

**Моисеева Анна Юрьевна**

## **Невозможные миры в моделировании убеждений**

Аннотация: В статье рассматривается один из подходов к моделированию убеждений, разработанный для преодоления проблем стандартной семантики возможных миров – семантика с невозможными мирами. Показано, что само по себе введение невозможных миров не позволяет справиться с указанными проблемами, пока миры остаются замкнутыми относительно какой-то принятой логики. Раскрыта суть задачи моделирования ограниченной рациональности и охарактеризован способ ее решения в семантике, разработанной М. Джэго и позднее развитой им совместно с Ф. Берто. Сделан вывод о том, что поставленная задача действительно была решена в семантике М. Джэго. Однако особенности этой семантики сделали трудным построение на ее базе эпистемической и доксистической логики и последующее доказательство теорем. Тем не менее перспективы данного подхода и в целом семантики убеждений с использованием невозможных миров оцениваются как благоприятные на фоне серьезных успехов и бурного развития всего этого направления в последнее время.

Ключевые слова: невозможные миры, семантика и логика убеждений, гиперинтенциональность.

### **Введение**

Моделирование убеждений – важная задача для философии языка, потому что ее решение позволит формулировать условия истинности для т. н. аскрипций – сложных предложений, в которых придаточное предложение управляется глаголом «убежден», «верит», «считает» или «полагает» (здесь все эти глаголы будут использоваться как синонимы англоязычного *believe*). Одновременно это важная задача для философии сознания, поскольку знать убеждения человека нужно для понимания мотивов его поступков. Существует множество семантических подходов, в разной степени пригодных для моделирования убеждений, и рассмотреть их все в рамках одной статьи невозможно. В настоящей статье внимание сконцентрировано лишь на одном подходе,

берущем начало в семантике возможных миров, а именно на подходе, использующем невозможные миры. Для понимания того, зачем вообще такие миры вводятся, мы рассмотрим сначала те проблемы, с которыми сталкивается стандартная семантика возможных миров при работе с доксистическими контекстами.

### Проблемы моделирования убеждений в семантике возможных миров

Как известно, в семантике возможных миров модальные высказывания интерпретируются на множестве возможных миров, каждый из которых представляет собой некоторое полностью определенное положение дел, которое могло бы быть реализовано, может быть реализовано или уже реализовано в действительности. Каждая модальность задается с помощью соответствующего ей отношения достижимости между миром, относительно которого оценивается истинность модального высказывания (обычно это просто мир, где высказывание произносится), и некоторым множеством миров так, что условие истинности высказывания выражает соотношение между достижимыми мирами и мирами, в которых истинно придаточное предложение высказывания. Например, предложение «Вася убежден, что снег бел» истинно, согласно этому анализу, тогда и только тогда, когда все миры, доксистически достижимые для Васи из действительного мира, – это миры, где снег бел.

Такой анализ позволяет обосновывать выводы, которые мы склонны делать из подобных высказываний. Ясно, например, что все миры, где снег бел, будут одновременно и мирами, где снег не черен<sup>1</sup>, таким образом, мы можем сделать заключение об истинности предложения «Вася убежден, что снег не черен». Вместе с тем мы получаем простое объяснение того, как убеждение, приписываемое нами Васе, должно влиять на его поведение: быть убежденным, что снег бел, значит при принятии решений о том, как действовать или размышлять в соответствующих ситуациях, ориентироваться только на те возможные варианты развития событий, которые предполагают близину снега<sup>2</sup>.

---

<sup>1</sup> При использовании в обычном значении предикатов «бел» и «черен», а также классической логики.

<sup>2</sup> Такая диспозициональная интерпретация доксистической модальности используется у Я. Хинтикки (Hintikka, 1962), Д. Льюиса (см., например, Lewis, 1975), Р. Столнейкера (Stalnaker, 1976, 1984) и др. Она не является ни единственно возможной, ни самой точной, поскольку, как показал Дж. Коэн (Cohen, 1992), диспозиция к определенному образу действий и мыслей более точно соответствует понятию *принятия* (*acceptance*), а не убеждения. Принятие и убеждение, как утверждается, в норме

При этом подход семантики возможных миров в большинстве случаев позволяет различать убеждения, истинность которых в действительном мире совпадает. Например, предложение «Вася убежден, что снег бел» и предложение «Вася убежден, что трава зелена» приписывают Васе разные убеждения. Хотя в действительном мире снег всегда бел и трава всегда зелена, в некоторых возможных мирах выполняется первое, но не выполняется второе, и наоборот. Значит, множество возможных миров, которые являются доксистически достижимыми для Васи из действительного мира, согласно первой аскрипции, будет не совпадать с множеством возможных миров, которые являются доксистически достижимыми для него же, согласно второй аскрипции. А поскольку мы моделируем содержания убеждений именно как множества возможных миров, мы должны заключить, что эти два убеждения различаются по своему содержанию. Это интуитивно убедительное заключение, поскольку на практике, очевидно, есть существенная разница между поведенческими склонностями человека, который ориентируется на белизну снега, и человека, который ориентируется на зелень травы. Например, первый будет склонен использовать белый камуфляж зимой, а второй – зеленый камуфляж летом, но не наоборот. Таким образом, семантика возможных миров, являющаяся по своей сути развитием экстенционального подхода к значению, оказывается способной учитывать различия в интенционалах.

Однако в тех случаях, когда придаточные предложения аскрипций истинны на одних и тех же множествах возможных миров, приписываемые такими аскрипциями убеждения неразличимы, с точки зрения данной семантики. Это создает контринтуитивные следствия, которым посвящена обширная литература. Мы здесь не будем разбирать все примеры, сконструированные различными авторами для того, чтобы показать, какие сбои дает семантика возможных миров при анализе убеждений, лишь перечислим основные случаи, в которых имеют место эти сбои.

1. *Аскрипции с кореферентными собственными именами.* Обычно собственные имена рассматриваются в семантике возможных миров как жесткие десигнаторы. Это означает, что замена одного собственного имени в аскрипции на другое, обозначающее тот же объект, не должна влиять на истинность аскрипции. Однако интуитивно это

---

сопутствуют друг другу, поэтому их часто путают, хотя это разные явления. Однако все эти детали не важны для нашей темы.

неприемлемый вывод, что первым показал Г. Фреге в своем знаменитом примере с Утренней звездой и Вечерней звездой<sup>3</sup>.

2. *Аскрипции с видовыми терминами, обладающими одним и тем же интенционалом.* Многие естественные виды имеют различные наименования в разных географических областях, в разных сферах человеческой деятельности и т. д. С точки зрения семантики возможных миров, два наименования взаимозаменяемы, если они имеют одинаковый интенционал, то есть если две функции, приписывающие значение этим терминам в каждом возможном мире, совпадают. Однако на практике некоторые агенты употребляют одно из наименований более уверенно, чем другое. В таких случаях замена первого на второе в аскрипции может привести к тому, что агент больше не будет согласен с приписыванием ему данного убеждения.

3. *Аскрипции с «пустыми» именами и дескрипциями.* Если в каком-то возможном мире нет ни одного объекта, который входил бы в экстенционал определенного имени или дескрипции в этом мире, то предложение, в котором на месте субъекта высказывания стоит данное имя или данная дескрипция, не является истинным в этом мире. Очевидно, что если имя или дескрипция, стоящие на месте субъекта высказывания, «пусты» во всех мирах<sup>4</sup>, то пусто и множество миров, в которых истинно такое предложение. Это означает, что ни одному агенту нельзя осмысленно приписать убеждение, выражаемое таким предложением. Однако интуитивно мы вполне понимаем смысл аскрипций типа «Маша верит, что Дед Мороз придет сегодня ночью» и полагаем, что некоторые агенты в некоторые периоды своей жизни действительно имеют подобные убеждения.

4. *Аскрипции с необходимо истинными придаточными предложениями.* Обратная ситуация имеет место, если какое-то предложение истинно во всех возможных мирах<sup>5</sup>. С точки зрения семантики возможных миров, все такие предложения равнозначны, и все они выражают убеждения, каждым из которых обладает любой агент в любое время. Очевидно, что это не так. Истинность некоторых предложений логики

---

<sup>3</sup> Примеры, построенные аналогично примеру Фреге, широчайшим образом распространены в литературе по семантике контекстов убеждения. Проблема, связанная с ними, носит название головоломки Фреге (*Frege's puzzles*). Всестороннее обсуждение и попытку решения этой проблемы см. в одноименной книге (Salmon, 1986), критику и другой подход к решению – в (Shiffer, 1978; 1992).

<sup>4</sup> Тезис о необходимом характере предложений, утверждающих отсутствие, обычно принимается применительно к фикциональным объектам.

<sup>5</sup> Здесь имеются в виду предложения без включенных в них модальных слов. Ради строгости можно также уточнить, что они необходимо истинны при данном наборе логических правил и/или математических аксиом.

и математики (например, Последней теоремы Ферма) остается не известной никому в течение веков, а потом получает строгое доказательство, которое способны понять лишь специалисты. Интуитивно представляется, что мы должны иметь возможность приписывать агентам неверие (*disbelieving*) в такие предложения.

Представленная здесь группировка не полностью отражает структуру головоломок (*puzzles*), возникающих в каждом случае, она используется просто для удобства. Следует понимать, что перечисленные случаи, за исключением случая 3<sup>6</sup>, порождаются одними и теми же особенностями семантики возможных миров, а потому в литературе они часто обсуждаются вместе под общим именем *проблемы логического всеведения*. Эта проблема фундаментальна для семантики пропозициональных установок. Можно сказать, что появление каждого нового проекта в этой области связано с тем, как тот или иной исследователь понимает источник и суть проблемы логического всеведения. Не претендуя здесь на разработку собственного решения, мы просто изложим один из подходов, предложенный в книге М. Джэго «Невозможное: эссе о гиперинтенциональности» (Jago, 2014) и позднее развитый в совместной работе Ф. Берто и М. Джэго «Невозможные миры» (Berto, Jago, 2019), а затем попытаемся осмыслить результаты, полученные в рамках этого подхода.

### Понятие и виды невозможных миров

Само существование проблемы логического всеведения показывает, что для адекватного анализа контекстов убеждений семантике недостаточно способности каким-то образом учитывать интенционалы. Она должна допускать *гиперинтенциональность*, т. е. чувствительность значения к логически эквивалентным подстановкам. Согласно обычному определению, оператор '*O*' называется гиперинтенциональным при условии, что для логически эквивалентных '*A*' и '*B*' не всегда выполняется материальная эквивалентность '*OA*' и '*OB*' (Jago, 2014, 50; так же в Berto, 2017). Существуют различные подходы к построению гиперинтенциональной семантики, но самым естественным путем, на наш взгляд, является введение понятия *невозможного мира*. Именно этот путь рассматривается в настоящей статье.

