Вестник Томского государственного университета. Философия. Социология. Политология. 2024. № 77. С. 32–52.
Tomsk State University Journal of Philosophy, Sociology and Political Science. 2024. 77. pp. 32–52.

Научная статья
УДК 16
doi: 10.17223/1998863X/77/3

ЛОГИКА ЭФФЕКТОВ ФРЕЙМИНГА Ф. БЕРТО И А. ОЗГЮН – НОВЫЙ ФОРМАЛИЗМ ДЛЯ РЕШЕНИЯ ПРОБЛЕМ СЕМАНТИКИ ПРОПОЗИЦИОНАЛЬНЫХ УСТАНОВОК

Анна Юрьевна Моисеева

Русское общество истории и философии науки, Москва, Россия, abyssian03@gmail.com

Аннотация. Логика эффектов фрейминга Ф. Берто и А. Озгюн – это доксастическая логика, разработанная ее авторами как способ формального описания так называемых эффектов фрейминга, которые не описываются корректно в рамках стандартного языка семантики возможных миров, что составляет один из аспектов проблемы логического всеведения. В настоящей статье, после содержательного введения в проблему, рассматривается семантика и аксиоматика данной логики, а также то, какими возможностями для решения проблемы эффектов фрейминга, в частности проблемы логического всеведения вообще, располагает этот формализм. Некоторые примечательные в контексте обозначенной проблематики свойства формально доказываются и философски комментируются.

Ключевые слова: содержание убеждений, семантика возможных миров, доксастическая логика, эффекты фрейминга, логическое всеведение, моделирование ограниченной рациональности.

Благодарности: подготовлено при поддержке РНФ, проект № 21-18-00496.

Для цитирования: Моисеева А.Ю. Логика эффектов фрейминга Ф. Берто и А. Озгюн – новый формализм для решения проблем семантики пропозициональных установок // Вестник Томского государственного университета. Философия. Социология. Политология. 2024. № 77. С. 32–52. doi: 10.17223/1998863X/77/3

Original article

THE LOGIC OF FRAMING EFFECTS BY FRANZ BERTO AND AYBÜKE ÖZGÜN AS A NEW FORMALISM FOR SOLVING PROBLEMS OF SEMANTICS OF PROPOSITIONAL ATTITUDES

Anna Yu. Moiseeva

Russian Society for the History and Philosophy of Science, Moscow, Russia, abyssian03@gmail.com

Abstract. The article discusses the problems of formalizing the content of propositional attitudes and how successfully these problems can be solved in possible worlds semantics. The focus of attention is, firstly, on the phenomena that in the psychological literature are called framing effects and in the semantic literature substitution violation in indirect contexts, and, secondly, the problem of logical omniscience. The first part of the article explains why framing effects create a problem for possible worlds semantics and why the agent is inevitably modeled as logically omniscient in this semantics. Next, a description is given of four approaches to the problem of framing effects in possible worlds semantics: metasemantic, pragmatic, multi-model, and the approach associated with modeling topics. For each of the first three, their main shortcomings are given in relation to this problem and

in general. The results of applying the latter approach are considered in more detail in the next part of the article using the example of one of the modern systems of doxastic logic, namely the logic of framing effects by Berto and Özgün. The syntax and semantic rules of this logic are outlined, after which it is explained how, according to these rules, the description of framing effects should look like. Models of situations are given in which an agent has both a belief in some proposition and a lack of belief in another proposition, which is necessarily equivalent to the first one. It is shown that the modeling method used does not lead to problems and is intuitively adequate. In the last part, the axiomatics of the logic of framing effects is presented and some theorems are given that are significant in the context of the problem of logical omniscience. An interpretation of these theorems is given, on the basis of which it is concluded that this logic does not cope with the problem of logical omniscience as successfully as with the framing effects themselves. In conclusion, a direction is proposed in which the logic of framing effects can be developed in order to more fully solve the problem of logical omniscience with its help, and the prospects for such development are discussed.

