым<sup>7</sup>.

Данная статья посвящена анализу формализованной модели эмоций **SDM** в контексте обсуждения диспозиционности эмоций.

### 1. Допущения, лежащие в основе модели эмоций **OCC**

*Я постараюсь показать, что слово «эмоция» используется для обозначения по меньшей мере трех или четырех явлений разного рода, которые я буду называть «наклонностями» (inclinations) (или «мотивами»), «настроениями» (moods), «возбуждениями» (agitations) (или «нервными потрясениями») и «чувствами» (feelings). Наклонности и настроения, включая возбуждения, не суть события, следовательно, не происходят ни публично, ни приватно. Они являются предрасположенностями (propensities), а не действиями или состояниями. Однако это предрасположенности разного рода, и различия между ними существенны. Чувства, с другой стороны, суть нечто происходящее, но место, которое они должны занимать в описаниях человеческого поведения, весьма отличается от того, которое отводят им стандартные теории. В отличие от мотивов, то подобно болезням и состояниям погоды, настроения или расположения духа представляют собой временные условия, которые определенным образом объединяют события, но сами по себе не являются некими дополнительными событиями.*

*Гилберт Райл. Понятие сознания<sup>8</sup>*

Многие авторы отмечают имеющиеся неопределенности и многозначность в отношении употребления термина «эмоция», а также других терминов, обозначающих различные психоэмоциональные состояния и процессы. Не вдаваясь в подробности психологических дискуссий относительно употребления этих терминов, отметим, что модель **OCC** использует следующие важные допущения в основаниях теории эмоций.

1.1. Временная продолжительность является одной из основных характеристик, позволяющих различать психоэмоциональные процессы и явления, такие как, например, аффекты (или импульсы), собственно эмоции, настроения, темперамент и характер. Эмоции попадают в темпоральные промежутки от нескольких секунд или минут до нескольких часов (в крайних случаях).

<sup>7</sup> Смирнов В. А. Логические методы анализа научного знания. — М.: «Наука», 1987. — 256 с.

<sup>8</sup> Райл Г. Понятие сознания. Перевод с англ. — М.: Идея-Пресс, Дом интеллектуальной книги, 1999. — С. 90.

1.2. Другой отличительной особенностью эмоций является то, что эмоция всегда относится к чему-то (кому-то), вызывается чем-то (кем-то), направлена на что-то (на кого-то), и в этом смысле интенциональна.

Обе эти характеристики использовались в классификационных схемах различных авторов, в том числе и при построении модели **ОСС** (Ортони, Клор и Коллинз отмечают классификации Гросса и Томпсона, а также Оутли и Дженкинса, на которые они опираются).

1.3. Третье допущение — принятие, выражаясь языком психологов, идеаторного характера эмоций. Последнее означает, что эмоции могут соотноситься не только с актуально существующими стимулами и быть направлены на актуально существующие объекты, ситуации и события, но и на воображаемые, предполагаемые, ожидаемые и т. п., однако последние являются, если можно так выразиться, «совозможными» с миром актуально существующих стимулов.

1.4. Четвертое допущение предполагает трехаспектную структуру эмоций, укладывающуюся в схему «Оценка — Переживание — Регуляция».

Как отмечает Б. Штейнбринк, «...процесс оценки объединяет перцепцию, заинтересованность и индивидуальные параметры, что приводит к набору выявленных (вызванных) эмоций. После присвоения интенсивностей этим вызванным эмоциям (с указанием их силы) результат процесса оценки представляет собой набор новых переживаемых эмоций, которые дополняют существующий набор переживавшихся эмоций. ... Во многих ситуациях можно захотеть регулировать свои эмоции, чтобы уменьшить переживание отрицательных эмоций и продлить переживание положительных эмоций»<sup>9</sup>.

1.5. В модели **ОСС** эмоциональными являются агенты, которые перцептивно и когнитивно воспринимают мир, ставят перед собой цели и целесообразно действуют. Эмоции представляют собой не только многоуровневый реактивный механизм на стимулы, но и механизм управления когнитивным мышлением посредством нейромодуляции (гипотеза соматического маркера Дамасио).

Штейнбринк следующим образом комментирует эту гипотезу: «Дамасио действительно признает, что эмоции иногда могут быть вредными для рассуждений, но жить без эмоций — удел гораздо худший»<sup>10</sup>.

1.6. В **ОСС** предполагается, что оцениваться могут: а) последствия событий; б) действия агентов; в) аспекты объектов. При этом последствия событий оцениваются сообразно принимаемым агентом целям, действия оцениваются сообразно допускаемым агентом стандартам действий, а аспекты объектов оцениваются сообразно отношению к ним агента. Такое деление дает три типа оценочных значений

---

<sup>9</sup> Steunebrink Bas. R. The Logical Structure of Emotions. — SIKS Dissertation Series № 2010-23 Printed by GVO drukkers & vormgevers B. V. j Ponsen & Looijen, Ede. 2010. P. 7.

<sup>10</sup> Steunebrink Bas. R. The Logical Structure of Emotions. — SIKS Dissertation Series № 2010-23 Printed by GVO drukkers & vormgevers B. V. j Ponsen & Looijen, Ede. 2010. P. 3.

(оценочных шкал) — «желательность/нежелательность», «похвальность/порицаемость» и «привлекательность/непривлекательность», и служит одним из оснований деления для получения классификационной иерархии типов эмоций **ОСС**. Допущения 1.4, 1.5 и 1.6 позволяют отнести **ОСС** к так называемым «когнитивно-аксиологическим» теориям эмоций.

1.7. Важнейшим допущением **ОСС** (восходящим к идеям Кэррола Изарда) является допущение конечного набора типов эмоций, который можно выявить на основе предыдущих допущений, в первую очередь в зависимости от типа оценки, входящей в структуру эмоции.

**ОСС** содержит такой конечный набор типов эмоций, который структурирован в виде подробной классификационной иерархии, включающей 22 типа эмоций с детальными спецификациями каждого типа эмоций:

радость (joy), горе (distress), радость за другого (happy-for), жалость (pity), обида (resentment), злорадство (gloating), надежда (hope), страх (fear), удовлетворение (satisfaction), подтвердившийся страх (fears-confirmed), облегчение (relief), разочарование (disappointment), гордость (pride), стыд (shame), восхищение (admiration), укоризна (reproach), благодарность (gratitude), гнев (anger), удовольствие (gratification), раскаяние (remorse), симпатия (liking), антипатия (disliking).

Спецификация в **ОСС** каждого типа эмоций из перечисленных выше состоит из следующих элементов.

1) Типовая метка — наиболее удобный термин для обозначения эмоций данного типа. Их-то мы как раз и перечислили выше в русском переводе вместе с англоязычным термином первоисточника, не претендуя на уникальность перевода и отдавая отчет в том, что не существует взаимно-однозначного соответствия между множествами эмотивных терминов двух естественных языков. 2) Собственно спецификация типа эмоции — описание условий, порождающих эмоции данного типа. 3) Список меток — перечисление эмотивных терминов (эмотивов), используемых для обозначения и характеристики эмоций данного типа. 4) Список параметров, влияющих на интенсивность процесса переживания эмоций данного типа. 5) Иллюстративное описание ситуации, в которой может возникнуть эмоция данного типа.

У Б. Штейнбринка иерархия, образуемая перечисленными типами эмоций (правда, в несколько модифицированном виде), рассматривается как иерархия, порождаемая отношением наследования в терминологии объектно-ориентированного программирования, но с логической точки зрения это, по сути, иерархия понятий, связанных родо-видовым отношением. Мы не будем здесь изображать иерархию в графическом виде, отсылая к источнику<sup>11</sup>. Отметим только, что эта иерархия, как и критерии ее построения, лежит в основе формализации, обозначенной нами как **SDM**.

## 2. Допущения, лежащие в основе **SDM**

<sup>11</sup> Steunebrink Bas. R. The Logical Structure of Emotions. — SIKS Dissertation Series № 2010-23 Printed by GVO drukkers & vormgevers B. V. j Ponsen & Looijen, Ede. 2010. 228 p.

*Между тем для некоторых целей необходимо выявлять логические взаимоотношения тех понятий, которые мы весьма успешно умеем использовать. Попытка осуществить это применительно к понятиям, относящимся к способностям, действиям и состояниям сознания, всегда была значительной частью задачи философов.*

*Гилберт Райл. Понятие сознания*<sup>12</sup>

2.1. Основным допущением, лежащим в основе подхода **SDM**, является убеждение его авторов в продуктивности и эффективности средств современного логико-синтаксического и логико-семантического анализа применительно к психологическим теориям эмоций.

Б. Штейнбринк следующим образом сформулировал основные вопросы, которые позволяет решить логико-синтаксическая и логико-семантическая формализация неформальных психологических теорий эмоций, и, в частности, формализация модели **ОСС**. Какова структура типов эмоций, пригодная для точного описания в формализованном языке? Как можно строгим и однозначным образом (в формализованном языке) описать условия, порождающие эмоции? Какие аспекты определяют процесс переживания эмоций и как их формализовать? Как можно формализовать влияние эмоций на принятие решений и поведение?

Значимым является также и такое соображение, выдвинутое Б. Штейнбринком, как то, что формализованный логический язык гораздо ближе языку программного кода, чем язык неформальных психологических теорий, и, «таким образом, формализация создает мост от психологических моделей эмоций к вычислительным моделям эмоций»<sup>13</sup>.

2.2. Второе допущение, принимаемое авторами подхода **SDM**, — это принятие так называемой парадигмы **BDI** (Belief, Desire, Intention — Убеждение, Желание, Намерение). Данная парадигма позволяет объяснять поведение агентов через выбор осуществляемых ими действий с использованием трех ключевых понятий: убеждения, желания и намерения. Эта парадигма широко используется для разработки архитектур систем искусственного интеллекта, и ее применение в качестве концептуальной основы для формализации психологических теорий и моделей эмоций кажется вполне оправданным.

2.3. Третье, также немаловажное, допущение связано с принятием модели **ОСС** как наиболее подходящей для ее переопределения в терминах парадигмы **BDI**. Этому также способствовало наличие детальных спецификаций типов эмоций в модели **ОСС**,

---

<sup>12</sup> Райл Г. Понятие сознания. Перевод с англ. — М.: Идея-Пресс, Дом интеллектуальной книги, 1999.

<sup>13</sup> Steunebrink Bas. R. The Logical Structure of Emotions. — SIKS Dissertation Series № 2010-23 Printed by GVO drukkers & vormgevers B. V. j Ponsen & Looijen, Ede. 2010. P. 11.

по существу, систематизирующих полуформальным образом как условия порождения эмоций каждого типа (из 22), так и условия их переживания.

2.4. Авторами **SDM** была выявлена неоднозначность в трактовках **OCC**, приводящая к отличающимся друг от друга иерархическим классификационным моделям типов эмоций в зависимости от принимаемых оснований деления эмоций. С целью устранения неоднозначности авторы **SDM** останавливаются на иерархической классификационной модели, где основанием деления выступают вызывающие эмоцию условия, порождающие соответствующую эмоциональную оценку. Б. Штейнбринк как один из авторов **SDM** специально обращает внимание на то, что такая иерархия, как и иллюстрирующая ее диаграмма, были согласованы с авторами **OCC** в личном общении с А. Ортони и Дж. Л. Клором<sup>14</sup>.

2.5. Однако в контексте нашего обсуждения диспозиционности эмоций одним из самых значимых допущений, принимаемых авторами **SDM**, является четкое проведение различия в структуре эмоциональных эпизодов между начальными условиями, порождающими, или, выражаясь языком авторов **SDM**, «включающими» эмоцию — отсюда и их название — «триггеры» эмоций<sup>15</sup>, и процессом переживания эмоции, причем последнее зависит от интенсивности переживания, зависящей от целого ряда параметров (разных у отличающихся типов эмоций).