Наиболее интуитивное определение невозможного мира гласит, что «если возможные миры представляют собой то, какими вещи могли бы быть, то невозможные миры представляют собой то, какими вещи не

---

<sup>6</sup> Случай 3 в нашем перечне стоит особняком, поскольку он связан не с логическими свойствами рассматриваемого типа предложений, а с их сугубо семантическими свойствами. Поэтому в данной статье мы практически не будем его касаться.

могли бы быть: они изображают нечто абсолютно невозможное как имеющее место» (Berto, 2017). При этом абсолютная невозможность первично определяется, исходя из понятия необходимости: если некоторое  $A$  необходимо, то  $\sim A$  абсолютно невозможно. Т. е. тот же смысл, который придается понятию необходимости, автоматически придается и понятию невозможности. Например, если мы вводим, помимо логической и математической необходимости, еще и метафизическую необходимость, то получаем класс метафизически невозможных миров. Поскольку обычно о метафизической необходимости говорят в связи с принципом необходимости тождества, если этот принцип признается, то невозможными в метафизическом смысле оказываются миры, где актуально тождественные индивиды являются различными (тогда как все другие логические и математические законы могут соблюдаться).

Миры, невозможные в силу нарушения в них различных законов и правил классической логики, называются *ненормальными мирами*. Можно различать миры, где выполняются разные наборы логических законов, и таким образом выделять различные специальные смыслы понятия ненормальности. Например, паранепротиворечивые и парাপолные логики порождают каждая собственный класс миров, которые могут использоваться для различных целей. Следует отметить, что, как и в случае с метафизической невозможностью, логическая невозможность является таковой именно из перспективы определенного взгляда на логику. Мы можем использовать неклассические теории вывода в каких-то специфических контекстах (эпистемических или доксистических, например), но при этом верить в то, что, глобально, классическая логика является истинным описанием того, что «бывает» в мире. В этом случае мы будем считать, что все ненормальные миры невозможны. Если же принять точку зрения, согласно которой сам мир в том или ином аспекте может быть таким, что, например, в нем выполняется  $A$  и  $\sim A$  одновременно (диалетизм), или таким, что в нем не выполняется ни  $A$ , ни  $\sim A$  (интуиционизм), то соответствующие миры попадут в множество возможных.

Кроме того, если в логике используются специфическим образом действующие модальные операторы, то опять же необходим соответствующий класс невозможных миров. Например, если в рамках эпистемической логики не принимается правило Геделя, то такая логика порождает класс миров, в которых «ненормальным» образом функционирует знание (как мы увидим ниже, такой является логика FDE). Достаточно широкий взгляд на рациональность позволяет говорить о

самых разных модальных логиках с самыми разными свойствами<sup>7</sup>. И поскольку каждой из модальностей соответствует свое собственное отношение достижимости, которое может обладать, вообще говоря, любыми свойствами в любой комбинации, потенциально каждая из них порождает особый класс невозможных миров. Дополнительно здесь существует опция задания специфического (неопределенного или градуированного) отношения достижимости между мирами, поскольку оно также определяет особое «поведение» модальных операторов, а значит, влияет и на логическую структуру миров<sup>8</sup>.

В целом, то, где конкретно мы проводим границу между возможными и невозможными мирами, может быть совсем не важно, если наш формализм позволяет обращаться с обоими множествами одним и тем же способом (например, как в FDE, см. ниже). Важнее то, какие возможные и невозможные миры у нас вообще есть. Разумно требовать, чтобы их было достаточно для любых наших целей. Иначе говоря, поскольку невозможные миры обычно рассматриваются как альтернативы актуальному миру для некоторого типа агентов, их множество должно быть таким, чтобы в нем были представлены все полноценные альтернативы, о которых способны осмысленно рассуждать такие агенты (Berto, 2017). Даже если мы признаем некоторую ситуацию невозможной из-за того, что в ней нарушается один из законов логики, но интересуемся тем, что было бы, если бы такая ситуация имела место, нам нужны соответствующие этой ситуации невозможные миры. Есть мнение (Priest, 2005), что в невозможных мирах можно допускать даже нарушение базового закона логики  $A \vdash A$ , хотя, на наш взгляд, не очевидно, что рассуждения, для моделирования которых могут потребоваться такие миры, действительно осмысленны.

Как утверждают Ф. Берто и М. Джэго в своем историческом обзоре (Berto, Jago, 2019, 112), первые шаги по направлению к семантике с невозможными мирами были сделаны Я. Хинтиккой в 1970-х гг. Эпистемическая логика, которую он развивал в своих пионерских работах предыдущего десятилетия (в первую очередь в Hintikka, 1962), сразу же столкнулась с проблемой логического всеведения. И хотя сначала Я. Хинтикка пытался обойти эту проблему, говоря, что его логика просто формализует другую модальность («защищаемость» (*defensibility*), «информированность» или что-то подобное)<sup>9</sup>, позже он пришел к

---

<sup>7</sup> В качестве примера можно привести работы Ф. Берто по логике воображения (Berto, 2017).

<sup>8</sup> В последнем разделе настоящей статьи мы рассмотрим пример такого отношения.

<sup>9</sup> Обсуждение этой стратегии см. в (Jago, 2007).

выводу, что необходимо все же модифицировать сам формализм. Так, в его статье 1975 г. «В защиту невозможных возможных миров» провозглашается необходимость создания такой эпистемической логики, в которой отбрасывалось бы допущение, что каждый эпистемически возможный мир является логически возможным, потому что принятие этого допущения неизбежно приводит к логическому всеведению. «Именно потому, что люди... вовсе не стремятся вывести все логические следствия *ad infinitum* из того, что они знают, они могут воображать картины, которые только кажутся возможными, но содержат в себе скрытые противоречия» (Цит. по русскому переводу Хинтиikka, 1980, 233). Более того, Я. Хинтиikka даже намечает подход к решению этой задачи. Однако стоит сказать, что «полнокровным» понятием невозможного мира в рамках своего подхода он все же не оперирует<sup>10</sup>.

Другим источником, к которому часто обращаются, говоря о раннем развитии семантики с невозможными мирами, является работа Н. Решера и Р. Брэндома «Логика несовместимости. Исследование семантики и онтологии нестандартных возможных миров» (Rescher N., Brandom R. B. *The Logic of Inconsistency. A Study in Non-Standard Possible-Worlds Semantics and Ontology*) (Излагается по Berto, Jago, 2019, 103–104). Н. Решер и Р. Брэндом дополняют стандартное множество возможных миров множеством т. н. «нестандартных» миров, которые образуются путем комбинирования возможных миров с применением рекурсивных операций схематизации и суперпозиции. Нестандартными миры называются потому, что в них в силу свойств операций, с помощью которых их создают, необычным образом работают конъюнкция и дизъюнкция. Так, среди них имеются миры, где истинно  $A$  и истинно  $B$ , но неверно, что истинно  $A \wedge B$ , а также миры, где истинно  $A \vee B$ , но неверно, что истинно  $A$  или истинно  $B$ . Кроме того, нестандартные миры допускают противоречие. При этом выводы с одной посылкой работают на этих мирах совершенно обычным образом, т. е. полной логической «анархии» в них все же не наблюдается.

Более радикально подходит к семантике с невозможными мирами Г. Прист, который давно использует сходные построения для формализации своих философских взглядов. В его работах с конца 1970-х гг. по настоящее время разрабатываются различные неклассические подходы к семантике и логике. Так, в книге 1987 г. «В противоречии: исследование трансконсистентного» (второе издание Priest, 2006) он строит релевантную логику для формализации

---

<sup>10</sup> Данный подход, носящий название урновых моделей, будет кратко рассмотрен в следующем разделе.

диалектического взгляда на истину. С помощью своей релевантной логики Г. Прист далее формализует, в частности, парадоксы, которые в данной интерпретации являются не чем иным, как предложениями, которые являются истинными и ложными одновременно. Позже он пишет, что наиболее подходящей для такой логики семантикой является именно семантика с невозможными мирами, в которых нестандартно действует операция отрицания (Priest, 2006, 270). Для решения других задач он привлекает миры, где не выполняется закон исключенного третьего и, в общем-то, любой другой закон логики. Таким образом Г. Прист приходит к семантике *открытых миров*, т. е. миров, не замкнутых относительно каких бы то ни было логических правил<sup>11</sup>.

В работе «К небытию: логика и метафизика интенциональности» (Priest, 2005) Г. Прист развивает уже весьма близкую к нашей тему: он рассуждает об интенциональности и содержании интенциональных ментальных состояний. Он намеренно строит свой подход с расчетом на семантику с невозможными мирами, хотя сначала принимает ограничения, соответствующие обычным возможным мирам. Далее объектом его атаки является, как нетрудно догадаться, логическое всеведение, а также интенциональный вариант формул Баркан – известных модальных теорем, касающихся онтологии возможных объектов<sup>12</sup>. Г. Прист показывает, что принятые ограничения ведут к контринтуитивным следствиям применительно к семантике интенциональных операторов, и отсюда заключает, что необходимо эти ограничения снять. В оставшейся части книги он развивает семантику с открытыми мирами и релевантную логику, показывая, как можно применять этот аппарат к широкому кругу проблем, связанных с содержанием интенциональных ментальных состояний.

Таким образом, к настоящему моменту семантика с невозможными мирами имеет уже достаточно внушительную историю и большое количество разнообразных приложений. Количество работ в этом направлении, появляющихся в последние годы, таково, что их невозможно охватить даже специалисту. Тем не менее работ, посвященных именно семантике убеждений, не так уж много. По-видимому, сложность этой темы состоит в том, что убеждения, в отличие от большинства других интенциональных ментальных состояний, явным

---

<sup>11</sup> В дальнейшем нашем изложении открытые миры будут играть существенную роль.

<sup>12</sup> Именно формулы Баркан являются наиболее неприемлемыми для Г. Приста как «сторонника небытия», поскольку они фактически запрещают приписывать агентам интенциональные отношения к фикциям. В нашем перечне эта проблема соответствует случаю 3.