Keywords: content of beliefs, possible worlds semantics, doxastic logic, framing effects, logical omniscience, modeling of bounded rationality

Acknowledgments: The paper is supported by the Russian Science Foundation, Project No. 21-18-00496.

For citation: Moiseeva, A.Yu. (2024) The logic of framing effects by Franz Berto and Aybüke Özgün as a new formalism for solving problems of semantics of propositional attitudes. *Vestnik Tomskogo gosudarstvennogo universiteta. Filosofiya. Sotsiologiya. Politologiya – Tomsk State University Journal of Philosophy, Sociology and Political Science.* 77. pp. 32–52. (In Russian). doi: 10.17223/1998863X/77/3

Введение в проблематику

Трудность задачи формализации содержания пропозициональных установок состоит в том, что, как давно доказали психологи (см., например: [1, 2]), склонность агента признавать истинность пропозиции в значительной степени зависит от того, в каком контексте и в какой форме высказана эта пропозиция. В частности, два предложения, являющихся логически эквивалентными, могут получить различную оценку агента, если одно из них связано с темой, о которой агент размышлял недавно или любит размышлять, а другое – нет. Например, утверждение «Вероятность выживания пациента через месяц после операции составляет 90%» воспринимается как более весомое основание рекомендовать эту операцию, чем утверждение «Смертность в течение одного месяца после операции составляет 10%». Обычно (когда нет особых причин рассуждать иначе) эксплицитная оценка агентом предложения рассматривается как достаточное условие для приписывания ему пропозициональной установки, содержанием которой является пропозиция, выраженная в этом предложении. Таким образом, в подобных случаях имеются как будто достаточные условия для приписывания агенту двух несовместимых пропозициональных установок с одним и тем же содержанием. Эту проблему в настоящей статье я буду называть проблемой эффектов фрейминга (так подобные эффекты называются в психологической литературе[1]).

[1] В формальной семантике данная проблема более известна как «проблема нарушения подстановочности в косвенных контекстах» или «головоломка Фреге». Под каждым из этих названий данная проблема имеет свою собственную историю и свою собственную специфическую постановку. Существует также так называемая проблема логического всеведения, которая является отчасти аспектом проблемы эффектов фрейминга, а отчасти самостоятельной проблемой. О проблеме логического всеведения также пойдет речь в настоящей статье.

В контексте формальной семантики и логики суть проблемы эффектов фрейминга можно представить следующим образом. В логиках, использующих стандартную семантику возможных миров для моделирования содержания пропозициональных установок (представленной, например, в [3]), это содержание моделируется посредством множества возможных миров, достижимых в определенном смысле для данного агента. Например, если моделируется содержание убеждений, то используется отношение доксастической достижимости.[1] Каждое убеждение агента делит множество всех возможных миров на те, которые «согласны» с этим убеждением, и те, которые «не согласны» с ним; в дальнейшем для простоты я буду говорить, что на первом подмножестве миров содержание этого убеждения истинно, а на втором – ложно[2]. Информативность предложения моделируется в этой семантике как доля миров, которые исключаются из числа доксастически достижимых для данного агента при принятии им «на веру» содержания этого предложения, по причине того что содержание этого предложения ложно в этих мирах.

Такой способ моделирования имеет своим следствием то, что логически эквивалентные предложения могут быть взаимозаменяемы в любых приписываниях агенту пропозициональных установок (для определенности я по-прежнему буду говорить об убеждениях) salva veritate – ведь их содержание истинно на одном и том же множестве миров, а значит, они не исключают никакие миры из числа достижимых. Иначе говоря, для всякой пропозиции *p*, агент должен или иметь все убеждения, содержанием которых является *p*, или не иметь ни одного из них. Однако, как показывают психологические исследования и даже повседневный опыт, это не так, поскольку существуют эффекты фрейминга. Далее, если агент убежден в некоторой пропозиции, он должен быть убежден во всех ее логических следствиях[3]; и также он должен быть убежден во всех логических тавтологиях – ведь тавтологии следуют из пустого множества пропозиций. Это свойство получило в литературе название логического всеведения. Из логического всеведения следует, что ни один логический вывод не является информативным, поскольку агент, убежденный в посылках, всегда убежден и в заключении. Излишне говорить, что реальные агенты такого свойства не демонстрируют.