2.6. Соответственно, такое различие вводится и в синтаксис языка формализма, в котором отдельно определяются предикатные выражения, обозначающие триггеры эмоций различных типов, и предикатные выражения, обозначающие эмоциональное переживание. Далее мы детально опишем синтаксис **SDM** в «русифицированном» нами варианте (для удобства прочтения и ассоциирования), в отличие от беглого описания нами базовых семантических конструкций, для которых мы оставим в силу их абстрактного теоретико-множественного характера первоначальные авторские обозначения.

---

<sup>14</sup> Steunebrink Bas. R. The Logical Structure of Emotions. — SIKS Dissertation Series № 2010-23 Printed by GVO drukkers & vormgevers B. V. j Ponsen & Looijen, Ede. 2010. P. 23.

<sup>15</sup> По этой же причине авторы пользуются такими оборотами речи, как «включенная эмоция», «условия включения эмоций» и т. п. Конечно, можно было бы заменить подобные выражения при переводе другими, скажем, «порожденная эмоция», «условия, инициирующие эмоции» и т. п., но учитывая технический характер приложений обсуждаемого исследования, мы сохранили дословный перевод.

2.7. Триггеры эмоций синтаксически конструируются из предикатных констант (предикаторов)<sup>16</sup> — Позитив<sup>T</sup>, Негатив<sup>T</sup>, Одобрение<sup>T</sup>, Неодобрение<sup>T</sup>, Довольство<sup>T</sup>, Недовольство<sup>T</sup>, Симпатия<sup>T</sup>, Антипатия<sup>T</sup>, Надежда<sup>T</sup>, Страх<sup>T</sup>, Радость<sup>T</sup>, Горе<sup>T</sup>, Гордость<sup>T</sup>, Стыд<sup>T</sup>, Восхищение<sup>T</sup>, Укоризна<sup>T</sup>, Удовольствие<sup>T</sup>, Раскаяние<sup>T</sup>, Благодарность<sup>T</sup>, Гнев<sup>T</sup>, Удовлетворение<sup>T</sup>, Подтвердившийся-страх<sup>T</sup>, Облегчение<sup>T</sup>, Разочарование<sup>T</sup>, Радость-за-другого<sup>T</sup>, Жалость<sup>T</sup>, Обида<sup>T</sup>, Злорадство<sup>T</sup>, — а также индивидуальных констант и переменных, в стандартной функциональной записи с помощью скобок обозначающих приложение предикатора (как функтора) к аргументам. Аргументы (индивидуальные константы и переменные) могут быть различных сортов и даже типов, о чем будет сказано ниже подробнее.

Выражение **Эмоция<sup>T</sup>** относится к выражениям метаязыка и используется как метапеременная для обозначения любой из перечисленных выше предикатных констант. Верхний индекс «**T**» указывает на то, что предикатные константы используются для построения триггеров эмоций. Перечисленные предикатные константы отличаются друг от друга не только сортом и типом возможных аргументов, но и местностью, однако всегда одним из возможных аргументов является индивидуальная константа, указывающая на какого-либо агента, или индивидуальная переменная, «пробегающая» по множеству агентов, обычно она у авторов **SDM** (по соглашению) синтаксически выглядит как приписанный предикатной константе нижний индекс.

<sup>16</sup> Здесь мы используем термины «предикатор» и «предикатная константа» как синонимичные (поскольку в синтаксисе формализма нет предикатных переменных) и отличающиеся от термина «предикат». Если определять последний точно, то это правильно построенная формула в языке исчисления предикатов, содержащая свободные (не связанные кванторами) индивидуальные переменные того или иного типа и сорта. Такие формулы строятся стандартным для языка исчисления предикатов образом: определяется понятие термина; простые термины — это индивидуальные константы или индивидуальные переменные, составные термины строятся из простых терминов и функциональных констант (и функциональных переменных, если они есть); атомарные формулы получаются из предикаторов и терминов с помощью опять-таки стандартного функционального способа записи, когда предикатор выполняет роль функтора или оператора, индивидуальные переменные или константы, а возможно, и более сложно построенные термы — роль аргументов, а сама операция приложения оператора к аргументам изображается с помощью скобок; сложные формулы получаются из более простых с помощью пропозициональных связок или с помощью кванторов. В силу модального характера многих выражений из синтаксиса формализации **SDM** они могут иметь двойную трактовку — их можно рассматривать и как предикаты, и как модальные операторы. Ситуация усложняется еще и тем, что язык формализации **SDM**, а это язык комбинированной логики, по всей видимости, разрабатывался таким образом, чтобы быть практически полезным (или реализуемым) на основе парадигм логического программирования. В этом случае приходится учитывать, как можно сконструировать модальные операторы в терминах языка логики предикатов, используемого в качестве средства программирования. Авторы **SDM** сами используют различную терминологию для получаемых синтаксических конструкций — «триггер» (только применительно к перечисленным выше), «конструкт», «оператор» и даже «макрос». Мы далее будем использовать термин «предикатор», учитывая то, что предикаторы можно рассматривать как функторы, участвующие в синтаксическом построении и предикатов, и замкнутых формул (не содержащих свободных переменных), и модальных операторов, и операторов языка (логического) программирования.

Более подробно семантику и, соответственно, онтологические допущения, лежащие в основе **SDM**, мы обсудим ниже.

Соответственно, выражения метаязыка «*i*» и «*j*» используются как метапеременные для обозначения индивидуальных констант или переменных, указывающих на агентов (принимающих значения из некоторого фиксированного множества агентов); « $\alpha$ » и « $\beta$ » используются как метапеременные для обозначения индивидуальных констант или переменных, указывающих на действия; « $\pi$ » — как метапеременная для обозначения индивидуальных констант или переменных, указывающих на планы, конструируемые из действий; « $i:\alpha$ » и « $i:\beta$ » как метапеременные для обозначения индивидуальных констант или переменных, указывающих на действия и планы, выполненные указанными агентами; « $c$ » — как метапеременная для обозначения индивидуальных констант или переменных, указывающих на «последствия событий»; « $X$ » и « $Y$ » — как метапеременные для обозначения индивидуальных констант или переменных произвольного типа и сорта. Например, выражение вида **Одобрение<sup>T</sup><sub>i</sub>(j:α)** читается как «эмоция одобрения агента *j* за выполненное им действие  $\alpha$  включилась у агента *i*». В силу того, что в записи **Одобрение<sup>T</sup><sub>i</sub>(j:α)** используются метапеременные, мы и говорим «выражение вида», а не просто «выражение».

2.8. Двадцати восемью триггерам эмоций соответствуют предикаты переживания эмоций, конструируемые синтаксически аналогичным образом из двадцати восьми предикатных констант, отличающихся от перечисленных выше только отсутствием верхнего индекса «*T*». Если **Эмоция<sup>T</sup>** есть предикатная константа (предикатор), используемая для синтаксического построения триггера эмоций, то **Эмоция** есть предикатная константа (предикатор), используемая для синтаксического построения предиката переживания эмоций. Соответственно, выражение вида **Одобрение<sup>T</sup><sub>i</sub>(j:α)** читается как «эмоция одобрения агента *j* за выполненное им действие  $\alpha$  переживается агентом *i*», или, по-другому, «агент *i* испытывает эмоцию одобрения в отношении агента *j* за выполненное агентом *j* действие  $\alpha$ ».

2.9. Триггеры эмоций, как и предикаты переживания эмоций, являются определяемыми понятиями в рамках **SDM**. Поскольку понятия, соответствующие триггерам эмоций, образуют родо-видовую иерархию, авторы **SDM** конструируют систему определений, начиная с триггеров, соответствующих наиболее общим понятиям, и распространяя систему определений на триггеры, соответствующие видовым понятиям, добавляя в состав дефиниенсов отличительные признаки<sup>17</sup>. Аналогичным образом (от корня классификационного дерева к его листьям) определяются и предикаты переживания эмоций.

<sup>17</sup> См. определения (3.1.) — (3.31) в Steunebrink Bas. R. The Logical Structure of Emotions. — SIKS Dissertation Series № 2010-23 Printed by GVO drukkers & vormgevers B. V. j Ponsen & Looijen, Ede. 2010. 228 p.

2.10. Триггеры эмоций определяются в терминах предикатов перцептивного восприятия и в терминах предикатов оценки<sup>18</sup>, поэтому в синтаксис добавляются следующие предикатные константы:

а) для конструирования предикатов перцептивного восприятия: **Перцепция**, **ПерцепцияПоследствий**, **ПерцепцияДействий**, **ПерцепцияОбъектов**, **ПерцепцияСвязи**;

б) для конструирования предикатов оценки: **Желательно**, **Нежелательно**, **ДостойноПохвалы**, **ЗаслуживаетПорицания**, **Привлекательно**, **Непривлекательно**;

в) дополнительные предикатные константы, позволяющие конструировать предикаты, являющиеся терминами парадигмы **BDI**: **Ожидается-в-Будущем**, **ОжидаетсяСейчас**, **КогнитивноОбъединено**, **Подтверждает**, **Отвергает**, **Предполагает**;

г) темпоральный оператор **Было-или-Есть-истинно**.

В этих терминах конструируются дефиниенсы определений (3.1.) — (3.31) в работе Б. Штейнбринка «Логическая структура эмоций»<sup>19</sup>.

Как утверждает Б. Штейнбринк — и это, конечно же, еще одно допущение в основаниях формализма **SDM**, — «если спецификации... точны, то порождающие условия для типов эмоций модели **OCC** строятся с помощью не более чем семнадцати представлений, сформулированных в нашей формализации на основе семнадцати конструкторов, указанных выше»<sup>20</sup>.

2.11. Здесь следует отметить, что в качестве одного из допущений авторы **SDM** принимают тезис, в целом характерный для оснований когнитивной психологии, в соответствии с которым отождествляется перцептивное ощущение и осознание субъектом этого ощущения. Для этого авторы используют достаточно многосмысленный глагол «to perceive» (и производные от него), который можно перевести не только как «воспринимать», «чувствовать», «ощущать», но и как «понимать», «осознавать», «постигать». Отсюда — оригинальные названия таких предикаторов, как **Perceive**, **PerceiveConseq**, **PerceiveAction**, **PerceiveObject**, которые мы «русифицировали» выше (см. п. 2.10, подп. а)), используя термин «перцепция». В примечании 4 на стр. 7. своей диссертации Б. Штейнбринк пишет: «Поскольку нас интересуют эмоции, возникающие в результате когнитивных оценок, мы будем рассматривать только когнитивные результаты перцепции»<sup>21</sup>.

---

<sup>18</sup> Из последних конструируются дефиниенсы определений (3.1.) — (3.31) у Б. Штейнбринка.

<sup>19</sup> Steunebrink Bas. R. The Logical Structure of Emotions. — SIKS Dissertation Series № 2010-23 Printed by GVO drukkers & vormgevers B. V. j Ponsen & Looijen, Ede. 2010. 228 p.

<sup>20</sup> Steunebrink Bas. R. The Logical Structure of Emotions. — SIKS Dissertation Series № 2010-23 Printed by GVO drukkers & vormgevers B. V. j Ponsen & Looijen, Ede. 2010. P. 57.

<sup>21</sup> Steunebrink Bas. R. The Logical Structure of Emotions. — SIKS Dissertation Series № 2010-23 Printed by GVO drukkers & vormgevers B. V. j Ponsen & Looijen, Ede. 2010. P. 7.

2.12. Обобщенные предикаты положительной и отрицательной оценок строятся с помощью предикатных констант **Хорошо** и **Плохо** и определяются как дизъюнкции следующего вида<sup>22</sup>.