образом подчиняются требованиям рациональности. Хотя очевидно, что реальные эпистемические агенты не демонстрируют логического всеведения, в норме от них ожидается, что они не будут продолжать придерживаться противоречивых убеждений, по крайней мере после того, как противоречие обнаружено. На этом ожидании основана, в частности, практика логической аргументации, без учета которой невозможно понять функцию убеждений в человеческой культуре<sup>13</sup>. Т. е. исследователь семантики убеждений, с одной стороны, должен отбросить идеализированные нормы рациональности, навязываемые семантикой возможных миров, но, с другой стороны, не должен разрушать само представление о рациональности. Задача, которая стоит перед ним, настолько нетривиальна, что непонятно даже, как ее ясно сформулировать. Далее, по мере нашего продвижения в последующих разделах, мы увидим, как меняется понимание этой задачи, и достижение ее правильной постановки становится важным этапом ее решения в рамках того подхода, на котором будет сконцентрировано наше основное внимание.

### **Семантика с замкнутыми мирами и варианты ослабления логики**

Основываясь на сказанном выше, можно предположить, что для получения логики, которая адекватно отражала бы свойства модальности убеждения, нужно просто взять логику, которая задается семантикой возможных миров, и попытаться исключить некоторые ее правила или заменить их более слабыми, а потом посмотреть, какими должны быть миры, на которых выполнялась бы эта логика. Но если так, встает вопрос: насколько именно должна быть слабой искомая логика? Очевидно, что, хотя нужно избежать логического всеведения, в то же время некоторые умозаключения имеет смысл считать тривиальными и «автоматически» производящимися всяким рациональным агентом, если мы хотим иметь хоть какое-то представление о нормах рационального мышления. Целью данного раздела является прояснение имеющихся вариантов ответа на этот вопрос с использованием аппарата семантики с невозможными мирами.

Если мы посмотрим на модальную логику, которая поддерживается стандартной семантикой возможных миров, то увидим, что

---

<sup>13</sup> Правда, отдельные агенты тем не менее могут защищать явно противоречивые позиции (как делают те же диалектисты), однако они, как правило, делают это не спонтанно, а с опорой на специфические философские и/или практические установки применительно к специфическому контексту. Например, религиозный человек может утверждать противоречие о Боге, чтобы подчеркнуть, что «факты» такого рода не описываются категориями человеческого разума.

замыкание множества предложений, в которых убежден агент<sup>14</sup>, согласно данной формализации, относительно применения всех экстенциональных операций в любом сочетании и любое количество раз обеспечивается принятием всего одного правила и одной аксиомы, а именно

$$(Nec) \frac{A}{KA} \text{ – правила Геделя}$$

и

$$(K) K(A \rightarrow B) \rightarrow (KA \rightarrow KB) \text{ – аксиомы Крипке<sup>15</sup>}.$$

Правило Геделя гласит, что агент убежден во всех общезначимых предложениях. Отсюда можно получить, в частности, убежденность агента во всех логических и математических истинах, как бы сложны они ни были (то, что мы выше обозначили как случай 4). Аксиома Крипке гласит, что убеждение агента в импликации совместно с его убеждением в антецеденте этой импликации влечет его убеждение также и в консеквенте. Отсюда можно получить замкнутость системы убеждений относительно применения агентом *modus ponens*, опять же как бы сложны ни были импликации, с которыми ему придется иметь дело. Совместное же рассмотрение аксиомы Крипке и правила Геделя дает замкнутость системы убеждений относительно логического следования и логической эквивалентности. Именно она ответственна за появление головоломок, в которых используется замена в аскрипции

---

<sup>14</sup> Как мы помним, с точки зрения семантики возможных миров, убеждения агента – это предложения, истинные на всем множестве возможных миров, достижимых из данного мира. Среди этих предложений некоторые истинны вообще на всем пространстве миров, т. е. имеют характер глобальных посылок. Таковы, например, утверждения математики. Другие истинны только в достижимых мирах, а в недостижимых мирах могут оказаться ложными. Такие предложения называются локальными посылками. Строго говоря, выводимость из глобальных и локальных посылок логически задается по-разному. Так, в локальном выводе не используется правило Геделя. Однако для решения проблемы логического всеведения это различие ничего не дает, поэтому мы здесь не фокусируемся на нем. В контексте проблемы логического всеведения важным здесь является то, что, коль скоро при данном множестве (локальных и глобальных) посылок из них (в данной логике) выводится определенное заключение, это заключение считается убеждением агента, даже если длина вывода такова, что никакой реальный агент (по крайней мере, ни один человек) не смог бы не только построить его, но и рассмотреть шаг за шагом.

<sup>15</sup> Здесь мы используем нотацию эпистемической логики, но будем интерпретировать *K* как оператор убеждения. В действительности применительно к проблеме логического всеведения между этими двумя модальностями нет почти никакой разницы.

коррелятивных имен и видовых терминов с одним и тем же интенционалом (то, что мы выше обозначили как случаи 1 и 2). Таким образом, даже логика К является намного более сильной, чем та, которую можно было бы считать адекватной логикой убеждения.

Ясно, однако, что принятие правил логики К является неизбежным следствием правил самой семантики возможных миров. Так, если мы получаем значения истинности для сложных формул в мирах не иначе как путем применения к более простым формулам определенных семантических правил, то очевидно, что ни в одном мире не может возникнуть ситуация, когда эти сложные формулы будут иметь иное значение истинности, чем то, которое предполагается правилами. Это значит, что, поскольку убеждения моделируются как множества миров, ни один агент не может иметь иных убеждений относительно сложных предложений, чем те, которые логически следуют из более простых предложений, в которых он уже убежден. И правило Геделя, и аксиома Крипке (при использовании стандартной семантики для импликации) обосновываются семантическим способом, а не берутся «с потолка». Чтобы была другая логика, нужно изначально иначе строить семантику.

Таким образом, мы должны сконцентрироваться на вопросе о том, какие есть альтернативные способы построения семантики, в частности способы задания значения истинности для сложных формул. В качестве наиболее очевидной альтернативы можно, например, попытаться ввести невозможные миры, в которых истинность всем формулам будет приписываться непосредственно функцией валуации. Одним из первых такой подход стал использовать В. Рантала в статье 1982 г. «Невозможные миры и логическое всеведение» (Impossible Worlds Semantics and Logical Omniscience) именно для того, чтобы более правдоподобно моделировать пропозициональные установки. Другим новшеством, введенным В. Ранталой в семантику возможных миров с этой целью, стали т. н. *урновые модели* – подход, основанный на интерпретации первопорядковых кванторов как выборов из варьирующейся популяции индивидов. Этот подход, во многом пересекающийся с работами Я. Хинтикки в области теоретико-игровой семантики, был далее развит самим Я. Хинтиккой в вышеупомянутой статье 1975 г. (Русский перевод Хинтикка, 1980).

Мы не будем здесь излагать формальный аппарат, используемый в теории урновых моделей, поскольку это не касается напрямую нашей темы. Заметим, однако, что, хотя сама по себе идея оценки сложности логического вывода посредством измерения его квантификационной «глубины» представляется интересной и имеющей большие перспективы, именно для построения правдоподобной семантики и логики

убеждений она оказывается малополезной. Как показывает М. Джэго в своей статье и позже в книге (Jago, 2007; Jago, 2014, 180–184), избежать проблемы логического всеведения с помощью этой идеи можно только до тех пор, пока рассматриваются предложения с кванторами. Для предложений пропозициональной логики от введения урновых моделей ничего не меняется, агенты по-прежнему моделируются как всеведущие. Более того, данный подход даже выводу с кванторами не позволяет приписать информативность, если его квантификационная глубина не превышает уровня логической компетентности агента. Это означает, что все выводы, которые мы только способны сделать, применив максимум своих знаний и навыков в области логики, рассматриваются в семантике урновых моделей как «автоматические», что в высшей степени контринтуитивно. Многие наши выводы нужно считать информативными по той простой причине, что, прежде чем мы их сделали, мы понятия не имели об истинности их заключений. Короче говоря, семантика с урновыми моделями поддерживает логику, которая является все еще намного более сильной, чем нам нужна. Фактически, это та же самая логика, что соответствует обычной семантике возможных миров, только с ограничением на глубину вывода.

Чтобы получить существенно более слабую логику, можно, помимо введения невозможных миров, модифицировать само понятие логического следования. Ф. Берто и М. Джэго приводят вариант семантики, в которой применяется такой подход. Эта семантика называется FDE, поскольку экстенциональный фрагмент логики, которую она поддерживает, оказывается при ближайшем рассмотрении *логикой первоуровневого вывода* (first-degree entailment), предложенной в 1976 г. Дж. Данном (Dunn J. Intuitive semantics for first-degree entailments and coupled trees) и независимо в 1977 г. Н. Белнапом (Belnap N. A useful four-valued logic) (Излагается по: Berto, Jago, 2019, 114–118). Мы рассмотрим ее более подробно, поскольку она дает четкое представление о том, как модификации условий истинности вкпе с понятием логического следования, предпринимаемые в рамках семантики, влияют на законы, которые действуют в логике, моделью для которой такая семантика является.

Фрейм в FDE задается стандартным образом, как пара  $\langle W, R \rangle$ , где  $W$  – множество миров, а  $R$  – отношение достижимости на этом множестве. Множество  $W$  рассматривается как гомогенное, т. е. возможные миры в нем не отличаются от невозможных. Главной особенностью FDE-модели является использование отношения валуации  $\rho$  вместо более привычной функции валуации. Отношение  $\rho$  связывает атомарные формулы, миры из  $W$  и элементы множества истинностных значений

$\{0, 1\}$  так, что атомарная формула  $p$  может быть истинной в мире  $w$  (обозначается ' $\rho_w p1$ '), ложной в  $w$  (обозначается ' $\rho_w p0$ '), одновременно истинной и ложной в  $w$ , а также ни истинной, ни ложной в  $w$ .

Для формул с логическими связками принимаются следующие условия истинности:

$$(S1\sim) \rho_w(\sim A)1 \text{ е. т. е. } \rho_w A0,$$

$$(S2\sim) \rho_w(\sim A)0 \text{ е. т. е. } \rho_w A1,$$

$$(S1\wedge) \rho_w(A \wedge B)1 \text{ е. т. е. } \rho_w A1 \text{ и } \rho_w B1,$$

$$(S2\wedge) \rho_w(A \wedge B)0 \text{ е. т. е. } \rho_w A0 \text{ или } \rho_w B0,$$

$$(S1\vee) \rho_w(A \vee B)1 \text{ е. т. е. } \rho_w A1 \text{ или } \rho_w B1,$$

$$(S2\vee) \rho_w(A \vee B)0 \text{ е. т. е. } \rho_w A0 \text{ и } \rho_w B0,$$

Можно видеть, что, хотя FDE-миры могут быть противоречивыми и неполными, они имеют определенную логическую структуру. Так, среди них нет таких миров, где была бы истинна конъюнкция, но неверно, что истинны оба конъюнкта, или истинна дизъюнкция, но неверно, что истинен хотя бы один дизъюнкт. Они также поддерживают законы ослабления и навешивания / снятия двойного отрицания.