За немалую историю развития семантики пропозициональных установок сформировалось достаточно большое количество подходов и программ, претендующих на то, чтобы решить или снять проблему эффектов фрейминга, а также связанную с ней проблему логического всеведения. Существуют отдельные направления семантики, при разработке которых изначально имелись в виду эти задачи (в качестве примера можно назвать ситуационную семантику Дж. Барвайса и Дж. Перри.) В настоящей статье такие направления рассматриваться не будут. Я сконцентрируюсь исключительно на том, как данные проблемы можно решить средствами семантики возможных миров. В первой части статьи кратко будут рассмотрены некоторые старые подходы

[1] То, что мир является доксастически достижимым, означает, что в нем выполняется все, в чем убежден агент. Соответственно, то, что мир является доксастически недостижимым, означает, что в нем выполняется что-то, противоречащее некоторому убеждению агента.

[2] При условии, что есть только два значения истинности. В семантике возможных миров может использоваться значение истинности «не определено», но в данном случае я игнорирую эту возможность.

[3] Это свойство я буду называть замкнутостью системы убеждений относительно необходимой импликации.

к объяснению эффектов фрейминга, существующие в семантике возможных миров, а потом покажу, как она решается в семантике и логике, недавно разработанных Ф. Берто и А. Озгюн специально для этой цели [4]. В конце статьи будут приведены и проинтерпретированы несколько теорем логики эффектов фрейминга в контексте вопроса о том, может ли данная логика столь же успешно справляться с проблемой логического всеведения, как она справляется с эффектами фрейминга. В качестве базовой пропозициональной установки, на примере которой будут рассмотрены интересующие меня вопросы, будет использоваться убеждение, поскольку именно оно моделируется в семантике Берто и Озгюн.

Четыре подхода к проблеме эффектов фрейминга

В рамках семантики возможных миров сформировалось несколько различных подходов к решению или снятию проблемы эффектов фрейминга. Один из первых подходов, который я буду называть *метасемантическим*, представлен в ранних работах Р. Сталнэкера. Он полагает, что различия в обыденных оценках приписываний, в которых используются логически эквивалентные предложения, объясняются недопониманием в области семантики этих предложений. То же самое происходит и с приписываниями, которые обычно оцениваются как ложные, несмотря на то что пропозиция, убеждение, которые приписываются агенту, являются логической тавтологией. Он пишет: «Кажущаяся неспособность видеть, что пропозиция необходимо истинна или что пропозиции необходимо эквивалентны, должна объясняться неспособностью видеть, какие пропозиции высказываются в данных выражениях» [5]. Поскольку Сталнэкер – сторонник диспозиционализма в трактовке пропозициональных установок (см., например, [6]), это действительно кажется простым и логичным объяснением. Если пропозициональное содержание установки соответствует диспозиции агента вести себя определенным образом, в частности, соглашаться с предложениями, которые он понимает как выражающие данную пропозицию, и если в определенном случае эта диспозиция не проявляется, это может объясняться тем, что агент просто не понимает или не полностью понимает данное предложение.

Сам феномен семантического недопонимания хорошо известен: имея дело со сложным по структуре предложением, мы часто не можем уловить смысл, хотя знаем, что означает каждое слово в отдельности. В классической литературе можно найти множество таких предложений. Подобный эффект возникает часто при восприятии структурно сложных предложений логики и математики. Однако не все предложения математики, истинность которых неочевидна, по крайней мере, некоторым агентам, сформулированы таким образом. В качестве примера можно привести теорему Ферма: предложение «$x^n + y^n = z^n$ для $n > 2$ не имеет решений в целых положительных числах» сформулировано на языке, хорошо знакомом каждому по школьному курсу математики, и имеет достаточно простую структуру, чтобы быть понятным, однако это не дает нам «автоматического» знания, что данное предложение истинно. Да и простые школьные задачи типа «$\sqrt{1089} = ?$»[1] не вызывали бы ни у кого затруднений, если бы истинность математических предложений,

[1] Пример из работы [7].