**Опр. 1:**

**Хорошо<sub>i</sub>(X) =<sub>df</sub>**

**=<sub>df</sub> Желательно<sub>i</sub>(X) ∨ ДостоинноПохвалы<sub>i</sub>(X) ∨ Привлекательно<sub>i</sub>(X)**

«Агент *i* оценивает *X* как «хорошо», по определению это означает, что агент *i* оценивает *X* как желательное последствие какого-то события, или как достойное похвалы действие, совершенное им самим или другим агентом, или как привлекательный объект (аспект объекта)», соответственно, *X* — метапеременная для обозначения индивидуальной константы или индивидуальной переменной одного из трех сортов — сорт «последствия событий», сорт «агент-действие», сорт «объект».

**Опр. 2:**

**Плохо<sub>i</sub>(X) =<sub>df</sub>**

**=<sub>df</sub> Нежелательно<sub>i</sub>(X) ∨ ЗаслуживаетПорицания<sub>i</sub>(X) ∨ Непривлекательно<sub>i</sub>(X)**

«Агент *i* оценивает *X* как «плохо», что по определению означает, что агент *i* оценивает *X* как нежелательное последствие какого-то события, или как заслуживающее порицания действие, совершенное им самим или другим агентом, или как непривлекательный объект (аспект объекта)», соответственно, *X* — по-прежнему метапеременная для обозначения индивидуальной константы или индивидуальной переменной одного из трех сортов — сорт «последствия событий», сорт «агент-действие», сорт «объект».

2.13. Обобщенный предикат перцептивного восприятия конструируется с помощью предикатной константы **Перцепция** (см. выше) как дизъюнкция предикатов следующего вида<sup>23</sup>.

**Опр. 3:**

**Перцепция<sub>i</sub>(X) =<sub>df</sub>**

**=<sub>df</sub> ПерцепцияПоследствий<sub>i</sub>(X) ∨ ПерцепцияДействий<sub>i</sub>(X) ∨**

**∨ ПерцепцияОбъектов<sub>i</sub>(X)**

«Агент *i* перцептивно воспринимает *X*, что по определению означает, что агент *i* перцептивно воспринимает либо последствие события, либо действие, либо объект».

<sup>22</sup> Определения (3.4.) и (3.5.) в Steunebrink Bas. R. The Logical Structure of Emotions. — SIKS Dissertation Series № 2010-23 Printed by GVO drukkers & vormgevers B. V. j Ponsen & Looijen, Ede. 2010. 228 p.

<sup>23</sup> Определение (3.3.) в Steunebrink Bas. R. The Logical Structure of Emotions. — SIKS Dissertation Series № 2010-23 Printed by GVO drukkers & vormgevers B. V. j Ponsen & Looijen, Ede. 2010. 228 p.

2.14. Триггеры позитивной и негативной эмоций, соответствующие наиболее общим понятиям и стоящие в корне родо-видовой иерархии, определяются в терминах предикатов перцепции и оценки как конъюнкции следующего вида<sup>24</sup>.

**Опр. 4:**

**Позитив<sup>T<sub>i</sub></sup>(X) =<sub>df</sub> Перцепция<sub>i</sub>(X) ∧ Хорошо<sub>i</sub>(X)**

«Положительная эмоция включается у агента *i* относительно *X*, что по определению означает, что агент *i* перцептивно воспринимает *X* и оценивает *X* как «хорошо».

**Опр. 5:**

**Негатив<sup>T<sub>i</sub></sup>(X) =<sub>df</sub> Перцепция<sub>i</sub>(X) ∧ Плохо<sub>i</sub>(X)**

«Отрицательная эмоция включается у агента *i* относительно *X*, что по определению означает, что агент *i* перцептивно воспринимает *X* и оценивает *X* как «плохо».

2.15. Видовые триггеры специфицируются в системе определений **SDM** с помощью дополнительных предикатов, выполняющих роль отличительных признаков в дефиниенсе определений, конструируемых из 17 предикаторов, перечисленных выше. В качестве примеров приведем определения триггеров таких негативных эмоций, как недовольство, горе, страх и подтвердившийся страх<sup>25</sup>.

**Опр. 6:**

**Недовольство<sup>T<sub>i</sub></sup>(c) =<sub>df</sub> ПерцепцияПоследствий<sub>i</sub>(c) ∧ Нежелательно<sub>i</sub>(c)**

**Опр. 7:**

**Горе<sup>T<sub>i</sub></sup>(c) =<sub>df</sub> Недовольство<sup>T<sub>i</sub></sup>(c) ∧ ОжидаетсяСейчас(c)**

**Опр. 8:**

**Страх<sup>T<sub>i</sub></sup>(c) =<sub>df</sub> Недовольство<sup>T<sub>i</sub></sup>(c) ∧ Ожидается-в-Будущем<sub>i</sub>(c)**

**Опр. 9:**

**Подтвердившийся-страх<sup>T</sup>(c) =<sub>df</sub>**

**=<sub>df</sub> Горе<sup>T<sub>i</sub></sup>(c) ∧ Было-или-Есть-истинно(Страх<sup>T<sub>i</sub></sup>(c`)) ∧ Подтверждает<sub>i</sub>(c, c`)**

**Опр. 9** можно редуцировать за счет подстановок до следующего:

<sup>24</sup> Определения (3.1.) и (3.2.) в Steunebrink Bas. R. The Logical Structure of Emotions. — SIKS Dissertation Series № 2010-23 Printed by GVO drukkers & vormgevers B. V. j Ponsen & Looijen, Ede. 2010. 228 p.

<sup>25</sup> Определения (3.7), (3.15), (3.13), (3.25) из Steunebrink Bas. R. The Logical Structure of Emotions. — SIKS Dissertation Series № 2010-23 Printed by GVO drukkers & vormgevers B. V. j Ponsen & Looijen, Ede. 2010. 228 p.

**Опр. 10:**

**Подтвердившийся-страх** $T(c) =_{df}$

$=_{df}$  **ПерцепцияПоследствий** $_i(c) \wedge$  **Нежелательно** $_i(c) \wedge$  **ОжидаетсяСейчас** $(c) \wedge$   
 $\wedge$  **Было-или-Есть-истинно** (**ПерцепцияПоследствий** $_i(c') \wedge$  **Нежелательно** $_i(c') \wedge$   
**Ожидается-в-Будущем** $_i(c')$ )  $\wedge$  **Подтверждает** $_i(c, c')$ )

Такое определение по существу эксплицирует в терминах **SDM**-формализма **ОСС**-трактовку на основе парадигмы **BDI** эмоции подтвердившегося страха как недовольства подтверждением ожидания нежелательного события.

Аналогичные редукции можно осуществить и для определений триггеров недовольства, горя и страха, а также для триггеров прочих эмоций, не взятых нами в рассмотрение.

2.16. Легко увидеть, что в данном определении триггер эмоции подтвердившегося страха встречается в явном виде только в определяемой на основе парадигмы **BDI** части определения. Правая часть определения, то есть дефиниенс, в явном виде не содержит терминов триггеров. В принципе, термины, из которых оказываются построенными дефиниенсы таких редуцируемых (за счет систем подстановок) определений, можно было бы принять в качестве базисных в данном формализме, однако авторы **SDM** на этом не останавливаются, но развивают систему определений, продолжая процесс редукции теперь уже терминов, из которых построены определения, приводимые выше, до определений, в которых предикаты перцепции, предикаты оценок, а также ряд терминов, соответствующих парадигме **BDI**, определяются в терминах языка комбинированной докситической + динамической логики, тем самым уточняя и саму парадигму **BDI**.

Забегая вперед, отметим, что язык используемой авторами логики содержит в качестве исходных модальный оператор убеждения-полагания  $B_i$  («агент  $i$  убежден в том, что...»), дуальный ему оператор определяется через отрицание как  $\neg B_i \neg$  («агент  $i$  полагает, что...»). Для первого оператора подоператорное выражение — формула — семантически рассматривается как истинная во всех мирах (состояниях), эпистемически возможных с миром агента  $i$ , для его дуала подоператорное выражение рассматривается как формула, истинная по крайней мере в одном эпистемически возможном мире (состоянии). Предполагается возможность произвольной итерации эпистемических модальных операторов (возможно с различными агентными индексами).

В качестве базового модального оператора динамической логики используется оператор действия  $[i:\alpha]$  «после выполнения агентом  $i$  действия  $\alpha$  с необходимостью имеет место, что...» и его дуал  $\langle i:\alpha \rangle$  «после выполнения агентом  $i$  действия  $\alpha$ , возможно, имеет место, что...», определяемый как  $\neg[i:\alpha]\neg$ .

Синтаксис этих операторов немного отличается от использовавшегося ранее нами тем, что указание на агента в нашем варианте выглядело как нижний индекс, приписанный оператору<sup>26</sup>, а в **SDM** используются термины вида «i:α» или «i:π» для указания на пары «агент-действие» или «агент-план». Соответственно, определения модальных операторов действий доопределяются и на случай планов.

2.17. Как уже говорилось выше, язык синтаксического аспекта формализации **SDM** — это комбинированный язык, состоящий из двух подязыков, названных разработчиками соответственно **ZAPL** (**An Abstract Agent Programming Language** — абстрактный язык программирования агентов, то есть собственно язык динамической доклатической логики с эпистемическими модальными операторами и модальными операторами действий в качестве основных) и **KARO** — расширение **ZAPL** специально для синтаксической формализации парадигмы **BDI**. В **KARO** планы (образующие множество планов **Plans**) конструируются из действий (из множества действий **ACT**) и планов как их последовательная композиция.

Предикаты перцепции, конструируемые на основе предикаторов **ПерцепцияПоследствий**, **ПерцепцияДействий**, **ПерцепцияОбъектов**, **ПерцепцияСвязи**, а также предикаты, конструируемые из предикаторов **Ожидается-в-Будущем**, **ОжидаетсяСейчас**, **Было-или-Есть-так-что**, **Предполагает**, определяются авторами **SDM** на основе синтаксиса **ZAPL**.

Предикаты оценки (то есть конструируемые из предикаторов **Желательно**, **Нежелательно**, **ДостойноПохвалы**, **ЗаслуживаетПорицания**, **Привлекательно**, **Непривлекательно**), предикаты (не)подтверждения (на основе предикаторов **Подтверждает**, **Отвергает**) и предикат когнитивного объединения (на основе предикатора **КогнитивноОбъединено**) определяются авторами **SDM** в языке **KARO**.

2.18. Отдельно следует обсудить допущения, принимаемые в формализации **SDM** относительно темпоральных отношений и характеристик. Содержательно-интуитивные темпоральные допущения в конечном счете сводятся к тому, что синтаксис предназначен прежде всего для описания эпистемических и эмоциональных состояний агентов и для описания переходов от одних состояний к другим в результате выполняемых агентами действий. Длительности действий и состояний на данном этапе формализации не учитываются, поэтому темпоральность задается исключительно порядком выполняемых действий и, как следствие, порядком реализуемых в результате действий состояний. На структуру времени оказывают влияние такие допущения, как детерминированность/недетерминированность действий, их дискретный характер, а также допущение/недопущение распараллеливания действий. В общем случае, и Б. Штейнбринк неоднократно это подчеркивает, формализм **SDM** предполагает дискретную временную структуру, линейную в прошлое и ветвящуюся в будущее, для описания которой в синтаксис вводятся темпоральные операторы **Было-или-Есть-**

---

<sup>26</sup> См. Синицкий Д. А. Экспликация понятия «диспозиционный предикат» (логические подходы) // Vox. Философский журнал. Выпуск № 22. Июнь, 2017. <https://vox-journal.org/html/issues/358/398>.