Для отображения модальности в FDE вводится оператор  $K$ , формулы с которым имеют следующие условия истинности:

$$(S1K) \rho_w(KA)1 \text{ е. т. е. для всех } w' \in W \text{ таких, что } Rww', \rho_w A1,$$

$$(S2K) \rho_w(KA)0 \text{ е. т. е. для некоторого } w' \in W \text{ такого, что } Rww', \rho_w A0.$$

Видно, что, хотя используется обозначение, традиционное для эпистемической логики, здесь  $K$  по смыслу соответствует скорее докстатической, чем эпистемической модальности. Причем даже как докстатический оператор  $K$  в FDE имеет относительно слабые свойства. Так, согласно семантике этого оператора, вполне возможна ситуация  $\rho_w(K(A \wedge \sim A))1$ , то есть агент может быть убежден в явном противоречии. Однако такая ситуация не влечет за собой убеждения в чем угодно, т. е. «взрыва». Этого не происходит потому, что логическое

следование в FDE определяется как сохранение истинности во всех (возможных и невозможных) мирах всех моделей:

$\Gamma \models A$  е. т. е. во всех моделях  $M = \langle W, R, \rho \rangle$  и во всех мирах  $w \in W$  выполняется, что если  $\rho_w B1$  для всех  $B \in \Gamma$ , то  $\rho_w A1$ .

Таким образом, даже если  $\rho_w(A \wedge \sim A)1$  во всех мирах  $w' \in W$  таких, что  $Rww'$ , необязательно, чтобы в них было также  $\rho_w B1$  для произвольного  $B$ , а значит

$$K(A \wedge \sim A) \not\equiv KB.$$

Аналогично можно показать, что в FDE не выполняется ни правило Геделя, ни аксиома Крипке, ни даже элементарный *modus ponens*<sup>16</sup>. Это означает, что отсутствует возможность делать нетривиальные выводы, причем не только о тех мирах, которые являются достижимыми из данного мира, но и вообще о каких бы то ни было мирах. Отсюда можно было бы заключить, что FDE в качестве логики убеждений предъявляет минимальные требования к рациональности агента.

Однако, как отмечают Ф. Берто и М. Джэго (Berto, Jago, 2019, 117), логическое всеведение возвращается в FDE с неожиданной стороны. Как сказано выше, семантика экстенциональных операторов FDE все же поддерживает применение некоторых правил, в частности правила ослабления. Это означает, что, зная о наличии у агента убеждения в некотором  $A$ , мы имеем право вывести его убеждение в  $A \vee B$  для произвольного  $B$ . Такой вывод представляется сомнительным, ведь  $B$  может включать термины, которыми наш агент в действительности не владеет. Если так, то он не согласится с приписыванием ему убеждения в  $A \vee B$ . Некоторые сомнения вызывает и правило навешивания / снятия двойного отрицания. Представим ситуацию, что предложение, убеждение в котором приписывается агенту, включает отрицание, повторенное миллион раз. Очевидно, что в силу ограниченных когнитивных способностей агент не способен оперировать выражениями такой длины и поэтому не смог бы быть убежден в истинности данного предложения, если бы ему пришлось это предложение рассмотреть, даже при условии,

---

<sup>16</sup> Аксиома К не может быть даже сформулирована в FDE, поскольку она предполагает вложенные импликации, тогда как FDE работает с ограниченным синтаксисом, в котором вложенных импликаций нет. *Modus ponens* же «проваливается» потому, что импликация определяется как в пропозициональной логике:  $A \rightarrow B$  е. т. е.  $\sim A \vee B$ . Поскольку миры могут быть противоречивыми, не в каждом мире, где истинно  $\sim A \vee B$  и одновременно истинно  $A$ , будет истинно  $B$ .

что он убежден в истинности того же самого предложения без отрицаний.

В целом, FDE предполагает замкнутость системы убеждений относительно того понятия следования, которое там принимается. И хотя это понятие следования отличается от классического, Ф. Берто и М. Джэго заключают, что оно в равной степени непригодно для правдоподобного моделирования пропозициональных установок. Ни один реальный агент не способен рассмотреть бесконечное количество следствий из своих убеждений. Именно в этом состоит главная проблема, а не в том, что слишком много слишком сильных правил признаются действующими в рамках логики убеждений.

Таким образом, исследователь докстастических контекстов, как кажется, встает перед неприятной дилеммой: либо принять логическое всеведение, пусть в ослабленном виде, либо отказаться от всяких притязаний на формализацию рациональности агента. В книге М. Джэго (Jago, 2014) эта дилемма называется *проблемой ограниченной рациональности*. Задачей следующего раздела будет рассмотреть, как она разрешается в той семантике, которую сам М. Джэго предлагает в качестве лучшего кандидата на роль модели для убеждений ограниченно рационального агента<sup>17</sup>. Затем последует раздел, посвященный непосредственно модальностям.

---

<sup>17</sup> Другой подход к решению проблемы ограниченной рациональности, использующий невозможные миры, развивается в рамках динамической эпистемической логики (Bjerring, Skipper, 2019; Rasmussen, 2015). В этой логике используется не одна модель, а множество моделей, каждая из которых включает свое множество возможных и невозможных миров. Модели могут рассматриваться как эквивалентные данной для агента из мира  $w$  при условии, что он может использовать дедуктивные рассуждения длиной не более чем в  $n$  шагов. Если в какой-то из таких эквивалентных моделей во всех мирах истинно некоторое  $A$ , это означает, что агент может за  $n$  шагов прийти к убеждению, что  $A$  (записывается  $\langle n \rangle A$ ). Однако, как замечают по этому поводу Ф. Берто и М. Джэго (Berto, Jago, 2019, 121), такая трактовка динамического оператора не позволяет нам сказать, что агент действительно придет к какому бы то ни было убеждению. Если же мы будем использовать более сильное требование истинности  $A$  во всех мирах *всех* эквивалентных моделей, означающее, что агент *должен* за  $n$  шагов прийти к убеждению, что  $A$  (записывается  $[n]A$ ), то окажется, что единственным таким обязательным убеждением для агента, начинающего с убеждения в множестве предложений  $\Gamma$ , будет само это  $\Gamma$ . В самом деле, поскольку существует множество альтернативных цепей рассуждения, легко понять, что совпадающими в них всех будут только начальные посылки. Более того, если в  $\Gamma$  имеется скрытое противоречие, динамический подход не позволяет даже описать ситуацию, когда агент должен обнаружить это противоречие за  $n$  шагов. Таким образом, он дает только половину решения проблемы ограниченной рациональности, а именно позволяет различить убеждение в  $p$  и убеждение в  $q$ , если  $q$  логически следует из  $p$ . Однако он не позволяет описать структуру ограничений, накладываемых на множество убеждений какими бы то ни было нормами рациональности.

### Семантика с открытыми мирами и моделирование пошагового вывода

Чтобы рассмотреть семантику с невозможными мирами М. Джэго, нам придется начать с его онтологии. Принято считать, что стандартный подход к построению семантики является в целом нейтральным по отношению к онтологии возможных и невозможных миров, которую принимает исследователь. Вне зависимости от избранной точки зрения на то, благодаря чему некоторое предложение языка может быть истинным о некотором мире, условия истинности «работают» одинаково: предложение истинно, когда истинна пропозиция, которую оно означает. Таким образом, истинность нечувствительна по крайней мере к некоторым (эквивалентным, согласно принятой логике) переформулировкам предложений. В случае семантики Джэго дело обстоит совершенно иначе.

С точки зрения М. Джэго, только актуальный мир существует в обычном смысле этого слова. Он состоит из фактов, позитивных и негативных<sup>18</sup>. Все остальные миры состоят из «эрзац-фактов» (термин наш, а не Джэго – А. М.), которые структурно подобны фактам и тоже могут быть позитивными и негативными, однако они представляют собой не настоящие сущие, а лишь интеллектуальные конструкторы, поэтому их истинность в мирах зависит от того, как именно они мыслятся. Каждый факт и «эрзац-факт» представлен в виде некоторого предложения специального *мирообразующего языка*<sup>19</sup>. Этот язык содержит список всех актуально существующих объектов в качестве имен и список всех свойств и отношений, экземплифицированных в актуальном мире, в качестве предикатов. Каждое имя и каждый предикат рассматриваются как обозначающие сами себя; таким образом, актуально истинное предложение буквально представляет собой факт, который оно обозначает. Вместе с тем данный факт структурирован совершенно определенным образом, соответствующим структуре предложения, и при любой переформулировке мы получим уже другой факт, истинность которого может отличаться от истинности первоначального факта. С «эрзац-фактами» ситуация точно такая же: любые переформулировки могут влиять на истинность.

---

<sup>18</sup> Следует сказать, что М. Джэго признает за негативными фактами все признаки существования, включая каузальную силу, конституирующую способность, а также восприимчивость когнитивными агентами.

<sup>19</sup> Факты представляются в виде таких предложений постольку, поскольку они могут быть предметом мысли. Следует помнить, что семантика М. Джэго предназначена в первую очередь для моделирования пропозициональных установок, а не для описания «объективной» части реальности.

Кроме предикатов, имеющих экземпляфикации в актуальном мире, в мирообразующий язык включаются предикаты, которые не экземплифицированы актуально, но экземплифицированы их логические дополнения. Например, свойство «быть кентавром» является логическим дополнением свойства «не быть кентавром», которое экземплифицировано в каждом объекте актуального мира, поэтому мы можем получить «возможного кентавра» (Jago, 2014, 138–144)<sup>20</sup>. Несуществующие партикулярии тоже включаются, будучи определенными как «имена-пучки» (*bundle-names*) – жесткие десигнаторы, обозначающие объект, который существует в некотором возможном мире в качестве уникального носителя заданной совокупности свойств. Как поясняет М. Джэго, имена-пучки являются аналогами выражений с каплановским оператором DTHAT. Так, имя-пучок  $b = \{F_1, \dots, F_n\}$  ведет себя в точности как выражение ‘DTHAT (the  $F_1 \wedge \dots \wedge F_n$ )’ (Jago, 2014, 152)<sup>21</sup>. Таким образом, на мирообразующий язык оказываются адекватно переводимыми все субъект-предикатные конструкции естественного языка вне зависимости от того, описывают ли они нечто актуальное, только возможное или невозможное. Это значит, что можно приступать к построению семантики.