простых по структуре, могла непосредственно усматриваться нами. Подобные примеры показывают, что метасемантических рассуждений как минимум недостаточно, чтобы объяснить, в чем состоит информативность необходимо истинных предложений, а как максимум они вообще не имеют отношения к делу.

Некоторые авторы предполагают, что в обыденной речи мы оцениваем приписывание пропозициональных установок не только на истинность, но и на соответствие прагматическому требованию, чтобы это приписывание было сформулировано «максимально близко к собственным словам агента, если нет особых причин для отклонения от них» [8]. Такой подход можно было бы назвать *прагматическим*, и он широко используется в расселианстве и неорасселианстве. В неорасселианстве предложения (естественного языка или внутреннего кода) выступают в качестве «масок» (guices) пропозиций (пример такого анализа см. у Н. Сэлмона [9]), и фраза «*x* убежден, что *p*», где *x* – имя агента, *p* – пропозиция, рассматривается как утверждение, что существует такое предложение, выражающее пропозицию *p*, с которым агент *x* согласен. Соответственно, «*x* не убежден, что *p*»[1], рассматривается как утверждение, что существует такое предложение – «маска» пропозиции *p*, – с которым агент не согласен. Формально это выглядит следующим образом:

$$B(x, p) := (\exists g)\,[BEL(x, p, g) \wedge G(x, p, g)],$$

где *B* – условие истинности приписывания агенту *x* убеждения в пропозиции, что *p*; *BEL* – отношение убежденности агента в пропозиции под некоторой «маской» *g*; *G* – отношение «схватывания» (grasping) агентом пропозиции под этой «маской».

Такой анализ позволяет согласовать между собой приписывания, в которых агенту приписываются одновременно две противоположных доксастических установки с одним и тем же содержанием, например, утверждать одновременно «Вася убежден, что количество учеников в классе меньше 33» и «Вася не убежден, что количество учеников в классе меньше √1089». Затруднение здесь устраняется тем, что формулы

$$(\exists g)\,[BEL\,(x, p, g) \wedge G\,(x, p, g)]$$

и

$$(\exists g)\,[\neg BEL\,(x, p, g) \wedge G\,(x, p, g)]$$

не противоречат друг другу.

У прагматического подхода есть проблемы в объяснении некоторых естественных выводов с использованием приписываний пропозициональных установок. Например, представим себе ситуацию, когда Маша, подруга Васи, которая в курсе его убеждений касательно количества учеников, думает о

[1] Более точно, Сэлмон использует предикат «воздерживается от убеждения, что *p*» (withholds from belief that *p*), который позволяет ему формализовать такие предложения способом, отличным от простого отрицания предложений вида «*x* убежден, что *p*» [9. P. 111]. За контринтуитивность этого решения применительно к некоторым контекстам его критикует С. Шиффер, который взамен предлагает свою версию прагматического подхода, известную как теория скрытых индексикалов [10, 11]. Теория скрытых индексикалов избегает наиболее очевидных проблем теории «масок», однако она имеет собственные проблемы, самой серьезной из которых, на мой взгляд, является необходимость «вчитывать» в содержание приписываний компоненты, призванные характеризовать контекст употребления агентом того предложения, с помощью которого передается его убеждение в данном приписывании. При этом неочевидно, каким именно из (потенциально бесконечного) множества возможных способов этот контекст должен характеризоваться, а значит, содержание приписывания всегда остается недоопределенным. Подробнее об этом см.: [12].