**истинно, НепосредственноНакануне-истинно, Истинно-или-Будет-истинным, Будет-истинным-но-не-сейчас.**

В принципе, эти операторы являются достаточно стандартными операторами языка временной логики, частично взаимоопределимыми друг через друга<sup>27</sup>. Эти темпоральные выражения активно используются в синтаксисе определений различных выражений, включая триггеры эмоций, предикаты перцепции, оценочные предикаты и предикаты, связанные с парадигмой **BDI**.

2.19. Теперь проследим до конца синтаксис системы определений триггера эмоции подтвердившегося страха. В дефиниенсе определения **Опр. 10** можно найти темпоральный оператор **Было-или-Есть-истинно**, и, кроме него, там стоят такие выражения, как **ПерцепцияПоследствий<sub>i</sub>(с)**, **Нежелательно<sub>i</sub>(с)**, **ОжидаетсяСейчас(с)**, **Ожидается-в-Будущем<sub>i</sub>(с`)**, **Подтверждает<sub>i</sub>(с,с`)**. Посмотрим на то, как они, за исключением темпорального оператора, переопределяются в терминах **ЗAPL** и **KARO**.

Обратим внимание на то, что у авторов **SDM** в предикатных выражениях на аргументных местах стоят метапеременные **с** и **с`**, замещающие индивидуальные константы или индивидуальные переменные, относящиеся к сорту «последствия событий». Поэтому возникает вопрос относительно их интерпретации. Что значит «последствие событий»? Этот термин перекочевал в формализацию **SDM** из **OCC**, причем любопытно примечание, которое делает Б. Штейнбринк на странице, где он объясняет неформально со ссылкой на **OCC** этот термин: «На странице 18 [Ortony et al., 1988] **OCC** говорят, что «события — это просто конструкты людей о том, что происходит». Однако с вычислительной точки зрения мы бы сказали, что это далеко не просто!»<sup>28</sup>

Но в конечном счете последствия событий оказываются в **SDM** также событиями, которые агенты принимают за последствия первых. А это означает, что события описываются формулами языка. Отсюда двойственность интерпретации аргументов — либо это имена некоторых событий (единичные или общие — отсюда использование индивидуальных констант или переменных), либо описания событий, и тогда это — либо пропозиции, либо пропозициональные переменные. В первом случае перечисленные выше выражения могут интерпретироваться как предикатные выражения, во втором — как модальные темпорализированные операторы, аргументами которых служат формулы.

2.20. Допущения относительно формул. Предполагается, что заданы множество имен агентов **AGT** (**i, j** — метапеременные для имен агентов), множество элементарных действий **EACT** и множество атомарных предложений **ATM** (**p** — метапеременная для обозначения произвольных атомарных предложений). Множество действий **ACT** ограничено элементарными действиями и их последовательными композициями (**α**,

<sup>27</sup> См. определения (4.2) — (4.4.) из Steunebrink Bas. R. The Logical Structure of Emotions. — SIKS Dissertation Series № 2010-23 Printed by GVO drukkers & vormgevers B. V. j Ponsen & Looijen, Ede. 2010. 228 p.

<sup>28</sup> Steunebrink Bas. R. The Logical Structure of Emotions. — SIKS Dissertation Series № 2010-23 Printed by GVO drukkers & vormgevers B. V. j Ponsen & Looijen, Ede. 2010. P. 19.

$\beta$  — метапеременные для действий из АСТ):  $EACT \subseteq ACT$ , если  $\alpha \in ACT$  и  $\beta \in ACT$ , то  $\alpha; \beta \in ACT$  (здесь точка с запятой обозначает операцию последовательной композиции действий (планов)). Множество литералов есть множество атомарных предложений вместе с их отрицаниями:  $LIT = ATM \cup \{\neg p \mid p \in ATM\}$ . Множество формул  $FORM$  включает в себя множество  $LIT$ , и если  $\varphi \in FORM$  и  $\psi \in FORM$ , то  $\neg(\varphi) \in FORM$ ,  $(\varphi \wedge \psi) \in FORM$ ,  $(\varphi \vee \psi) \in FORM$ ,  $(\varphi \rightarrow \psi) \in FORM$ ,  $(\varphi \leftrightarrow \psi) \in FORM$ . Кроме того, если  $i \in AGT$ ,  $\alpha \in ACT$  и  $\varphi \in FORM$ , то  $\forall_i \varphi \in FORM$  и  $[i:\alpha]\varphi \in FORM$ . Принимается стандартное соглашение о пропуске скобок.  $\forall_i \varphi$  читается как «агент  $i$  убежден, что  $\varphi$  (истинно)».  $[i:\alpha]\varphi$  читается как «после выполнения действия  $\alpha$  агентом  $i$  (необходимо) имеет место  $\varphi$ ».  $\langle i:\alpha \rangle \varphi$  определяется как  $\neg[i:\alpha]\neg\varphi$  и читается как « $\varphi$  является возможным результатом выполнения действия  $\alpha$  агентом  $i$ ». Дуал для  $\forall_i \varphi$ , то есть  $\neg\forall_i \neg\varphi$  специально в языке не обозначается, но читается как «агент  $i$  полагает  $\varphi$  возможным».

С учетом допущений из п. 2.18 (см. выше), если  $\varphi \in FORM$ , то **Было-или-Есть-истинно**( $\varphi$ )  $\in FORM$ , **НепосредственноНакануне-истинно**( $\varphi$ )  $\in FORM$ , **Истинно-или-Будет-истинным**( $\varphi$ )  $\in FORM$ , **Будет-истинным-но-не-сейчас**( $\varphi$ )  $\in FORM$ . Кроме того,  $T \in FORM$ ,  $\perp \in FORM$ , где  $T$  и  $\perp$  — пропозициональные константы, обозначающие всегда истинную и всегда ложную формулы.

2.21. Как сокращения вводятся следующие выражения ( $\varphi$  — произвольная формула):

**Новость** ( $\varphi$ ) — « $\varphi$  не была истинна в предыдущем состоянии, но  $\varphi$  истинна в текущем состоянии» определяется как<sup>29</sup>:

**Опр. 11:**

**Новость** ( $\varphi$ ) =<sub>df</sub>  $\varphi \wedge \neg$  **НепосредственноНакануне-истинно** ( $\varphi$ )

**Опр. 12:**

**ОбновлениеУбеждений** <sub>$i$</sub> ( $\varphi$ ) =<sub>df</sub> **Новость**( $\forall_i \varphi$ ) =<sub>df</sub>

=<sub>df</sub>  $\forall_i \varphi \wedge \neg$  **НепосредственноНакануне-истинно** ( $\forall_i \varphi$ )

Вспомогательное выражение **Будет-истинным-но-не-сейчас**( $\varphi$ ) читается как «когда-нибудь в будущем, но не сейчас  $\varphi$  может быть истинным» и определяется следующим образом<sup>30</sup>.

**Опр. 13:**

**Будет-истинным-но-не-сейчас**( $\varphi$ ) =<sub>df</sub>  $\neg\varphi \wedge$  **Истинно-или-Будет-истинным**( $\varphi$ )

<sup>29</sup> В соответствии с (4.11) из Steunebrink Bas. R. The Logical Structure of Emotions. — SIKS Dissertation Series № 2010-23 Printed by GVO drukkers & vormgevers B. V. j Ponsen & Looijen, Ede. 2010. 228 p.

<sup>30</sup> В соответствии с (4.1) из Steunebrink Bas. R. The Logical Structure of Emotions. — SIKS Dissertation Series № 2010-23 Printed by GVO drukkers & vormgevers B. V. j Ponsen & Looijen, Ede. 2010. 228 p.

На основе **Опр. 13** вводится выражение **БудущееОбновлениеУбеждений<sub>i</sub>(φ)**, которое прочитывается как «агент *i* пришел к убеждению: φ — это неправда сейчас, но существует будущее, в котором φ будет истинно»:

**Опр. 14:**

**БудущееОбновлениеУбеждений<sub>i</sub>(φ) =<sub>df</sub>**

**=<sub>df</sub> Новость (В<sub>i</sub> Будет-истинным-но-не-сейчас(φ)) =<sub>df</sub>**

**=<sub>df</sub> (В<sub>i</sub> Будет-истинным-но-не-сейчас(φ)) ∧**

**∧ ¬ НепосредственноНакануне-истинно (В<sub>i</sub> Будет-истинным-но-не-сейчас(φ))**

Еще одно эпистемическое выражение **НеопределенноеОбновлениеУбеждений<sub>i</sub>(φ)** в соответствии с (4.18) из работы Б. Штейнбринка<sup>31</sup> вводится как сокращение следующим образом:

**Опр. 15:**

**НеопределенноеОбновлениеУбеждений<sub>i</sub>(φ) =<sub>df</sub>**

**=<sub>df</sub> Новость (¬В<sub>i</sub>φ ∧ ¬В<sub>i</sub>¬φ) =<sub>df</sub>**

**=<sub>df</sub> (¬В<sub>i</sub>φ ∧ ¬В<sub>i</sub>¬φ) ∧ ¬ НепосредственноНакануне-истинно (¬В<sub>i</sub>φ ∧ ¬В<sub>i</sub>¬φ)**

Наконец, выражения **Ожидается-в-Будущем<sub>i</sub>(φ)** и **ОжидаетсяСейчас(φ)**, находящиеся в дефиниенсе определения подтвержденного страха (см. **Опр. 10** выше), вводятся с помощью следующих определений:

**Опр. 16:**

**Ожидается-в-Будущем<sub>i</sub>(φ) =<sub>df</sub>**

**=<sub>df</sub> БудущееОбновлениеУбеждений<sub>i</sub>(φ) ∨**

**∨ НеопределенноеОбновлениеУбеждений<sub>i</sub>(φ) =<sub>df</sub>**

**=<sub>df</sub> ((В<sub>i</sub> Будет-истинным-но-не-сейчас(φ)) ∧**

**∧ ¬ НепосредственноНакануне-истинно (В<sub>i</sub> Будет-истинным-но-не-сейчас(φ))) ∨**

**∨ ((¬В<sub>i</sub>φ ∧ ¬В<sub>i</sub>¬φ) ∧ ¬ НепосредственноНакануне-истинно (¬В<sub>i</sub>φ ∧ ¬В<sub>i</sub>¬φ))**

**ОжидаетсяСейчас(φ) =<sub>df</sub>**

**=<sub>df</sub> ОбновлениеУбеждений<sub>i</sub>(φ) =<sub>df</sub>**

<sup>31</sup> Steunebrink Bas. R. The Logical Structure of Emotions. — SIKS Dissertation Series № 2010-23 Printed by GVO drukkers & vormgevers B. V. j Ponsen & Looijen, Ede. 2010. 228 p.

=<sub>df</sub>  $B_i\varphi \wedge \neg$  **НепосредственноНакануне-истинно** ( $B_i\varphi$ )

Как пишет Б. Штейнбринк, «**Ожидается-в-Будущем**<sub>*i*</sub>( $\varphi$ ) и **ОжидаетсяСейчас**<sub>*i*</sub>( $\varphi$ ) теперь определены так, что они охватывают оба случая для перцептивного восприятия последствий событий, которые различаются в модели ОСС. Это означает, что **ПерцепцияПоследствий**<sub>*i*</sub>( $\varphi$ ), которую мы использовали для выражения того, что агент *i* перцептивно воспринимает  $\varphi$  как следствие (какого-то) события, может быть определена как дизъюнкция **Ожидается-в-Будущем**<sub>*i*</sub>( $\varphi$ ) и **ОжидаетсяСейчас**<sub>*i*</sub>( $\varphi$ )»<sup>32</sup>.