В семантике М. Джэго работает со *структурами* вида  $\langle W, V^+, V^-, \# \rangle$ , где  $W$  – множество миров,  $V^+$  и  $V^-$  – функции, дающие для каждого мира множество мирообразующих предложений, истинных и ложных в этом мире соответственно, а  $\#$  – функция ранга от миров к множеству  $N \cup \{\omega\}$ , которая будет определена ниже (Jago, 2014, 237). Функции  $V^+$  и  $V^-$  заданы независимо друг от друга, так что для любых предложений  $A, B$  и любого мира  $w$  возможна любая из следующих ситуаций:  $A$  истинно в  $w$ , а  $B$  ложно в  $w$ ;  $A$  и  $B$  оба истинны в  $w$ ;  $A$  и  $B$  оба ложны в  $w$ ; ни одно из них ни истинно, ни ложно в  $w$ . Чтобы избежать логического всеведения, М. Джэго изначально отказывается от стратегии наложения на миры условий замкнутости относительно каких бы то ни было логических правил. Иначе говоря, ему требуется множество

---

<sup>20</sup> При таком подходе мы имеем разные репрезентации для разных свойств (посколько их дополнения на всем множестве миров не совпадают: скажем, объект, актуально не являющийся ни кентавром, ни химерой, в мире  $w$  является кентавром, но не химерой, а в мире  $w'$  – наоборот), что отвечает требованию гиперинтенциональности.

<sup>21</sup> В этом месте изложение М. Джэго не вполне ясно. Как можно понять, только возможные и невозможные носители имен являются, с его точки зрения, не настоящими объектами, но конструктами из свойств, на которые можно указывать как на объекты. Скажем, Супермена как индивида не существует ни в одном мире, но имя-пучок «Супермен» является жестким десигнатором, обозначающим такой конструкт. Поэтому можно осмысленно сказать, что Супермен – если бы он существовал – мог бы не уметь летать (Jago, 2014, 157).

открытых миров. Именно поэтому он избирает подход, при котором запрещены любые переформулировки.

Кажется, что на таких «анархических» мирах невозможно моделировать никакую рациональность, однако это ошибочное впечатление. Трюк состоит в следующем: вместо того чтобы определять эпистемическую (или, аналогично, докстическую<sup>22</sup>) возможность мира, исходя из какой-то его внутренней структуры, М. Джэго определяет ее, основываясь на внешних отношениях между этим миром и другими мирами (Jago, 2014, 194). Несколько упрощая, можно сказать, что с точки зрения его подхода, каждый мир представляет собой множество предложений, которые могут составлять знание или убеждение агента на одном конкретном шаге логического вывода. Например, сначала агент рассматривает  $A$  и  $B$ , но не рассматривает их конъюнкцию, а на следующем шаге рассматривает  $A$ ,  $B$  и  $A \wedge B$ . В этом случае мы говорим, что из мира  $w = \{A, B\}$  тривиально выводится<sup>23</sup> мир  $w' = \{A, B, A \wedge B\}$ .

Согласно второй столь же важной для семантики М. Джэго идее, эпистемическая возможность мира зависит от того, какое минимальное количество тривиальных шагов потребуется, чтобы перейти из этого мира в явно противоречивый мир, т. е. мир, содержащий  $A \wedge \sim A$ . Сам М. Джэго дает следующее пояснение: «Чем больше шагов нужно предпринять для раскрытия противоречия, тем тоньше (*subtle*) исходное противоречие. Эпистемически возможными являются миры, в которых противоречия (если они вообще есть) все носят тонкий характер» (Jago, 2014, 200). Чтобы формализовать эту идею, используется семантика, в которой строится *логическая структура на мирах*, выявляющая степень их противоречивости. Для ее построения принимаются специальные структурные правила, сформулированные в стиле секвенциального исчисления, со следующей интерпретацией: множество предложений из левой части секвенции интерпретируется как истинное в некотором мире, а множество предложений из ее правой части – как ложное в том же мире. Сама секвенция означает противоречие: в ней утверждается несовместимость истинности предложений в левой части с ложностью предложений в правой части. Таким образом, вместо истины мы доказываем ложь, а точнее то, что определенный мир противоречив.

---

<sup>22</sup> Все, что будет говориться далее об эпистемической возможности, справедливо и относительно докстической возможности, даже если это явно не указано в тексте.

<sup>23</sup> Понятие тривиальной выводимости (*trivial consequence*) играет большую роль в теории М. Джэго, однако это роль скорее вспомогательного средства, чем цели. Поэтому, чтобы не перегружать статью, мы не будем здесь приводить его определение и свойства, а постараемся, насколько это возможно, обойтись без него. С его полной формализацией можно ознакомиться в (Jago, 2014, 238–240).

Выбирая структурные правила, М. Джэго сначала принимает эквиваленты всех правил стандартного секвенциального исчисления (см. Рис. 1):

$$\begin{array}{c}
 \frac{}{A \vdash A} \text{ IDENTITY} \qquad \frac{\Gamma, A, A \vdash B, B, \Delta}{\Gamma, A \vdash B, \Delta} \text{ CONTRACTION} \\
 \\
 \frac{\Gamma, A, B \vdash C, D, \Delta}{\Gamma, B, A \vdash D, C, \Delta} \text{ EXCHANGE} \qquad \frac{\Gamma \vdash \Delta}{\Gamma, A \vdash B, \Delta} \text{ THINNING} \\
 \\
 \frac{\Gamma, A \vdash \Delta \quad \Gamma \vdash A, \Delta}{\Gamma \vdash \Delta} \text{ CUT}
 \end{array}$$

Рис. 1 (Jago, 2014, 210).

Основным структурным правилом является IDENTITY, которое фактически фиксирует ситуацию явного противоречия (когда некоторое  $A$  одновременно выполняется и не выполняется). Впоследствии оно переформулируется в более общем виде как

$$\frac{}{\Gamma, A \vdash A, \Delta} \text{ IDENTITY}^+,$$

чтобы противоречия сохранялись при переходе от менее полного мира к более полному. Остальные структурные правила при ближайшем рассмотрении оказываются возможным отбросить<sup>24</sup>.

Вывод в такой системе строится снизу вверх, от утверждения, противоречивость которого нужно доказать, и целью является получение какой-то из инстанций IDENTITY<sup>+</sup>. Общий смысл процесса, как поясняется в (Berto, Jago, 2019, 200), состоит в том, что каждый элементарный шаг вывода

$$\frac{\Gamma_1 \vdash \Delta_1}{\Gamma_2 \vdash \Delta_2}$$

---

<sup>24</sup> CONTRACTION и EXCHANGE отбрасываются на основании того, что  $\Gamma$  и  $\Delta$  являются множествами предложений, в которых порядок и повторения не важны. THINNING отбрасывается, когда IDENTITY заменяется на IDENTITY<sup>+</sup>. CUT отбрасывается со ссылкой на доказательство Генценом того, что все выводы, где используется CUT, можно построить и без этого правила ценой увеличения их длины (Jago, 2014, 212).

создает связь между мирами  $w_1$  и  $w_2$  такими, что в  $w_1$  истинно  $\Gamma_1 \cup \Gamma_2$  и ложно  $\Delta_1 \cup \Delta_2$ , а в  $w_2$  истинно  $\Gamma_2$  и ложно  $\Delta_2$ . Группа миров, связанных между собой таким образом, образует направленный граф, являющийся *деревом*<sup>25</sup> (пример см. на Рис. 2). Каждое поддерево этого дерева соответствует какому-то дедуктивному выводу, который может рассматриваться изолированно. Каждый из нижних миров является корнем некоторого поддерева, т. е. множеством посылок вывода, а каждый из верхних миров является листом некоторого количества поддеревьев, т. е. заключением для соответствующих этим поддеревам выводов. Пары соединенных стрелок означают, что два соответствующих шага вывода принадлежат к одному и тому же доказательству, т. е. что их нельзя помешать в отдельные поддеревья.

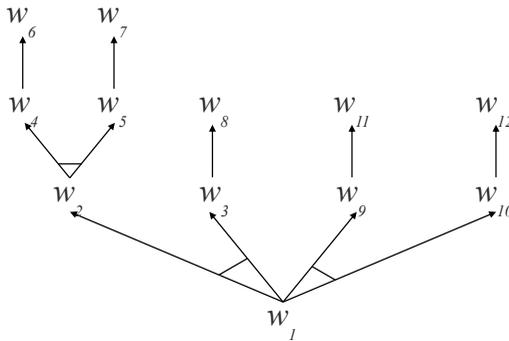


Рис. 2.

Как было сказано выше, завершение любого вывода в данном варианте секвенциального исчисления — это доказательство противоречия. По определению, противоречие является доказанным в поддереве  $T$ , если и только если оно удовлетворяет следующим трем условиям (формулировка из Verto, Jago, 2019, 200; нотация слегка модифицирована):

1. Для каждого мира-листа  $w$  из  $T$ , множества  $V^+w$  и  $V^-w$  перекрываются (т. е. некоторое предложение  $A$  содержится в них обоих).

<sup>25</sup> В теории графов деревом называется связный ациклический граф, в котором между любыми двумя узлами существует ровно один путь.

2. Из каждого мира, не являющегося листом, выходит, как максимум, два ребра.
3. Ребра  $\langle\langle w_1, w_3 \rangle\rangle$  и  $\langle\langle w_2, w_3 \rangle\rangle$  содержатся в  $T$ , только если система доказательств включает инстанциацию некоторого правила

$$\frac{|w_1| \vdash |w_1| \quad |w_2| \vdash |w_2|}{|w_3| \vdash |w_3|},$$

где запись  $|w| \vdash |w|$  означает, что в левой и правой частях секвенции находятся предложения из множеств  $V^+w$  и  $V^-w$  соответственно.

Поскольку каждое такое поддерево доказывает противоречивость мира, находящегося в корне, оно называется «миродоказательством» (*world-proof*) для этого мира. Таким образом, каждый мир, для которого существует миродоказательство, противоречив<sup>26</sup>, причем противоречие является тем более тонким, чем больше длина этого миродоказательства, т. е. количество вершин поддерева, не являющихся листьями. Если длина равна 0, т. е. если корень совпадает с листом, то такой мир явно противоречив; если длина больше, то он скрыто противоречив.

На полученной древесной структуре определяется функция *ранга* #, сопоставляющая каждому миру  $w$  некоторое натуральное число, соответствующее минимальной длине его миродоказательства, если оно существует, либо  $\omega$ , если такое доказательство не завершается, т. е. если мир  $w$  непротиворечив (формальное определение функции # см. в Jago, 2014, 237). Эпистемически возможными в таком случае будут только те миры, ранг которых не меньше некоторого заданного значения  $n$ , которое мы считаем минимальным числом шагов, необходимым для выявления тонкого противоречия. Само значение  $n$  может быть разным в зависимости от того, насколько логически компетентных агентов мы хотим рассматривать, и в значительной степени произвольным. Можно сказать лишь, что для формализации естественных рассуждений разумным является выбор по крайней мере  $n > 1^{27}$ . Таким образом,

---

<sup>26</sup> Это следует из корректности системы правил и того, что миродоказательство строится из инстанциаций этих правил (Jago, 2014, 215).