них. Естественно предположить, что Маша будет использовать те самые формулировки, которые указаны выше. Однако Маша, в отличие от Васи, знает и помнит о том, что $\sqrt{1089} = 33$, поэтому она может заключить о тождестве пропозиции, что количество учеников в классе меньше $\sqrt{1089}$, и пропозиции, что количество учеников в классе меньше 33. Исходя из условий примера, у нас есть все основания сказать о ней «Маша убеждена, что Вася убежден, что *p*» и одновременно «Маша убеждена, что Вася не убежден, что *p*». Конъюнкция таких приписываний как будто ставит под сомнение рациональность уже не Васи, а Маши. Эта проблема, носящая название проблемы итерированных приписываний, а также другие проблемные случаи подробно обсуждаются в развернутой по этому поводу полемике между С. Шиффером и Н. Сэлмоном [9, 13–15].

Еще один подход пытается справиться с проблемой эффектов фрейминга на уровне модели. Вместо одной структуры на мирах предлагается рассмотреть альтернативные структуры, различающиеся между собой валюацией пропозициональных переменных. Каждая валюация соответствует какому-то одному способу мыслить из используемых агентом, что можно назвать предметной или концептуальной рамкой. Такой способ моделирования позволяет считать агента, игнорирующего логическую эквивалентность предложений, рациональным в локальном смысле – в пределах какой-то одной модели. Я буду называть этот подход *мультимодельным*. Одним из сторонников мультимодельного подхода является Д. Льюис, который приводит в качестве пояснения того, как работает данный подход, следующий пример: «Раньше я был убежден, что улица Нассау проходит примерно с востока на запад; что железная дорога поблизости идет примерно с севера на юг; и при этом что они примерно параллельны друг другу. …Так что каждое предложение в неконсистентной тройке было истинным, согласно моим убеждениям, но неверно, что все вместе были истинными, согласно моим убеждениям. …Моя система убеждений была разбита на (перекрывающиеся) фрагменты. В разных ситуациях активировались разные фрагменты, и никогда не проявлялась вся система убеждений сразу» [16. P. 436].

Главным недостатком мультимодельного подхода, на мой взгляд, является то, что он не способен описать систему убеждений за рамками каждого отдельного фрагмента, т.е. рассмотреть агента как целое. Между тем иногда полезно знать, в каком доксастическом состоянии был бы агент, если бы активировал все свои убеждения вместе. Кроме того, в некоторых ситуациях естественно предполагать, что все имеющие отношение к вопросу убеждения активированы, но ответ на вопрос, логически следующий из них, все равно оказывается не получен агентом. В качестве примера можно привести следующую ситуацию[1]: в контексте некоторой шахматной партии агент рассматривает позицию и выбирает ход. Предположим, что он полностью знает правила игры, имеет цель выиграть и способен рационально действовать для достижения этой цели. Предположим также, что он видит все возможные варианты ходов, в том числе ход Qe7, который является началом стратегии, приводящей в его случае к гарантированному выигрышу[2]. Однако он делает другой ход и проигрывает. Эту ситуацию, аналоги которой сплошь

[1] Пример из работы [17].

[2] Выигрышность стратегии в шахматной партии всегда можно просчитать математически.

и рядом встречаются в жизни, нельзя объяснить иначе как незнанием агентом того, что выигрышная стратегия начинается с хода Qe7. Между тем все условия, необходимые для получения им соответствующего знания, были выполнены.

Последний подход из тех, что я буду рассматривать, и наиболее перспективный, на мой взгляд, связан с *моделированием тематики* предложений, рассматриваемых агентом посредством введения в семантику специальных элементов – собственно тем или топиков. С формальной точки зрения топик – что-то вроде маски пропозиции, но это элемент теории значения, а не языка, теории сознания или онтологии (хотя есть подходы, в которых тематичность рассматривается как элемент онтологии – подробнее см., например, [18]). Подход, использующий топики в моделировании содержания пропозициональных установок, реализован в [19] применительно к логике воображения, в [20] применительно к эпистемической логике и в [4] непосредственно для моделирования содержания убеждений.