Соответственно, имеем следующее определение **ПерцепцияПоследствий**<sub>*i*</sub>( $\varphi$ ):

**Опр. 18:**

**ПерцепцияПоследствий**<sub>*i*</sub>( $\varphi$ ) =<sub>df</sub>

=<sub>df</sub> **Ожидается-в-Будущем**<sub>*i*</sub>( $\varphi$ )  $\vee$  **ОжидаетсяСейчас**<sub>*i*</sub>( $\varphi$ ) =<sub>df</sub>

=<sub>df</sub> ((( $B_i$  **Будет-истинным-но-не-сейчас**( $\varphi$ ))  $\wedge$

$\wedge \neg$  **НепосредственноНакануне-истинно** ( $B_i$  **Будет-истинным-но-не-сейчас**( $\varphi$ )))  $\vee$

$\vee$  (( $\neg B_i\varphi \wedge \neg B_i\neg\varphi$ )  $\wedge \neg$  **НепосредственноНакануне-истинно** ( $\neg B_i\varphi \wedge \neg B_i\neg\varphi$ )))

$\vee$

$\vee$  ( $B_i\varphi \wedge \neg$  **НепосредственноНакануне-истинно** ( $B_i\varphi$ ))

2.22. Теперь, если выполнить до конца последовательность подстановок из дефиниенсов определений, получаем в качестве определения подтвердившегося страха следующее выражение:

**Опр. 19:**

**Подтвердившийся-страх**<sup>T</sup>( $c$ ) =<sub>df</sub>

=<sub>df</sub> ((( $B_i$  **Будет-истинным-но-не-сейчас**( $c$ ))  $\wedge$

$\wedge \neg$  **НепосредственноНакануне-истинно** ( $B_i$  **Будет-истинным-но-не-сейчас**( $c$ )))  $\vee$

$\vee$  (( $\neg B_i c \wedge \neg B_i\neg c$ )  $\wedge \neg$  **НепосредственноНакануне-истинно** ( $\neg B_i c \wedge \neg B_i\neg c$ )))  $\vee$

$\vee$  ( $B_i c \wedge \neg$  **НепосредственноНакануне-истинно** ( $B_i c$ )))  $\wedge$

$\wedge$  **Нежелательно**<sub>*i*</sub>( $c$ )  $\wedge$  ( $B_i c \wedge \neg$  **НепосредственноНакануне-истинно** ( $B_i c$ ))  $\wedge$

$\wedge$  **Было-или-Есть-истинно**

(( $B_i$  **Будет-истинным-но-не-сейчас**( $c$ )))  $\wedge$

<sup>32</sup> Steunebrink Bas. R. The Logical Structure of Emotions. — SIKS Dissertation Series № 2010-23 Printed by GVO drukkers & vormgevers B. V. j Ponsen & Looijen, Ede. 2010. P. 65.

$$\begin{aligned} & \wedge \neg \text{НепосредственноНакануне-истинно} \quad (B_i \quad \text{Будет-истинным-но-не-} \\ & \text{сейчас}(c`))) \vee \\ & \vee ((\neg B_i c` \wedge \neg B_i \neg c`) \wedge \neg \text{НепосредственноНакануне-истинно}(\neg B_i c` \wedge \neg B_i \neg \\ & c`))) \vee \\ & \vee (B_i c` \wedge \neg \text{НепосредственноНакануне-истинно} (B_i c`)) \wedge \\ & \wedge \text{Нежелательно}_i(c`) \wedge \\ & \wedge ((B_i \text{Будет-истинным-но-не-сейчас}(c`)) \wedge \\ & \wedge \neg \text{НепосредственноНакануне-истинно} \quad (B_i \quad \text{Будет-истинным-но-не-} \\ & \text{сейчас}(c`))) \vee \\ & \vee ((\neg B_i c` \wedge \neg B_i \neg c`) \wedge \neg \text{НепосредственноНакануне-истинно} (\neg B_i c` \wedge \neg B_i \neg \\ & c`)) \wedge \\ & \wedge \text{Подтверждает}_i(c, c`) \end{aligned}$$

В таком определении в дефиниенсе осталось только два предикатных выражения, которые требуют своего доопределения. Это оценочный предикат **Нежелательно<sub>i</sub>(X)** и двуместный предикат подтверждения одного последствия другим **Подтверждает<sub>i</sub>(X, Y)**. Для их доопределения нам понадобится рассмотреть используемый авторами **SDM** формализм **KARO**, что, естественно, требует анализа некоторых из явных и неявных допущений, лежащих в основаниях теперь уже этого формализма.

### 3. Формализация целеполагания и мотивации

*Но возбуждения предполагают мотивы, или, скорее, они предполагают общие тенденции поведения, мотивы которых представляют для нас наибольший интерес.*

*Гилберт Райл. Понятие сознания<sup>33</sup>*

3.1. Как пишет Б. Штейнбринк, «каркас **KARO** представляет собой смесь динамической логики, эпистемической/доксастической логики и нескольких дополнительных (модальных) операторов для работы с мотивационными аспектами искусственных агентов»<sup>34</sup> (выделено курсивом нами).

3.2. Одним из базовых понятий в **KARO** является понятие цели. Для описания собственно целей используется подмножество множества формул, которое получается из множества формул за счет ряда ограничительных условий, которые мы также отнесем к допущениям, лежащим в основе **SDM**.

<sup>33</sup> Райл Г. Понятие сознания. Перевод с англ. — М.: Идея-Пресс, Дом интеллектуальной книги, 1999. — С. 104.

<sup>34</sup> Steunebrink Bas. R. The Logical Structure of Emotions. — SIKS Dissertation Series № 2010-23 Printed by GVO drukkers & vormgevers B. V. j Ponsen & Looijen, Ede. 2010. P. 81.

А именно — для описания целей используются формулы, являющиеся непустыми непротиворечивыми конъюнкциями литералов. По сути, отождествление целей с формулами такого логического вида позволяет авторам, во-первых, легко структурировать цели, как состоящие из подцелей (как подконъюнкций), во-вторых, не будем забывать о том, что язык **KARO** разрабатывался для формализации целеполагания искусственных агентов, для которых характерна последовательная обработка целей и подцелей.

Вслед за **SDM** обозначим множество целей (множество непротиворечивых конъюнкций литералов) — **ccl**. На этом множестве задаются два бинарных отношения:  $\mathbf{sub} \subseteq \mathbf{ccl}^2$ ,  $\mathbf{inv} \subseteq \mathbf{ccl}^2$ . Первое означает быть подцелью какой-либо цели, второе — быть инверсией какой-либо цели. Инверсированная цель получается заменой каждой атомарной формулы в конъюнкции литералов на ее отрицание и заменой каждого отрицания атомарной формулы на саму атомарную формулу.

3.3. Язык **KARO** надстраивается над языком **ZAPL** за счет добавления модальных операторов **G**, **A**, **Com** и соответствующего доопределения понятия правильно построенной формулы. Если  $\varphi \in \mathbf{ccl}$ ,  $\pi$  есть план<sup>35</sup>, то  $\mathbf{G}_i(\varphi)$ ,  $\mathbf{A}_i(\pi)$ ,  $\mathbf{Com}_i(\pi)$  — правильно построенные формулы. Выражения  $\mathbf{G}_i(\varphi)$ ,  $\mathbf{A}_i(\pi)$ ,  $\mathbf{Com}_i(\pi)$  читаются как «агент  $i$  имеет провозглашенной целью  $\varphi$ », «агент  $i$  способен выполнить план  $\pi$ », «агент  $i$  намерен выполнять  $\pi$ » (или, другими словами, «план  $\pi$  включен в повестку дня агента  $i$ »).

3.4. Очень важное допущение, лежащее в основе формализации **SDM**, заключается в принятии процедурной интерпретации, связывающей понятие убеждения и понятие достижения цели. Если  $\varphi$  — формула, описывающая цель, то «говорят, что цель была достигнута, когда агент  $i$  убежден, что  $\varphi$  выполняется, т. е. когда истинно  $\mathbf{B}_i\varphi$ »<sup>36</sup>.

Для оперирования с целями вводится набор специальных операторов, например, оператор удаления цели **drop**( $\varphi$ ), однако мы их не будем специально анализировать, просто отсылая читателя к источнику<sup>37</sup>.

Кроме того, вводятся синтаксические аналоги отношений  $\mathbf{sub} \subseteq \mathbf{ccl}^2$  и  $\mathbf{inv} \subseteq \mathbf{ccl}^2$ , определенных выше: если  $\varphi \in \mathbf{ccl}$ ,  $\psi \in \mathbf{ccl}$ , то  $\varphi \subseteq \psi$  есть формула (читается — « $\varphi$  является (логической) частью  $\psi$ ») и  $\varphi \ominus \psi$  есть формула (читается — « $\varphi$  есть инверсная конъюнкция  $\psi$ »). Мы (в отличие от Б. Штейнбринка) используем знак  $\subseteq$  в расчете на то, что у читателя не возникнет путаницы его прочтения указанным способом,

<sup>35</sup> Напомним, что в **KARO** планы конструируются из действий и планов как их последовательная композиция.

<sup>36</sup> Steunebrink Bas. R. The Logical Structure of Emotions. — SIKS Dissertation Series № 2010-23 Printed by GVO drukkers & vormgevers B. V. j Ponsen & Looijen, Ede. 2010. P. 83.

<sup>37</sup> Steunebrink Bas. R. The Logical Structure of Emotions. — SIKS Dissertation Series № 2010-23 Printed by GVO drukkers & vormgevers B. V. j Ponsen & Looijen, Ede. 2010. 228 p.

в отличие от обозначения теоретико-множественного отношения нестроого включения, в зависимости от контекста.

Б. Штейнбринк семантически определяет эти синтаксические аналоги следующим образом:

$$\models \varphi \subseteq \psi \text{ если и только если } \langle \varphi, \psi \rangle \in \mathbf{sub};$$

$$\models \varphi \ominus \psi \text{ если и только если } \langle \varphi, \psi \rangle \in \mathbf{inv}.$$

Здесь знак  $\models$  у авторов **SDM** обозначает классическую общезначимость формул, не требующую для определения крипковской семантики.

Отметим еще раз в качестве принимаемого допущения достаточно сильное ограничение на формулы, являющиеся подоператорными выражениями в синтаксических конструкциях, рассматриваемых нами в следующем пункте. Они являются целями, то есть элементами множества **ccl**.