<sup>27</sup> Никаких других ясных критериев для выбора значения  $n$ , по-видимому, не может быть, поскольку логическая компетентность агентов – вещь крайне неопределенная и вдобавок сильно зависящая от контекста. Ситуация здесь ровно такая же, как в семантике более привычных нечетких предикатов типа «маленький» или «много».

понятие эпистемически возможного мира в данной семантике является *нечетким*.

Кроме структурных правил, в системе доказательств присутствуют логические правила – обычные правила для связок. Чтобы система была полной, т. е. чтобы для каждого противоречивого мира имелось миродоказательство, требуется не менее двух правил на одну связку (Jago, 2014, 216)<sup>28</sup>. М. Джэго принимает такой их набор, который достаточен для задания классической пропозициональной логики со стандартными значениями логических констант, только вместо правил введения и исключения связок он использует «левые» и «правые» правила (см. Рис. 3). Содержательно, каждое из правил предполагает, что если секвенция над чертой представляют собой противоречия, то и секвенция под чертой представляет собой противоречие; и наоборот, если секвенция под чертой не есть противоречие, то хотя бы одна из секвенций над чертой тоже не есть противоречие.

$$\begin{array}{c}
 \frac{\Gamma, \neg A \vdash A, \Delta}{\Gamma, \neg A \vdash \Delta} \neg L \qquad \frac{\Gamma, A \vdash \neg A, \Delta}{\Gamma \vdash \neg A, \Delta} \neg R \\
 \\
 \frac{\Gamma, A \vee B, A \vdash \Delta \quad \Gamma, A \vee B, B \vdash \Delta}{\Gamma, A \vee B \vdash \Delta} \vee L \\
 \\
 \frac{\Gamma \vdash A, B, A \vee B, \Delta}{\Gamma \vdash A \vee B, \Delta} \vee R \qquad \frac{\Gamma, A \wedge B, A, B \vdash \Delta}{\Gamma, A \wedge B \vdash \Delta} \wedge L \\
 \\
 \frac{\Gamma \vdash A, A \wedge B, \Delta \quad \Gamma \vdash B, A \wedge B, \Delta}{\Gamma \vdash A \wedge B, \Delta} \wedge R \\
 \\
 \frac{\Gamma, A \rightarrow B \vdash A, \Delta \quad \Gamma, A \rightarrow B, B \vdash \Delta}{\Gamma, A \rightarrow B \vdash \Delta} \rightarrow L \\
 \\
 \frac{\Gamma, A \vdash B, A \rightarrow B, \Delta}{\Gamma \vdash A \rightarrow B, \Delta} \rightarrow R
 \end{array}$$

Рис. 3 (Jago, 2014, 211).

---

<sup>28</sup> Это следует из полноты такой системы относительно табличной семантики, которая, в свою очередь, следует из того, что система сразу строится как эквивалент стандартного секвенциального исчисления.

Эти логические правила плюс правило IDENTITY<sup>+</sup> являются нормативным базисом того понятия рациональности, которое моделирует М. Джэго. Резюмируя настоящий раздел, содержание этого нормативного базиса можно выразить примерно так: *во-первых*, ни один рациональный агент не должен быть убежден в явном противоречии; *во-вторых*, он не должен быть убежден в предложении, из которого путем менее чем за  $n$  элементарных логических шагов можно получить явное противоречие. Как видно, это далеко от требования логического всеведения и в то же время далеко от логической «анархии». Таким образом, проблема ограниченной рациональности каким-то образом решена.

Однако на вопрос о том, является ли это решение удовлетворительным, трудно ответить, не исследовав свойств той модальной логики, которая может быть построена на основе такого понятия рациональности. До сих пор в явном виде мы этого не сделали. Также можно заметить, что из всего сказанного сих пор не следует никаких выводов о том, в чем *должен* быть убежден (или что *должен* знать) рациональный агент, очерчена лишь область *возможных* знаний и убеждений. В следующем разделе, посвященном непосредственно моделированию понятий знания и убеждения, станет ясно, почему так происходит и в чем состоит основная проблема данного подхода, которую признает и сам М. Джэго. После этого мы сможем, наконец, подытожить наш разбор.

### **Градуированное отношение достижимости и эпистемические недосмотры**

До сих пор мы рассматривали только логическую структуру на мирах, отражающую отношения выводимости между ними. Теперь, поскольку мы приступаем к моделированию понятий знания и убеждения, нам нужно задать на них собственно модальную структуру, т. е. отношения эпистемической и докстической достижимости между ними. В данном разделе мы попытаемся разобраться, каким способом можно задать эти отношения и какими свойствами они должны обладать.

Первая мысль, которая приходит в голову по поводу отношений эпистемической и докстической достижимости, состоит в том, что они должны соответствовать принятому понятию эпистемически возможного мира. Если мы интерпретируем достижимые миры как такие альтернативы, которые рассматривает агент, исходя из своих знаний и убеждений, то очевидно, что ни один эпистемически невозможный мир такой альтернативой не является. Кроме того, исходя из структуры миров, ясно, что оба отношения должны быть интранзитивными и асимметричными, а отношение докстической достижимости должно

быть вдобавок иррефлексивным (а мы помним, что отношение эпистемической достижимости, как правило, задают как рефлексивное). Здесь нет ничего особенно нового, и оба этих тезиса действительно принимаются М. Джэго. Однако на этом очевидные вещи заканчиваются, и начинается относительная свобода построений.

Например, можно перенести на отношения достижимости некоторые из выводов, которые были сделаны нами в процессе исследования логической структуры на мирах. Так, кажется естественным потребовать, чтобы все миры, ранг которых меньше  $n$ , не входили в множества миров, достижимых (в любом из двух смыслов) из любого другого мира. Более того, кажется естественным потребовать, чтобы в множество миров, достижимых из некоторого мира  $w$ , входили по крайней мере те миры, в которые из  $w$  существует одношаговый дедуктивный переход – стрелка. Например, из мира  $w_1$  на Рис. 2 должен был бы быть достижим мир  $w_2$ , коль скоро оба этих мира являются эпистемически возможными (допустим, что для этого случая  $n = 2$ , тогда условия  $\#w_1 \geq n$  и  $\#w_2 \geq n$  выполняются). Однако неочевидно, как эти интуиции обоснованы и почему отношение достижимости не может быть каким-либо другим<sup>29</sup>. Как мы увидим ниже, в семантике, которую строит М. Джэго, они удовлетворяются лишь отчасти.

Ситуация становится еще интереснее, как только мы замечаем, что любые подобные соображения имеют смысл только до тех пор, пока предполагается, что в мирах, о которых мы ведем речь, модальности функционируют хоть каким-то регулярным образом. Ведь формулы с модальностями являются просто одним из видов формул, которые имеют определенные условия истинности и из которых, зная эти условия, мы можем делать определенные выводы. Но если миры, относительно которых мы рассматриваем истинность формул, допускают нарушения логики без модальностей, то у нас есть веские основания полагать, что и в области модальностей могут быть нарушения. Т. е. мы оказываемся способны с какой-то определенностью говорить о логических свойствах модальностей лишь при условии, что миры, относительно которых мы оцениваем истинность модальных высказываний, являются *мирами, возможными в логическом смысле*.

Таким образом, чтобы задать адекватную семантику для модальностей, нужно начинать с определения логически возможного мира. Именно это делает М. Джэго. Согласно неформально данному им ранее определению, мир  $w \in W$  является логически возможным, только если он дедуктивно замкнут относительно правил принятой логики  $L$  и при

---

<sup>29</sup> Кое-что косвенно касающееся второй интуиции будет сказано в конце раздела.

этом не содержит всех предложений мирообразующего языка (Jago, 2014, 159); теперь к этому добавляется требование, чтобы в нем выполнялись стандартные условия истинности для предложений с эпистемической и докстической модальностью. Будучи выражено более строго, это второе требование, если ввести в логику операторы  $K$  и  $B$  для знания и убеждения соответственно, будет заключаться в следующем (Jago, 2014, 241):

- 1) для всех формул вида  $KA$  и  $BA$ , истинных в  $w$ , во всех мирах, достижимых (в эпистемическом и докстическом смысле соответственно) из  $w$ , истинна формула  $A$ ;
- 2) для всех формул вида  $KA$  и  $BA$ , ложных в  $w$ , в некотором мире, достижимом (в эпистемическом и докстическом смысле соответственно) из  $w$ , ложна формула  $A$ ;
- 3) дополнительно выполняется условие фактивности для знания, т. е. если  $KA$  истинно в  $w$ , то и  $A$  истинно в  $w$ <sup>30</sup>.

Как видно, данное понятие логически возможного мира связывает воедино все три части модели, а именно: внутренние свойства миров, логику и отношения достижимости, которые, вообще говоря, зависят не столько от миров, сколько от агентов. Смысл здесь вполне прозрачный: отношения достижимости (достижимость в эпистемическом и докстическом смысле будем обозначать  $R^e$  и  $R^d$  соответственно) должны укладываться в наше представление о том, какие альтернативы может рассматривать рациональный агент в описанной эпистемической (или докстической) ситуации. Так, если в мире  $w$  имеет место  $BA$ , то не должно быть мира  $w'$  такого, что  $wR^d w'$  и при этом в  $w'$  формула  $A$  не является истинной. В обычной семантике возможных миров действует точно такое же требование. Однако здесь следует обратить внимание на один нюанс. В обычной семантике возможных миров формулы всегда либо истинны, либо ложны в мире, поэтому если формула  $A$  не истинна, то она ложна. Отсюда можно сделать вывод, что агент по крайней мере допускает в докстическом смысле  $\sim A$  (иногда такую ситуацию обозначают  $\hat{B}\sim A$ ). Это значимое модальное заключение, которое можно далее использовать в процессе какого-нибудь вывода. Но в семантике М. Джэго многие невозможные миры являются в высокой степени неполными, поэтому в мире, где  $A$  не

---

<sup>30</sup> В невозможных мирах, согласно М. Джэго, истинность и ложность могут при-  
сваиваться формулам с модальностями, как и другим сложным формулам, независимо ни  
от чего.

истинно, оно не обязательно будет ложно. Поэтому никакого значимого заключения из этой ситуации сделать нельзя.