Сама по себе идея, что тематика должна как-то учитываться при моделировании значения предложения, не является новой. В неформальном виде она встречается уже у К. Гемпеля при обсуждении его знаменитого «парадокса ворона». Действительно, с точки зрения условий истинности предложений «Все вороны черные» и «Ни одна нечерная вещь не является вороном» эквивалентны, однако они верифицируются по-разному, и мы интуитивно будем склонны считать, скажем, коричневый ботинок подтверждением второго, но не подтверждением первого. Сам Гемпель делает по этому поводу следующее предположение: «Возможно, впечатление парадоксальности [таких случаев] можно назвать вырастающим из ощущения, что гипотеза о том, что *все вороны черные*, говорит *о воронах*, а не о нечерных вещах и не обо всех вещах [вообще]» (цит. по: [18]; курсив мой. – *А.М.*). Отсюда можно сделать вывод, что тематика воспринимается как один из аспектов значения, по крайней мере, в некоторых случаях.

Рассуждая более строго, в значении предложения можно различить информационное содержание предложения – грубую или «толстую» (thick) пропозицию – и его эпистемическое содержание – «тонкую» (thin) пропозицию. «Тонкая» пропозиция интуитивно представляется как такая, которая «складывается» из тематики предложения и его логической структуры[1]. Предложения могут различаться по структуре, но иметь одну тематику и наоборот. Например, предложения «Снег бел» и «Неверно, что снег не бел» говорят о белизне снега, и оба они утверждают об этой белизне снега одно и то же – то, что она имеет место[2]. Однако структура второго предложения сложнее, поскольку она включает два отрицания, которых в первом предложении нет. Если же рассмотреть предложения «Снег бел» и «Снег черен», то видно, что структура у них одна и та же, но второе предложение говорит уже о черноте снега. Логически эквивалентные предложения так же могут иметь разную структуру и/или разную тематику. Получается, что предложения p и q

[1] В случае пропозициональной логики анализ структуры не идет дальше простых предложений, в случае логики первого порядка структура атомарные пропозиции также рассматриваются как структурированные.

[2] Опять же, это справедливо до тех пор, пока рассматриваются только модели, в которых всякая пропозиция является либо истинной, либо ложной.

должны считаться эквивалентными по своему эпистемическому содержанию всегда и только тогда, когда у них одна и та же тематика, либо одна и та же структура, либо такие структуры, установить логическую эквивалентность которых сможет любой рациональный агент. Именно такая эквивалентность, по мысли Ф. Берто и А. Озгюн [4], является основанием для того, чтобы ожидать от агента, убежденного в p, что он также будет убежден и в q, и наоборот.

Следующие части статьи посвящены анализу объяснительных возможностей логики эффектов фрейминга, представленной в [4], применительно к проблемам эффектов фрейминга и логического всеведения в семантике пропозициональных установок. Сначала будут изложены формальные основы этой логики, начиная с ее семантики, и показаны на примерах, как она моделирует убеждения ограниченно рациональных агентов. Далее я перейду к аксиоматике логики эффектов фрейминга и доказательству в этой аксиоматике некоторых теорем, имеющих философское значение в контексте рассматриваемых проблем. В заключении я подведу итог сказанному и затрону вопрос о том, какие перспективы развития есть у этой логики для того, чтобы расширить свои возможности на те случаи, которые она пока не в состоянии объяснить.

Логика эффектов фрейминга. Семантические правила и объяснение

Язык L, использующийся в логике эффектов фрейминга Ф. Берто и А. Озгюн, включает стандартный язык модальной пропозициональной логики и два пропозициональных оператора B_A и B_P для «активных» и «пассивных» убеждений. Активным убеждение называется тогда, когда оно актуализировано, т.е. его содержание «загружено» из долговременной памяти в рабочую память агента. Пассивным, соответственно, называется убеждение, содержание которого находится только в долговременной памяти, но не в рабочей. Семантически рабочая память агента в разных состояниях моделируется так называемыми *ячейками памяти* (см. ниже).