3.5. Язык **KARO** позволяет авторам определить с помощью модальных операторов **G**, **A**, **Com** целый ряд производных мотивационных предикатов, таких как **ПрактВозм<sub>i</sub>( $\pi, \varphi$ )** — «агент  $i$  имеет практическую возможность выполнить действие/план  $\pi$ , чтобы добиться  $\varphi$ », **Может<sub>i</sub>( $\pi, \varphi$ )** — «агент  $i$  может выполнить  $\pi$  для осуществления  $\varphi$ », **ЦельДостижения<sub>i</sub>( $\pi, \varphi$ )** — «у агента  $i$  есть цель достичь  $\varphi$ », **ВозмНамерение<sub>i</sub>( $\pi, \varphi$ )** — «у агента  $i$  есть возможное намерение выполнить  $\pi$  с целью достичь  $\varphi$ ». Ниже приведены их определения (5.1) — (5.4) из работы Б. Штейнбринка<sup>38</sup>, расширенные нами соответствующими редукциями дефиниенсов до выражений, содержащих только базовые операторы **KARO**:

**Опр. 20** соответствует определению (5.1):

$$\mathbf{ПрактВозм}_i(\pi, \varphi) =_{df} \mathbf{A}_i \pi \wedge \langle i: \pi \rangle \varphi$$

**Опр. 21** соответствует определению (5.2):

$$\mathbf{Может}_i(\pi, \varphi) =_{df} \mathbf{B}_i \mathbf{ПрактВозм}_i(\pi, \varphi) =_{df} \mathbf{B}_i \mathbf{A}_i \pi \wedge \langle i: \pi \rangle \varphi$$

**Опр. 22** соответствует определению (5.3):

$$\mathbf{ЦельДостижения}_i(\pi, \varphi) =_{df} \mathbf{B}_i (\neg \varphi \wedge \mathbf{G}_i \varphi)$$

Следующее **Опр. 23** соответствует определению Б. Штейнбринка (5.4), редуцированному нами до выражений, содержащих только операторы **3APL** и **KARO**:

**Опр. 23:**

$$\mathbf{ВозмНамерение}_i(\pi, \varphi) =_{df}$$

$$=_{df} \mathbf{Может}_i(\pi, \varphi) \wedge \mathbf{ЦельДостижения}_i(\pi, \varphi) =_{df}$$

<sup>38</sup> Steunebrink Bas. R. The Logical Structure of Emotions. — SIKS Dissertation Series № 2010-23 Printed by GVO drukkers & vormgevers B.V. j Ponsen & Looijen, Ede. 2010. 228 p.

$$=_{df} (B_i A_i \pi \wedge \langle i: \pi \rangle \varphi) \wedge B_i (\neg \varphi \wedge G_i \varphi)$$

3.6. Определения (5.17) и (5.18) Б. Штейнбринка позволяют редуцировать предикаты **Подтверждает**<sub>i</sub>( $\varphi$ ,  $\psi$ ) и **Отвергает**<sub>i</sub>( $\varphi$ ,  $\psi$ ) до выражений, содержащих эпистемический оператор убеждения-полагания и связки  $\subseteq$  и  $\ominus$ , являющиеся синтаксическими аналогами отношений **sub** и **inv**<sup>39</sup>:

**Опр. 24:**

$$\text{Подтверждает}_i(\varphi, \psi) =_{df} B_i (\varphi \subseteq \psi)$$

(Целевая формула  $\varphi$  подтверждает целевую формулу  $\psi$  для агента  $i$ , если и только если агент  $i$  убежден, что  $\varphi$  есть подцель.)

**Опр. 25:**

$$\text{Отвергает}_i(\varphi, \psi) =_{df} B_i ((\varphi \ominus \psi) \wedge (\varphi \subseteq \psi))$$

(Целевая формула  $\varphi$  отвергает целевую формулу  $\psi$  для агента  $i$ , если агент  $i$  убежден, что  $\varphi$  является инвертированной подцелью  $\psi$ .)

Мы предоставляем читателю возможность самостоятельно переписать **Опр. 19.**, заменив выражение **Подтверждает**<sub>i</sub>( $c$ ,  $c'$ ) на  $B_i(c \subseteq c')$ .

3.7. Для доопределения оценочных предикатов, в частности таких, как **Желательно**<sub>i</sub>( $X$ ) и **Нежелательно**<sub>i</sub>( $X$ ), авторы **SDM** используют понятия, относящиеся к семантике формализма, поэтому настало время показать, как используется семантика возможных миров для построения основных семантических конструкций авторами **SDM**.

#### 4. Семантические допущения и конструкции формализма SDM

*Положение о том, что способность оценить действие однотипно со способностью его исполнить, иллюстрирует только что доказанное утверждение, а именно то, что разумные способности являются не сингулярными диспозициями, а диспозициями, допускающими большое разнообразие более или менее непохожих практик.*

*Гилберт Райл. Понятие сознания*<sup>40</sup>

Как мы уже отмечали выше, семантика формализма является двухслойной, или двухуровневой. Первый уровень образуют модели для эпистемических операторов убеждения-полагания, второй уровень образует трансформационная семантика моделей для действий, изменяющих модели убеждения-полагания.

<sup>39</sup> Steunebrink Bas. R. The Logical Structure of Emotions. — SIKS Dissertation Series № 2010-23 Printed by GVO drukkers & vormgevers B. V. j Ponsen & Looijen, Ede. 2010. 228 p.

<sup>40</sup> Райл Г. Понятие сознания. Перевод с англ. — М.: Идея-Пресс, Дом интеллектуальной книги, 1999. С. 64.

Модели убеждений определяются стандартным для семантики эпистемической логики образом и имеют вид  $M = \langle S, R, V \rangle$ , где  $S$  есть непустое множество состояний (или «возможных миров»),  $R = \{R_i \mid i \in AGT\}$  есть множество отношений достижимости на  $S$ , по одному для каждого агента  $i$ , причем  $R_i \subseteq S^2$ , и  $V: S \rightarrow 2^{ATM}$  есть функция оценки.

Функция оценки указывает, какие из атомарных предложений имеют место в каждом состоянии. Каждое отношение  $R_i$  достижимости из  $R$  сериально, транзитивно и евклидово.

Как пишет Б. Штейнбринк, каждая модель подобного вида репрезентирует убеждения агентов в определенный момент времени. Динамика появляется тогда, когда агенты «могут «передумать», выполняя действия»<sup>41</sup>. И далее: «Мы можем моделировать действия, используя семантику возможных миров, но с моделями полагания в качестве возможных миров»<sup>42</sup>. В соответствии с этим подходом, модели действий в **SDM** определяются как реляционные системы следующего вида:

$M = \langle S, R, Aux, Emo \rangle$ , где  $S$  есть непустое множество возможных пар модель-состояние, соответственно, модель есть  $M = \langle S, R, V \rangle$ , как определено выше, а состояние есть состояние из множества состояний  $S$  в этой модели, то есть если  $\langle M, s \rangle \in S$  и  $M = \langle S, R, V \rangle$ , то  $s \in S$ .

$R = \{R_{i:\alpha} \mid i \in AGT, \alpha \in ACT\}$  есть множество отношений достижимости на  $S$ , соответствующих переходам от одной модели убеждения к другой для каждого агента, выполняющего соответствующее действие.

Также задается структура  $Aux = \langle Goals, Caps, Agd \rangle$  вспомогательных функций, которые указывают для каждого агента и каждой пары модель-состояние, какие цели (*Goals*), способности (*Caps*) и обязательства (*Agd*) имеет агент. Их теоретико-множественное представление следующее:  $Goals: AGT \times S \rightarrow 2^{cd}$ ,  $Caps: AGT \times S \rightarrow 2^{Plans}$ ,  $Agd: AGT \times S \rightarrow 2^{Plans}$ , то есть *Goals* есть функция, возвращающая набор целей, которые агент имеет для каждой пары модель-состояние, *Caps* есть функция, которая возвращает набор действий, которые агент способен выполнить для каждой пары модель-состояние, и *Agd* является функцией, которая возвращает набор действий, совершенных агентом (которые находятся в его, по выражению авторов **SDM**, «повестке дня») для каждой пары «модель — состояние».

$Emo = \langle Des, Undes, Praisew, Blamew, Appeal, Unappeal, CogUnit \rangle$  является набором функций, служащих для семантического определения оценочных предикатов и предиката когнитивного объединения, которые синтаксически строятся с помощью предикаторов **Желательно**, **Нежелательно**, **Достойно** **Похвалы**,

<sup>41</sup> Steunebrink Bas. R. The Logical Structure of Emotions. — SIKS Dissertation Series № 2010-23 Printed by GVO drukkers & vormgevers B. V. j Ponsen & Looijen, Ede. 2010. P. 85.

<sup>42</sup> Steunebrink Bas. R. The Logical Structure of Emotions. — SIKS Dissertation Series № 2010-23 Printed by GVO drukkers & vormgevers B. V. j Ponsen & Looijen, Ede. 2010. P. 85.

**ЗаслуживаетПорицания,  
КогнитивноОбъединено.**

**Привлекательно,**

**Непривлекательно,**

Эти семантические функции имеют следующее теоретико-множественное представление:  $Des:AGD \times S \rightarrow 2^{FORM}$ ,  $Undes:AGD \times S \rightarrow 2^{FORM}$ ,  $Praisew:AGD \times S \rightarrow 2^{AGT \times ACT}$ ,  $Blamew:AGD \times S \rightarrow 2^{AGT \times ACT}$ ,  $Appeal:AGD \times S \rightarrow 2^{OBJ}$ ,  $Unappeal:AGD \times S \rightarrow 2^{OBJ}$ ,  $CogUnit:AGD \times S \rightarrow 2^{AGT}$  (где FORM есть множество правильно построенных формул, OBJ — множество объектов, существование которых просто постулируется в семантике авторами формализма **SDM**, причем одно из онтологических допущений заключается в том, что все агенты являются объектами из  $AGD \subseteq OBJ$ ).

Пусть заданы структуры **M** и **M**. Формулы интерпретируются на произвольном состоянии *s* модели **M** с учетом того, что  $\langle M, s \rangle \in S$ , где *S* есть элемент структуры **M**.

**Опр. 26:**

1.  $\langle M \langle M, s \rangle \rangle \models p$  если и только если  $p \in V(s)$  и  $p \in ATM$ .
2.  $\langle M \langle M, s \rangle \rangle \models \neg \phi$  если и только если неверно, что  $\langle M \langle M, s \rangle \rangle \models \phi$ .
3.  $\langle M \langle M, s \rangle \rangle \models \phi \wedge \psi$  если и только если  $\langle M \langle M, s \rangle \rangle \models \phi$  и  $\langle M \langle M, s \rangle \rangle \models \psi$ .
4.  $\langle M \langle M, s \rangle \rangle \models \phi \vee \psi$  если и только если  $\langle M \langle M, s \rangle \rangle \models \phi$  или  $\langle M \langle M, s \rangle \rangle \models \psi$ .
5.  $\langle M \langle M, s \rangle \rangle \models \phi \rightarrow \psi$  если и только если или  $\langle M \langle M, s \rangle \rangle \models \neg \phi$  или  $\langle M \langle M, s \rangle \rangle \models \psi$ .
6.  $\langle M \langle M, s \rangle \rangle \models \phi \leftrightarrow \psi$  если и только если  $\langle M \langle M, s \rangle \rangle \models \phi \rightarrow \psi$  и  $\langle M \langle M, s \rangle \rangle \models \psi \rightarrow \phi$ .
7.  $\langle M \langle M, s \rangle \rangle \models \forall i \phi$  если и только если для любого  $s' \in S$ , такого, что  $\langle s, s' \rangle \in R_i$ , имеет место  $\langle M \langle M, s' \rangle \rangle \models \phi$ .
8.  $\langle M \langle M, s \rangle \rangle \models G_i \phi$  если и только если  $\phi \in Goals(i, \langle M, s \rangle)$ .
9.  $\langle M \langle M, s \rangle \rangle \models A_i \pi$  если и только если  $\pi \in Caps(i, \langle M, s \rangle)$ .
10.  $\langle M \langle M, s \rangle \rangle \models Com_i \pi$  если и только если  $\pi \in Agd(i, \langle M, s \rangle)$ .
11.  $\langle M \langle M, s \rangle \rangle \models \langle i: \pi \rangle \phi$  если и только если найдется по крайней мере одна пара  $\langle M', s' \rangle \in S$ , такая, что имеет место  $\langle M \langle M', s' \rangle \rangle \models \phi$  и  $\langle \langle M, s \rangle, \langle M', s' \rangle \rangle \in R_i: \pi \in R$ .
12.  $\langle M \langle M, s \rangle \rangle \models [i: \pi] \phi$  если и только если для любой пары  $\langle M', s' \rangle \in S$  имеет место  $\langle M \langle M', s' \rangle \rangle \models \phi$  и  $\langle \langle M, s \rangle, \langle M', s' \rangle \rangle \in R_i: \pi \in R$ .
13.  $\langle M \langle M, s \rangle \rangle \models$  **Истинно-или-Будет-истинным**( $\phi$ ) если и только если найдется по крайней мере одна пара  $\langle M', s' \rangle \in S$ , такая, что имеет место  $\langle M \langle M', s' \rangle \rangle \models \phi$  и  $\langle \langle M, s \rangle, \langle M', s' \rangle \rangle \in (\cup R)^*$ .
14.  $\langle M \langle M, s \rangle \rangle \models$  **Было-или-Есть-истинно**( $\phi$ ) если и только если найдется по крайней мере одна пара  $\langle M', s' \rangle \in S$ , такая, что имеет место  $\langle M \langle M', s' \rangle \rangle \models \phi$  и  $\langle \langle M', s' \rangle, \langle M, s \rangle \rangle \in (\cup R)^*$ .