Добавок, если ограничить отношения достижимости множеством тех миров, которые являются эпистемически возможными (или если потребовать, по крайней мере, чтобы из логически возможных миров были достижимы только такие миры), то, поскольку понятие эпистемической возможности является нечетким, сами отношения достижимости тоже станут нечеткими. Между определенно достижимыми и определенно недостижимыми мирами будет лежать обширная «серая» область миров, которые являются достижимыми при меньших значениях  $n$  и недостижимыми при больших. Это, в свою очередь, означает, что, определяя истинность какой-то модальной формулы, скажем,  $KA$ , в каком-то мире, скажем,  $w$ , мы можем столкнуться с ситуацией, когда все миры, в которых  $A$  ложно, попадут в такую «серую» область. В этом случае мы не сможем приписать  $KA$  ни истинность, ни ложность в  $w$ . Иначе говоря, понятия знания и убеждения тоже оказываются у нас нечеткими. В целом, можно признать этот результат вполне закономерным, ведь в жизни, особенно рассматривая сложные вопросы, мы далеко не всегда можем со всей определенностью сказать, в чем состоят наши знания и убеждения на данный счет. Тем не менее, с точки зрения удобства применения теории, было бы предпочтительно иметь возможность хотя бы в большинстве случаев получить какой-то информативный ответ.

Чтобы улучшить свои объяснительные возможности, как представляется, имеет смысл зайти с другой стороны и использовать *градуированное отношение достижимости*. Допустим, у нас имеется агент, логическая компетентность которого нам в точности не известна, и множество предложений  $\Gamma$  таких, что для всех  $A \in \Gamma$  имеется аскрипция  $BA$ , истинность которой мы считаем заданной (например, потому что каждое такое  $A$  есть мирообразующий перевод предложения, которые он сам декларировал в качестве своего убеждения). Тем самым у нас имеется информация, что во всех мирах, доксистически достижимых для этого агента из актуального мира, истинно  $\Gamma$ . Нам не известно, что еще истинно и ложно в этих мирах, поэтому у нас нет полной информации об отношении  $R^d$ , но мы можем принять это отношение как неизвестное и работать с ним в этом качестве. Далее, мы хотим получить информацию о том, рассматривает ли данный агент некоторое предложение  $Z$  как истинное или ложное. Мы можем попытаться это сделать, построив расширенное отношение достижимости  $R_{+1}^d$  путем добавления в гипотетическое множество достижимых миров всех миров, которые содержат множества предложений, выводимые из  $\Gamma$  за

один дедуктивный шаг (исключая те из миров, которые явно противоречивы). После этого мы проверяем, нет ли в получившемся множестве миров, которые содержат  $\sim Z$ . Если такие миры есть, это означает, что из своего первоначального убеждения агент за один дедуктивный шаг мог прийти к допущению  $\sim Z$ . Отсюда с высокой вероятностью следует, что он не убежден в  $Z$ . Если таких миров нет, мы строим отношение  $R_{+2}^d$  по той же схеме и проверяем, нет ли там миров, содержащих  $\sim Z$ , и, находя их, снова делаем заключение, что агент с достаточно высокой (хотя и чуть меньшей, чем в первом случае) вероятностью не убежден в  $\sim Z$ . И т. д.

Описанный прием не используется у М. Джэго, но выглядит логичным развитием принятого им подхода<sup>31</sup>. Кроме того, он задает схему рассуждений, которая представляется очень естественной. В действительности мы во многих случаях так и рассуждаем, если нам нужно вынести какое-то суждение об убеждениях окружающих, но спросить их самих мы не имеем возможности. Конечно, эта схема не является абсолютно надежной, и, если бы от истинности полученного суждения зависело что-то жизненно важное, большинство из нас не стало бы, наверное, на нее полагаться. Однако, поскольку суждения об убеждениях людей представляют собой огромную часть тех данных, на которые мы опираемся ежедневно при принятии простых бытовых решений, совсем исключить подобные рассуждения, по-видимому, не только невозможно, но и не нужно. В большинстве случаев они являются достаточно надежными, особенно если используется небольшое число шагов.

Таким образом, подход М. Джэго можно было бы считать более или менее успешным в деле правдоподобного моделирования знаний и убеждений, хотя получение интересных логических результатов, очевидно, сильно затрудняется тем обстоятельством, что многие используемые понятия являются нечеткими. Однако есть еще одна проблема, с которой необходимо справиться, прежде чем давать этому подходу положительную оценку. Дело в том, что, исходя из имеющихся формализмов, значение модальных формул в мире может быть не определено, даже если эта формула является тривиальной тавтологией! Это значит, что нам придется отказывать нашим агентам в некоторых из тех знаний и убеждений, которыми они, как кажется, точно должны обладать, будучи рациональными существами. Это явление М. Джэго называет *эпистемическим недосмотром* (*epistemic oversight*) (Jago, 2014, 242). С его точки зрения, в самом существовании эпистемических

---

<sup>31</sup> Особенно учитывая последующий результат – теорему 8.3 (см. ниже).

недосмотров нет ничего плохого, так и должно быть, коль скоро мы задали целью избежать логического всеведения; однако в том, что даже тривиальные тавтологии могут попасть в разряд эпистемических недосмотров, он все же видит проблему и пытается найти решение.

Технически то место книги, которое посвящено проблеме тривиальных эпистемических недосмотров, достаточно сложное, и мы не будем здесь излагать весь ход рассуждения, приведем только его схему. М. Джэго начинает с того, что строит понятие *эпистемической модели*, в которой изначально проводится различие между множеством возможных и невозможных миров, а также заданы отношения эпистемической достижимости (для  $k$  различных эпистемических агентов) в виде набора проективных функций  $f_i$  от мира к множеству миров, достижимых из него для  $i$ -го агента. После построения такой модели он вводит множество т. н. *альтернативных последовательностей*. Каждая альтернативная последовательность  $\bar{g}$  состоит из таких вариантов проективных функций  $\bar{g}_i$ , которые заменяют миры, где какая-то теорема принятой логики  $L$  не ложна, мирами, где она истинна. Содержательно альтернативные последовательности представляют собой, согласно М. Джэго, «более отчетливый вариант (*sharpening*) термина "эпистемическая достижимость" для данной модели» (Jago, 2014, 248). Сам набор проективных функций можно понимать как одну из альтернативных последовательностей, из которой получаются все остальные.

Далее, предложения, которые истинны в модели при любом выборе альтернативной последовательности, определяются как истинные в модели в классическом смысле, т. е. в таком, в котором во всякой модели  $M$  имеет место либо  $M \models A$ , либо  $M \models \sim A$ . Дополнительно определяется истинность и ложность в модели относительно альтернативной последовательности  $\bar{g}$  (их обозначения –  $M \models_{\bar{g}} A$  и  $M \models_{\bar{g}} \sim A$  – для нас бесполезны, но все же введем их ради наглядности). Как и обычная истинность и ложность, истинность и ложность в модели относительно  $\bar{g}$  для сложных формул в возможных мирах определяется рекурсивно, а в невозможных – напрямую функциями  $V^+$  и  $V^-$ . Затем определяется понятие истинности в *направленной (pointed)* модели  $M'$ , т. е. в паре из модели и выделенного возможного мира в ней (содержательно соответствует миру оценки выводимых предложений), как истинность в модели относительно альтернативной последовательности, совпадающей с набором проективных функций этой модели. Потом определяется  $M' \models \Gamma$  как истинность в направленной модели  $M'$  всех предложений из  $\Gamma$ .

Далее, в качестве завершения всего предыдущего построения, формулируется понятие *логического  $n$ -следования*  $\Gamma \models_n A$  для всякого  $n \in \mathbb{N}$ , которое имеет место, если и только если, для всякой направленной модели  $M'$  ранга не меньше  $n$ ,  $M' \Vdash \Gamma$ , только если  $M' \Vdash A$ . Как поясняет М. Джэго, понятие логического  $n$ -следования является обобщением классического понятия следования, а также понятия выводимости за  $n$  шагов, которое он использовал, когда рассматривал структурные связи между мирами (полную формализацию последнего мы здесь не приводили). Об этом М. Джэго доказывает теорему, которая у него проходит под номером 8.2. Доказательство данной теоремы позволяет ему утверждать, что дополнительная структура, которая была введена для получения эпистемической модели, «не вмешивается в отношения тривиального следования» (Jago, 2014, 249). Наконец, как главный результат главы, доказывается теорема 8.3, которая гласит: для всякого  $n \in \mathbb{N}$ , если  $A$  выводимо за  $n$  шагов из  $\Gamma$ , то

$$\{\Delta K_i B \mid B \in \Gamma\} \models_n \sim \Delta \sim K_i A,$$

где  $\Delta$  – это оператор, который означает «определено, что».

Говоря простым языком, в теореме 8.3 сказано, что за такое же количество шагов, за которое можно получить  $A$  из  $\Gamma$ , можно доказать, что агент, определенно знающий все предложения из  $\Gamma$ , не является определенно не знающим  $A$ <sup>32</sup>. Это очень важный тезис относительно модальной логики, основанной на использованной М. Джэго семантике с логической структурой на мирах, поскольку он позволяет соотнести свойства данной структуры со свойствами отношения достижимости. Можно сказать, что эта теорема – аналог аксиомы Крипке для данной логики. Хотя она и не позволяет нам делать позитивные суждения о знаниях и убеждениях агентов так, как это позволяет аксиома Крипке, она позволяет ограничивать приписывание агентам незнания. И она имеет вполне прозрачный интуитивный смысл. Действительно, из того, что агент знает все предложения из  $\Gamma$ , и того, что множество этих предложений, рассмотренных совместно, влечет  $A$  за  $n$  тривиальных шагов, мы можем заключить, что агент был *в состоянии* (*in a position*) получить знание  $A$ . В свою очередь, это означает, что в момент оценки он, возможно, уже знает  $A$ .

Из теоремы 8.3 есть частное следствие, касающееся проблемы тривиальных эпистемических недосмотров. Если  $A$  является тривиально

---

<sup>32</sup> Таким образом, наша интуиция о связи между понятиями одношагового перехода и достижимости (в эпистемическом или доксистическом смысле) получила строгое выражение.

истинным предложением, т. е. выводимо за  $n$  шагов из пустого множества предложений, то мы должны, на основании данной теоремы, любому рациональному агенту либо приписывать знание  $A$ , либо говорить, что не определено, имеет ли он это знание. Т. е. хотя мы знаем о существовании эпистемических недосмотров в принципе, у нас никогда не будет достаточных оснований, чтобы приписать конкретному агенту конкретный эпистемический недосмотр относительно такого предложения  $A$ . Рациональность, таким образом, оказывается спасена, хотя абстрактно мы все еще можем предположить, что некоторые рациональные агенты в некоторые моменты времени могут не знать истинность некоторых тривиальных тавтологий<sup>33</sup>.