Семантика логики эффектов фрейминга основана на стандартной модели семантики возможных миров, в которую Берто и Озгюн добавляют некоторые новые компоненты, необходимые для конструирования ячеек памяти и корректной работы с тематикой пропозиций. В итоге в их семантике всякая модель M представляет собой кортеж вида

$$<W, \Theta, T, \oplus, t, v>,$$

где W – непустое множество миров; Θ – непустое конечное множество таких O, что $O \subseteq W$, и $\Theta \neq \{\emptyset\}$; T – конечное непустое множество тем или топиков; $\oplus$ – бинарная операция слияния (fusion) на T, индемпотентная, коммутативная и ассоциативная; t: $Prop \cup \Theta \rightarrow T \cup 2^T$, где $Prop := \{p_1, p_2 \ldots\}$ – функция назначения топиков, такая, что каждый элемент счетного множества атомарных пропозиций $Prop$ она отображает на какой-то элемент множества топиков T, а каждый элемент Θ – на какое-то подмножество T; v: $Prop \rightarrow 2^W$ – функция валюации[1].

[1] Обозначения несколько модифицированы для удобства набора, но только буквенные, все остальные особенности нотации Берто и Озгюн сохранены без изменений.

Область определения функции t распространяется на весь язык следующим образом:

$$t(\varphi) = t(p_1) \oplus \ldots \oplus t(p_n),$$

где $\{p_1, \ldots p_n\} = \mathrm{Var}(\varphi)$ – множество переменных формулы[1] φ.

Далее определяется отношение части на топиках:

$$\forall a, b\ (a \sqsubseteq b, \text{ е.т.е. } a \oplus b = b).$$

Отношение $\sqsubseteq$ представляет собой частичный порядок. Интуитивно его можно представлять как отношение подразумевания между различными компонентами мыслительного содержания: так, обладание любой мыслью о белизне или черноте снега подразумевает обладание мыслью о снеге как таковом. Множество T вместе с отношением $\sqsubseteq$ представляет собой верхнюю полурешетку.

Ячейки памяти имеют обозначения вида O_a, где O – непустой элемент Θ, a – элемент $t(O)$. Истинность всех формул оценивается относительно мира $w \in W$ и ячейки памяти O_a. Интуитивно множество O можно понимать по аналогии с множеством доксастически достижимых миров в стандартной семантике возможных миров для логики убеждений, но с той разницей, что здесь имеется несколько таких множеств, соответствующих разным состояниям рабочей памяти. (Как видим, стандартного отношения достижимости в модели нет.) Все пропозиции, в которых активно убежден агент в соответствующем состоянии, истинны на всем множестве O, но не обязательно все пропозиции, которые истинны на O, являются содержанием активных убеждений агента в данном состоянии. По смыслу ситуация, когда пропозиция p истинна на O, но ее топик не является частью a, соответствует тому, что агент в этом состоянии просто не думает о том, о чем говорит эта пропозиция, а значит, ему в этом состоянии нельзя истинно приписать активное убеждение, что p. Для моделирования пассивных убеждений используется специальная квазиячейка памяти, представляющая собой пару из такого множества, что на нем истинны все пропозиции, в которых агент активно убежден, и такого топика, что его частью являются все топики, о которых агент думает хотя бы в одном из своих состояний (см. ниже).

Говоря формально, истинность оценивается по следующим правилам:

$M, (w, O_a) \Vdash p$, е.т.е. $w \in v(p)$.

$M, (w, O_a) \Vdash \neg\varphi$, е.т.е. $M, (w, O_a) \nVdash \varphi$.

$M, (w, O_a) \Vdash \varphi \wedge \psi$, е.т.е. $M, (w, O_a) \Vdash \varphi$ и $M, (w, O_a) \Vdash \psi$.

(На основании этих правил получаются правила для дизъюнкции и импликации, которые определяются через отрицание и конъюнкцию стандартным образом.)

$M, (w, O_a) \Vdash \Box\varphi$, е.т.е. $W \subseteq [[\varphi]]^{O_a}_M$, где $[[\varphi]]^{O_a}_M = \{w \in W\colon M, (w, O_a) \Vdash \varphi\}$.