15.  $\langle M \langle M, s \rangle \rangle \models \text{НепосредственноНакануне-истинно}(\varphi)$  если и только если найдется по крайней мере одна пара  $\langle M', s' \rangle \in S$ , такая, что имеет место  $\langle M \langle M', s' \rangle \rangle \models \varphi$  и  $\langle \langle M', s' \rangle, \langle M, s \rangle \rangle \in (\cup R)$ .
16.  $\langle M \langle M, s \rangle \rangle \models \text{Желательно}_i(\varphi)$  если и только если  $\varphi \in Des(i, \langle M, s \rangle)$ .
17.  $\langle M \langle M, s \rangle \rangle \models \text{Нежелательно}_i(\varphi)$  если и только если  $\varphi \in Undes(i, \langle M, s \rangle)$ .
18.  $\langle M \langle M, s \rangle \rangle \models \text{ДостойноПохвалы}_i(j:\alpha)$  если и только если  $\langle j, \alpha \rangle \in Praisew(i, \langle M, s \rangle)$ .
19.  $\langle M \langle M, s \rangle \rangle \models \text{ЗаслуживаетПорицания}_i(j:\alpha)$  если и только если  $\langle j, \alpha \rangle \in Blamew(i, \langle M, s \rangle)$ .
20.  $\langle M \langle M, s \rangle \rangle \models \text{Привлекательно}_i(x)$  если и только если  $x \in Appeal(i, \langle M, s \rangle)$ .
21.  $\langle M \langle M, s \rangle \rangle \models \text{Непривлекательно}_i(x)$  если и только если  $x \in Unappeal(i, \langle M, s \rangle)$ .
22.  $\langle M \langle M, s \rangle \rangle \models \text{КогнитивноОбъединено}_i(j)$  если и только если  $j \in CogUnit(i, \langle M, s \rangle)$ .

Эти семантические условия интерпретации формул формализма в семантике **KARO** полностью приводятся в разделе 5.3 диссертации Б. Штейнбринка<sup>43</sup>. Истинность в таком состоянии  $s$  произвольной формулы  $\varphi$  обозначается авторами **SDM** как  $\langle M \langle M, s \rangle \rangle \models \varphi$ <sup>44</sup>. Мы добавили в **Опр. 26** пп. 4, 5, 6, в отличие от семантических определений у Б. Штейнбринка, который вводит дизъюнкции, импликации и эквиваленции как сокращения, нами также добавлен и п. 12. для дуала оператора  $\langle \dots \rangle$ .

$\cup R$  есть отношение, являющееся результатом теоретико-множественного объединения всех отношений из **R**. Отсюда ясно, что бинарные отношения достижимости из **R** трактуются в данной семантике экстенционально, то есть как множества пар, являющиеся подмножествами множества  $S^2$ . Аналогично,  $(\cup R)^*$  рассматривается авторами **SDM** как бинарное отношение на  $S$ , являющееся рефлексивным транзитивным замыканием  $\cup R$ .

Как видно из пунктов 16–22 **Опр. 27**, существование семантических значений (областей истинности) оценочных предикатов просто постулируется в семантике языка формализма с помощью постулирования структуры функций **Емо** и использования семантических функций из **Емо** в указанных пунктах определения.

Но, как отмечает Б. Штейнбринк в конце раздела 5.3 своей диссертации, «однако остается открытым вопрос, каким образом множества в **Емо** (*Des* и т. д.) получают свои элементы»<sup>45</sup>. Мы рассмотрим, как этот вопрос решает Б. Штейнбринк для оценочного предиката **Нежелательно**<sub>*i*</sub>( $\varphi$ ), поскольку именно данный предикат входит

<sup>43</sup> Steunebrink Bas. R. The Logical Structure of Emotions. — SIKS Dissertation Series № 2010-23 Printed by GVO drukkers & vormgevers B. V. j Ponsen & Looijen, Ede. 2010. 228 p.

<sup>44</sup> Точнее говоря, Б. Штейнбринк подразумевает такую запись, но для упрощения сохраняет обозначение « $M, s \models$ ». В отличие него, мы используем для обозначения кортежей угловые скобки и сохраняем полный вид записи.

<sup>45</sup> Steunebrink Bas. R. The Logical Structure of Emotions. — SIKS Dissertation Series № 2010-23 Printed by GVO drukkers & vormgevers B. V. j Ponsen & Looijen, Ede. 2010. P. 89.

в дефиниенс определения триггера эмоции подтвержденного страха, который мы выбрали в качестве примера.

На самом деле, сформулировав в языке оценочные предикаты и постулировав существование соответствующих семантических функций (в интересующем нас случае — **Нежелательно**<sub>i</sub>(φ) и *Undes*), авторы **SDM** далее постулируют ряд ограничений на области определения введенных семантических функций. Так, для функции *Undes* вводится следующее ограничение<sup>46</sup>, означающее, что *Undes* содержит все инвертированные подцели:

**Ограничение 1:**  $inv \circ sub \circ Goals(i, \langle M, s \rangle) \subseteq Undes(i, \langle M, s \rangle)$ .

Здесь есть операция композиции как функций, так и бинарных отношений.

Такой подход Б. Штейнбринк оправдывает тем, что понятие достигаемой цели является только частью мотивационного аспекта, который в общем случае может быть связан и с другими понятиями, такими, например, как нормы, интересы, безопасность. Как пишет он сам: «Мы не будем определять подобные функции оценки; вместо этого мы будем ограничивать их таким образом, чтобы они содержали оценку только для агентов с целями достижения. Идея состоит в том, что можно просто добавить больше ограничений к этой оценочной функции, если структура обогащена большим количеством факторов»<sup>47</sup>. Таким образом, в основаниях **SDM**, связанных с трактовкой оценок, лежит очень сильное допущение, имеющее априорный характер.

## 5. Формализм SDM и диспозиционные предикаты

*Суждение, приписывающее предмету диспозициональное свойство, во многом, хотя и не во всем, сходно с суждением, подводящим предмет под действие закона. Обладать диспозициональным свойством не означает пребывать в определенном состоянии или претерпевать определенные изменения. Это значит быть готовым или быть обязанным принять определенное состояние или же претерпеть определенные изменения тогда, когда реализуется определенное условие, то же самое справедливо и в отношении особых диспозиций человека, таких, как качества его характера.*

*Гилберт Райл. Понятие сознания*<sup>48</sup>

---

<sup>46</sup> Ограничение (5.11) из Steunebrink Bas. R. The Logical Structure of Emotions. — SIKS Dissertation Series № 2010-23 Printed by GVO drukkers & vormgevers B. V. j Ponsen & Looijen, Ede. 2010. 228 p.

<sup>47</sup> Steunebrink Bas. R. The Logical Structure of Emotions. — SIKS Dissertation Series № 2010-23 Printed by GVO drukkers & vormgevers B. V. j Ponsen & Looijen, Ede. 2010. Pp. 89–90.

<sup>48</sup> Райл Г. Понятие сознания. Перевод с англ. — М.: Идея-Пресс, Дом интеллектуальной книги, 1999. — С. 52.

5.1. Так называемая *эмпирическая* интерпретация диспозиционных предикатов подробно обсуждалась нами наряду с другими подходами к проблеме логической экспликации диспозиционных предикатов ранее<sup>49</sup>, поэтому мы не будем на ней детально останавливаться, отметив только, что в рамках семантики, предложенной В. А. Смирновым, интерпретирующая функция сопоставляет каждой индивидуальной константе — объект из предметной области (универсума рассуждения), в то время как каждому предикатному символу интерпретирующая функция сопоставляет разрешимое свойство или отношение, а каждой функциональной константе — рекурсивную функцию. В основе такой трактовки терминов и предикатов наблюдения лежит очень ясная интуиция: «...для того, чтобы узнать, обладает ли объект **a** свойством **P**, необходимо проделать некоторую работу по применению алгорифма к этому объекту. Эмпирическое утверждение будет тогда просто отчетом о проделанной работе»<sup>50</sup>.

5.2. Развитие такого «динамического» подхода в уточнении диспозиционного предиката происходило на основе идеи, заключающейся в том, что иногда хотелось бы иметь синтаксические средства, позволяющие непосредственно в языке формализма (а не только в метаязыке на уровне семантического обсуждения и интерпретации) формулировать и выражать связи между утверждениями, носящими дескриптивный (декларативный, пикториальный) характер, и утверждениями о свойствах и структурах разрешающих процедур и алгорифмов (проверочных действий и инструкций). Или, как отмечал А. М. Анисов со ссылкой на В. А. Смирнова, «иногда интереснее сопоставлять предикатам наблюдения не сами функции, а именно конкретные инструкции, поскольку в противном случае мы абстрагируемся от технической стороны дела, которая в эмпирических исследованиях может быть существенной»<sup>51</sup>.

Эти идеи Смирнова В. А. позволили нам предложить<sup>52</sup> описывать редукционные предложения, выражающие связь между диспозиционными предикатами и предикатами наблюдения, в терминах динамической логики следующим образом:

**Опр. 27:**  $D(x) =_{df} [\pi]_i R(x),$

или (в рамках более широкой трактовки диспозиционности) сторонником которой был, например, Г. Райл:

**Опр. 28:**  $D(x) =_{df} \langle \pi \rangle_i R(x),$

<sup>49</sup> Синицкий Д. А. Экспликация понятия «диспозиционный предикат» (логические подходы) // Vox. Философский журнал. Выпуск № 22. Июнь, 2017. <https://vox-journal.org/html/issues/358/398>.

<sup>50</sup> Смирнов В. А. Логические методы анализа научного знания. — М.: «Наука», 1987. — С. 231.

<sup>51</sup> Анисов А. М. Концепция научной философии В. А. Смирнова // Философия науки. Вып. 2: Гносеологические и логико-методологические проблемы. — М.: ИФ РАН, 1996. — С. 21–22.

<sup>52</sup> См. Синицкий Д. А. Экспликация понятия «диспозиционный предикат» (логические подходы) // Vox. Философский журнал. Выпуск № 22. Июнь, 2017. <https://vox-journal.org/html/issues/358/398>, а также библиографию в этой статье.

где  $D(x)$  — диспозиционный предикат,  $\pi$  — последовательность проверочных действий, ассоциируемая с условиями наблюдения и выполняемая агентом  $i$ ,  $R(x)$  — предикат наблюдения (эмпирический предикат),  $[...]$  — оператор «после выполнения агентом действий ... с **необходимостью** имеет место, что...»,  $\langle \dots \rangle$  — дуальный ему оператор.