### Заключение

Семантика с невозможными мирами, как мы увидели, является мощным и бурно развивающимся в последние два десятилетия подходом к моделированию убеждений. Для исследователей привлекательным его делает, с одной стороны, понятность, поскольку большая часть терминологии и методов уже знакома по семантике возможных миров, с другой стороны, правдоподобие выводов, поскольку посредством введения невозможных миров можно избежать чересчур большой идеализации рациональных агентов. Как показал М. Джэго, семантика с невозможными мирами успешно справляется с самой серьезной проблемой семантики возможных миров в этой области – проблемой логического всеведения. Фактически, он «одним выстрелом убил трех зайцев»: метод, который он разработал главным образом для объяснения случая приписываний с необходимо истинными придаточными предложениями, оказался пригоден также для случая с заменой кореферентных имен (головоломки Фреге) и случая с видовыми терминами, обладающими одним и тем же интенционалом. Неясно, справляется ли его семантика в оригинальном варианте со случаем приписываний, содержащих фикциональные имена, столь же успешно, как и с остальными случаями. Однако в последующей совместной с Ф. Берто книге содержится многообещающее развитие его идей касательно этой темы (Berto, Jago, 2019, 239–266).

---

<sup>33</sup> Более того, существование диалектистов как будто показывает, что для некоторых рациональных агентов в некоторых контекстах оказываются допустимыми даже явные противоречия. Правда, сам М. Джэго считает, что это иллюзия, порожденная тем, что диалектисты (по крайней мере, когда они рассматривают определенные предложения) в другом значении используют отрицание (Jago, 2014, 223). Не претендуя здесь на обоснованное мнение по данному вопросу, выразим лишь удивление по поводу того, что, коль скоро они действительно хотя бы в некоторых случаях употребляют не то же самое отрицание, что и все остальные люди, они никогда не говорят об этом прямо.

В то же время следует отметить, что в части логики полученные результаты не настолько впечатляют, как может показаться при беглом знакомстве с темой. Получен важный результат, ограничивающий приписывание агентам эпистемических недосмотров, но это лишь начальный этап создания эпистемической (и докстатической) логики на основе принятой системы доказательств. Поэтому, как представляется, в этой области в ближайшее время следует ожидать дальнейшего развития. Параллельно разрабатываются интересные приложения семантики с невозможными мирами к другим задачам, связанным с моделированием пропозициональных установок. Поскольку в общем вся эта тема характеризуется целостностью и тесной взаимосвязью различных проблем друг с другом, можно надеяться, что семантика и логика убеждений получат также импульс к развитию по мере того, как будет происходить прогресс в этих параллельных областях. В целом, можно заключить, что перспективы данного направления в ближайшее время представляются весьма благоприятными.

### Список литературы

- Хинтиikka, Я. (1980). В защиту невозможных возможных миров. В В. М. Леонтьев (Ред.), Я. Хинтиikka. *Логико-эпистемологические исследования: Сборник избранных статей* (228–244). Москва: Прогресс.
- Berto, F. (2017). Impossible Worlds and the Logic of Imagination. *Erkenntnis*, 82(6), 1277–1297. <https://doi.org/10.1007/s10670-017-9875-5>
- Berto, F. & Jago, M. (2019). *Impossible Worlds*. Oxford: Oxford University Press. <http://dx.doi.org/10.1093/oso/9780198812791.001.0001>
- Bjerring, J. & Skipper, M. (2019). A dynamic solution to the problem of logical omniscience. *Journal of Philosophical Logic*, 48(3), 501–521. <http://dx.doi.org/10.1007/s10992-018-9473-2>
- Cohen, L. J. (1992). *An Essay on Belief and Acceptance*. Oxford: Clarendon Press. <http://dx.doi.org/10.1093/acprof:oso/9780198236047.001.0001>
- Hintikka, J. (1962). *Knowledge and Belief: An Introduction to the Logic of the Two Notions*. Ithaca, New York: Cornell University Press.
- Jago, M. (2007). Hintikka and Cresswell on Logical Omniscience. *Logic and Logical Philosophy*, (15), 325–354. <http://dx.doi.org/10.12775/LLP.2006.019>
- Jago, M. (2014). *The Impossible: An Essay on Hyperintensionality*. Oxford: Oxford University Press. <http://dx.doi.org/10.1093/acprof:oso/9780198709008.001.0001>

- Lewis, D. (1975). Language and languages. In K. Gunderson (Ed.), *Language, Mind and Knowledge* (3–35). Minneapolis: University of Minnesota Press.
- Priest, G. (2005). *Towards Non-Being: The Logic and Metaphysics of Intentionality*. Oxford: Oxford University Press.  
<https://doi.org/10.1093/0199262543.001.0001>
- Priest, G. (2006). In *Contradiction: A Study of the Transconsistent* (2nd expanded ed.). Oxford: Oxford University Press.  
<https://doi.org/10.1093/acprof:oso/9780199263301.001.0001>
- Rasmussen, M. (2015). Dynamic epistemic logic and logical omniscience. *Logic and Logical Philosophy*, (24), 377–399.  
<http://dx.doi.org/10.12775/LLP.2015.014>
- Salmon, N. U. (1986). *Frege's Puzzle*. Cambridge: MIT Press.
- Schiffer, S. (1992). Belief Ascription. *The Journal of Philosophy*, 89(10), 499–521. <http://dx.doi.org/10.2307/2941005>
- Schiffer, S. (1978). The Basis of Reference. *Erkenntnis*, 13(1), 171–206.  
<http://dx.doi.org/10.1007/BF00160893>
- Stalnaker, R. C. (1984). *Inquiry*. Cambridge: MIT Press, Bradford Books.
- Stalnaker, R. C. (1976). Propositions. In A. MacKay & D. Merrill (Eds.), *Issues in the Philosophy of Language* (79–91). New Haven: Yale University Press.

Информация об авторе: Моисеева Анна Юрьевна, кандидат философских наук, научный сотрудник Института философии и права Сибирского отделения Российской академии наук, Новосибирск, [ajumo@yandex.ru](mailto:ajumo@yandex.ru)

**Moiseeva Anna Yurievna**

## **Impossible Worlds in Belief Modeling**

**Abstract:** This article discusses one approach to belief modeling that has been developed to overcome the problems of the standard possible world semantics, namely the semantics with impossible worlds. It is shown that the introduction of impossible worlds in itself does not allow coping with these problems, as long as the worlds remain closed with respect to some accepted logic. The essence of the problem of modeling bounded rationality is revealed and the method of its solution in the semantics developed by M. Jago and later developed by him together with F. Berto is characterized. It is concluded that the task was indeed solved in the semantics of M. Jago. However, the

peculiarities of this semantics made it difficult to build epistemic and doxastic logics on its basis and the subsequent proof of theorems. Nevertheless, the prospects for this approach and, in general, the semantics of beliefs using impossible worlds are assessed as favorable against the backdrop of significant success and rapid development of this whole trend in recent times.

Key words: impossible worlds, semantics and logic of beliefs, hyperintensionality.

## References

- Hintikka, J. (1980). V zaschitu nevozmozhnyh vozmozhnyh mirov [Impossible possible worlds vindicated]. In V. M. Leontyev (Ed.), *Hintikka J. Logiko-epistemologicheskie issledovaniya* [Hintikka J. Logical-Epistemological Studies] (228–244). Moscow: Progress. (in Russian).
- Berto, F. (2017). Impossible Worlds and the Logic of Imagination. *Erkenntnis*, 82(6), 1277–1297. <https://doi.org/10.1007/s10670-017-9875-5>
- Berto, F. & Jago, M. (2019). *Impossible Worlds*. Oxford: Oxford University Press. <http://dx.doi.org/10.1093/oso/9780198812791.001.0001>
- Bjerring, J. & Skipper, M. (2019). A dynamic solution to the problem of logical omniscience. *Journal of Philosophical Logic*, 48(3), 501–521. <http://dx.doi.org/10.1007/s10992-018-9473-2>
- Cohen, L. J. (1992). *An Essay on Belief and Acceptance*. Oxford: Clarendon Press. <http://dx.doi.org/10.1093/acprof:oso/9780198236047.001.0001>
- Hintikka, J. (1962). *Knowledge and Belief: An Introduction to the Logic of the Two Notions*. Ithaca, New York: Cornell University Press.
- Jago, M. (2007). Hintikka and Cresswell on Logical Omniscience. *Logic and Logical Philosophy*, (15), 325–354. <http://dx.doi.org/10.12775/LLP.2006.019>
- Jago, M. (2014). *The Impossible: An Essay on Hyperintensionality*. Oxford: Oxford University Press. <http://dx.doi.org/10.1093/acprof:oso/9780198709008.001.0001>
- Lewis, D. (1975). Language and languages. In K. Gunderson (Ed.), *Language, Mind and Knowledge* (3–35). Minneapolis: University of Minnesota Press.
- Priest, G. (2005). *Towards Non-Being: The Logic and Metaphysics of Intentionality*. Oxford: Oxford University Press. <https://doi.org/10.1093/0199262543.001.0001>
- Priest, G. (2006). *In Contradiction: A Study of the Transconsistent* (2nd expanded ed.). Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199263301.001.0001>

- Rasmussen, M. (2015). Dynamic epistemic logic and logical omniscience. *Logic and Logical Philosophy*, (24), 377–399. <http://dx.doi.org/10.12775/LLP.2015.014>
- Salmon, N. U. (1986). *Frege's Puzzle*. Cambridge: MIT Press.
- Schiffer, S. (1992). Belief Ascription. *The Journal of Philosophy*, 89(10), 499–521. <http://dx.doi.org/10.2307/2941005>
- Schiffer, S. (1978). The Basis of Reference. *Erkenntnis*, 13(1), 171–206. <http://dx.doi.org/10.1007/BF00160893>.
- Stalnaker, R. C. (1984). *Inquiry*. Cambridge: MIT Press, Bradford Books.
- Stalnaker, R. C. (1976). Propositions. In A. MacKay & D. Merrill (Eds.), *Issues in the Philosophy of Language* (79–91). New Haven: Yale University Press.

Author Information: Moiseeva Anna Yurievna, Candidate of Philosophy, Researcher, Institute of Philosophy and Law of the Siberian Branch of Russian Academy of Sciences, Novosibirsk, [ajumo@yandex.ru](mailto:ajumo@yandex.ru)