5.3. В рамках формализации теории эмоций **ОСС**, обозначенной нами выше как **SDM**, связь между эмпирическими предикатами и диспозиционными предикатами представлена как связь между так называемыми триггерами эмоций (то есть предикатами, характеризующими условия возникновения или «включения» эмоций) и предикатами переживания эмоций. Как было проиллюстрировано на примере с эмоцией подтвержденного страха, первые определяются авторами **SDM** терминах предикатов перцептивного восприятия и оценки, конструируемых в языках **3APL** и **KARO**. Таким образом, в модели **ОСС** исходными являются предикаты, связанные с перцептивным восприятием и первичной оценкой, которые в формализме **SDM** интерпретируются как действия агентов (перцептивные или ментальные), переводящие в конечном счете всю мультиагентную систему из одних состояний в другие, а предикаты переживания эмоций являются диспозициями, степень проявления которых (от нуля до некоторого порога) характеризуется не только качественно, но и через количественные показатели (например, общую интенсивность переживания эмоции).

5.4. Авторы **SDM**, и в частности, Б. Штейнбринк<sup>53</sup>, неоднократно отмечают важность проводимого ими различия между триггерами эмоций и предикатами переживания эмоций. Однако проведение такого различия требует и определения того, как эти концепты оказываются связанными друг с другом. В общем виде эта связь выражается следующей закономерностью: эмоция переживается тогда и только тогда, когда (1) эмоция была «включена» когда-то в прошлом и (2) общая интенсивность переживания эмоции положительна. Следующая схема определений<sup>54</sup> выражает эту связь в языке формализма **SDM**:

**Опр. 29:**  $\text{Эмоция}_i(o \rightarrow) =_{\text{df}} \text{Было-или-есть Эмоция}^T_i(o \rightarrow) \wedge \text{OFI}(\text{Эмоция}, i, o \rightarrow)$

Здесь **Эмоция**, **Эмоция<sup>T</sup>** есть метапеременные для предикаторов переживания эмоций и предикаторов триггеров эмоций,  $o \rightarrow$  есть метапеременная, называемая Б. Штейнбринком «вектором аргументов» для соответствующих предикаторов, **OFI** — предикатор, с помощью которого конструируются предикаты общей интенсивности переживания вида **OFI(Эмоция, i, o<sup>→</sup>)**. Как видно из схемы определения, одним из аргументов **OFI** является какой-либо из предикаторов переживания эмоций, что делает

<sup>53</sup> Steunebrink Bas. R. The Logical Structure of Emotions. — SIKS Dissertation Series № 2010-23 Printed by GVO drukkers & vormgevers B. V. j Ponsen & Looijen, Ede. 2010. P. 137.

<sup>54</sup> Соответствующая схеме определений (6.30) из Steunebrink Bas. R. The Logical Structure of Emotions. — SIKS Dissertation Series № 2010-23 Printed by GVO drukkers & vormgevers B. V. j Ponsen & Looijen, Ede. 2010. P. 128.

выражение **OFI(Эмоция,  $i, o \rightarrow$ )** предикатом более высокого порядка, чем первый. Однако это только одна из возможных интерпретаций данного выражения.

Если мы подставим вместо метапеременных соответствующие предикаторы для выбранной нами в качестве примера эмоции подтвердившегося страха, то получим следующее конкретное определение предиката переживания подтвердившегося страха как диспозиционного:

**Опр. 31:**

**Подтвердившийся-страх**  $i(c)$  =<sub>df</sub>

=<sub>df</sub> **Было-или-есть Подтвердившийся-страх**  $T_i(c) \wedge$

$\wedge$  **OFI(Подтвердившийся-страх,  $i, c$ )**

В дефиниендуме этого определения находится предикат переживания эмоции подтвердившегося страха, который мы интерпретируем как диспозиционный, а в дефиниенсе этого определения имеем триггер эмоции подтвердившегося страха, который, как было показано выше, редуцируется до выражений, содержащих только операторы **ЗАПЛ** и **KARO**, то есть описывается в терминах языков эпистемической-динамической логик в соответствии с идеей эмпирической интерпретации диспозиционных предикатов. Предикат **OFI(Подтвердившийся-страх,  $i, c$ )** связывает качественное определение диспозиционного предиката с количественными характеристиками и числовыми функциями, соответствующими такому параметру, как общая интенсивность переживания эмоции подтвердившегося страха. Определение и обсуждение предиката **OFI(Подтвердившийся-страх,  $i, c$ )** и общего выражения **OFI(Эмоция,  $i, o \rightarrow$ )** в контексте введения количественных характеристик можно найти в первоисточнике<sup>55</sup>. Мы же отметим, что наличие количественных характеристик в дефиниенсе определения иллюстрирует общую идею экспликации диспозиционности через подведение под законоподобную связь, имеющую как качественный, так и количественный характер.

### Вместо заключения

Как мы отмечали в самом начале данной статьи, принятие трехаспектной структуры эмоций, укладывающейся в схему «Оценка — Переживание — Регуляция», характеризует и модель эмоций **OCC**, и ее формализацию авторами **SDM**. С этой точки зрения мы рассмотрели только один из аспектов диспозиционности, укладывающийся в отношении «Оценка — Переживание», другой аспект диспозиционности возникает при анализе второй части этой схемы, а именно отношения «Переживание — Регуляция», где регуляция описывается в терминах действий, а порождение действий — в терминах тенденций действий. Анализ того, как формализуется этот тип диспозиционности авторами **SDM**, и насколько такая формализация соответствует эмпирической интерпретации диспозиционных предикатов, мы оставляем на будущее.

<sup>55</sup> Steunebrink Bas. R. The Logical Structure of Emotions. — SIKS Dissertation Series № 2010-23 Printed by GVO drukkers & vormgevers B. V. j Ponsen & Looijen, Ede. 2010. P. 137.

## Литература

*Bas R. Steunebrink, Mehdi Dastani, and John-Jules Ch. Meyer.* A formal model of emotions: Integrating qualitative and quantitative aspects. In Ghallab Mali, Constantine D. Spyropoulos, Nikos Fakotakis, and Nikos Avouris, editors, Proceedings of the 18th European Conference on Artificial Intelligence (ECAI'08). IOS Press, 2008. Pp. 256–260.

*Bas R. Steunebrink, Mehdi Dastani, and John-Jules Ch. Meyer.* A logic of emotions for intelligent agents. In Proceedings of the Twenty-Second Conference on Artificial Intelligence (AAAI'07). AAAI Press, 2007. Pp. 142–147.

*Bas R. Steunebrink, Mehdi Dastani, and John-Jules Ch. Meyer.* The OCC model revisited. In Dirk Reichardt, editor, 4<sup>th</sup> Workshop on Emotion and Computing, Paderborn, Germany, 2009. 8 p.

*Ortony A., Clore G. L., Collins A.* The Cognitive Structure of Emotions. — Cambridge University Press. 1990. 224 p.

*Picard R. W.* Affective Computing. — MIT Press, 1997. 292 p.

*Steunebrink Bas. R.* The Logical Structure of Emotions. — SIKS Dissertation Series № 2010-23 Printed by GVO drukkers & vormgevers B. V. j Ponsen & Looijen, Ede. 2010. 228 p.

*Анисов А. М.* Концепция научной философии В. А. Смирнова. // Философия науки. Вып. 2: Гносеологические и логико-методологические проблемы. — М.: ИФ РАН, 1996. — С. 5–27.

*Райл Г.* Понятие сознания. Перевод с англ. — М.: Идея-Пресс, Дом интеллектуальной книги, 1999. — 408 с.

*Синицкий Д. А.* Экспликация понятия «диспозиционный предикат» (логические подходы) // Vox. Философский журнал. Выпуск № 22. Июнь, 2017. <https://vox-journal.org/html/issues/358/398>.

*Смирнов В. А.* Логические методы анализа научного знания. — М.: «Наука», 1987. — 256 с.

## References

Anisov A. M. Kontseptsiya nauchnoi filosofii V. A. Smirnova. [The Concept of Scientific Philosophy of V. A. Smirnov] // Filosofiya nauki. Vyp. 2. Gnoseologicheskie i logiko-metodologicheskie problemy. M.: IF RAN, 1996. S. 5–27 [Philosophy of Science. Issue. 2: Epistemological and Logical-Methodological Problems. Moscow: IF RAS, 1996. — Pp. 5–27].

Bas R. Steunebrink, Mehdi Dastani, and John-Jules Ch. Meyer. A formal model of emotions: Integrating qualitative and quantitative aspects. In Ghallab Mali, Constantine D.

Spyropoulos, Nikos Fakotakis, and Nikos Avouris, editors, Proceedings of the 18th European Conference on Artificial Intelligence (ECAI'08). IOS Press, 2008. Pp. 256–260.

Bas R. Steunebrink, Mehdi Dastani, and John-Jules Ch. Meyer. A logic of emotions for intelligent agents. In Proceedings of the Twenty-Second Conference on Artificial Intelligence (AAAI'07). AAAI Press, 2007. Pp. 142–147.

Bas R. Steunebrink, Mehdi Dastani, and John-Jules Ch. Meyer. The OCC model revisited. In Dirk Reichardt, editor, 4<sup>th</sup> Workshop on Emotion and Computing, Paderborn, Germany, 2009. 8 p.

Ortony A., Clore G. L., Collins A. The Cognitive Structure of Emotions. — Cambridge University Press. 1990. 224 p.

Picard R. W. Affective Computing. — MIT Press, 1997. 292 p.

Ryle G. Ponyatie soznaniya. Pervod s angl. — M.: Ideya-Press, Dom intellektual'noi knigi, 1999. [G. Ryle. The Concept of Mind. N.Y.: Barnes & Noble, 1949 Russ. ed.: Ponyatie soznaniya. Translated from English. M.: Ideya-Press. 1999. 408 p.]

Sinitskiy D. A. Jeksplikacija ponjatija «dispozicionnyj predikat» (logicheskie podhody) [Sinitskiy Dmitriy. An Explication of the Concept of Dispositional Predicate (The Logical Approaches) (in Russian)] // Vox. Filosofskij zhurnal. Vypusk № 22. Ijun', 2017. <https://vox-journal.org/html/issues/358/398>. DOI: 10.24411/2077-6608-2017-00005.

Smirnov V. A. Logicheskie metody analiza nauchnogo znaniya. [Logical Methods of Analysis of Scientific Knowledge. (in Russian)] — M.: «Nauka», 1987. 256 p.

Steunebrink Bas. R. The Logical Structure of Emotions. — SIKS Dissertation Series № 2010-23 Printed by GVO drukkers & vormgevers B. V. j Ponsen & Looijen, Ede. 2010. 228 p.

## Emotions and Dispositional Predicates

*Sinitskiy Dmitriy,*

**Obninsk Institute for Nuclear Power Engineering (IATE MEPHI)**

**Abstract:** Affective computing and affective calculus are currently one of the rapidly developing areas of interdisciplinary research. The aim of the article is to examine the approach to building a formalized theory of emotions developed by the Dutch logicians B. Steunebrink, M. Dastani and J.-J. Ch. Meyer, as well as the analysis of this theory for the use of dispositional predicates in it. Reduction definitions for observation predicates and dispositional predicates in the analyzed theory of emotions are constructed in terms of emotion triggers and emotion experience predicates, which are then reduced to expressions of the language of epistemic and dynamic logics.

**Keywords:** affective computing, affective calculus, emotion, dispositional predicate.

**Синицкий Дмитрий Анатольевич**  
Кандидат философских наук,  
доцент кафедры философии и социальных наук ИАТЭ НИЯУ МИФИ,  
Обнинский институт атомной энергетики НИЯУ МИФИ  
249040, Россия, Калужская обл., г. Обнинск, ул. Энгельса, 24, кв. 35  
das\_sid@mail.ru  
Моб. тел. +7-910-709-67-33

**Sinitskiy Dmitriy Anatolyevich**  
Candidate of Philosophical Sciences, Assoc. Prof.  
Obninsk Institute for Nuclear Power Engineering (IA TE MEPHI),  
Russia, Kaluga region, Obninsk  
das\_sid@mail.ru