(Модальность возможности определяется через модальность необходимости также стандартно.)

$$M, (w, O_a) \Vdash B_A\,\varphi, \text{ е.т.е. } O \subseteq [[\varphi]]^{O_a}_M \text{ и } t(\varphi) \sqsubseteq a.$$

$$M, (w, O_a) \Vdash B_P\,\varphi, \text{ е.т.е. } O^{\cap} \subseteq [[\varphi]]^{O_a}_M \text{ и } t(\varphi) \sqsubseteq \boldsymbol{\beta},$$

где $O^{\cap} := \cap\,\Theta$ и $\boldsymbol{\beta} := \oplus\,(\cup_{O \in \Theta}\, t(O))$.

---

[1] Множество формул в L определяется рекурсивно:

$\varphi := p_i \mid \neg\varphi \mid \varphi \wedge \psi \mid \Box\varphi \mid B_A\,\varphi \mid B_P\,\varphi$,

где p_i для любого натурального i обозначает i-й элемент *Prop*.

(Здесь, поскольку $\cup_{O \in \Theta} t\ (O)$ конечно, мы гарантированно получим $\boldsymbol{\beta} \in T$.)

Как можно видеть, истинность формул, в составе которых нет выражений с операторами B_A и B_P, не зависит от ячейки памяти, а истинность модальных формул не зависит также и от мира. Истинность формул вида $B_A\,\varphi$ зависит от ячейки памяти (что соответствует идее релятивизировать активные убеждения к одному из состояний рабочей памяти), но не зависит от мира. Истинность формул вида $B_P\,\varphi$ не зависит ни от ячейки памяти, ни от мира.

Логическое следование для однопосылочного случая определяется так:

$\varphi \vDash \psi$ е.т.е. при любых M, w, O_a, если $M, (w, O_a) \Vdash \varphi$, то $M, (w, O_a) \Vdash \psi$.

Существует более общее определение следования из множества посылок, однако оно мне не понадобится, поэтому я не буду его приводить.

Чтобы было понятнее, рассмотрим действие условий истинности формул с доксастическими операторами на примерах. Допустим, нужно оценить на истинность высказывание «Алиса (здесь и сейчас) убеждена, что инопланетяне существуют». Для этого нам нужно задать, во-первых, множество возможных миров. Пусть это будут все миры, не являющиеся логически противоречивыми. Поскольку существование инопланетян не противоречит никаким логическим принципам, в некоторых из этих миров инопланетяне будут существовать. После этого нам нужно задать множество топиков и отношение части на нем. Допустим, мы это каким-то образом сделали. Далее требуется определить элементы Θ, т.е. рассечь множество возможных миров на фрагменты, соответствующие различным состояниям Алисы, а затем определить функцию t, присваивающую топик каждой пропозиции и множество топиков каждому элементу Θ. Теперь у нас есть возможность указать, какая ячейка памяти является активной сейчас, когда мы хотим оценить на истинность наше приписывание Алисе убеждения в существовании инопланетян. Допустим, это ячейка O_a. Чтобы узнать, истинно ли данное приписывание относительно актуального мира («здесь») и ячейки памяти O_a («сейчас»), нам нужно, во-первых, узнать, во всех ли мирах множества O существуют инопланетяне, во-вторых, является ли тема существования инопланетян частью a, т.е. того списка тем, который Алиса обдумывает, будучи в состоянии, моделируемом O_a. Если ответ на оба вопроса «да», то приписывание истинно. Интересно, что поскольку истинность приписывания независима от мира, если окажется, что Алиса действительно убеждена в существовании инопланетян относительно O_a («сейчас»), то она убеждена в нем не только актуально («здесь»), но и потенциально, т.е. во всех других возможных мирах. Это объясняется тем, что, зафиксировав ячейку памяти O_a, мы тем самым ограничили оценку так, что, в каком бы возможном мире ни находилась наша потенциальная Алиса, у нее будет то же состояние рабочей памяти, что и в актуальном мире.

С пассивными убеждениями ситуация аналогична, за двумя исключениями. Во-первых, вместо того чтобы проверять истинность содержания приписываемого убеждения в мирах активной ячейки памяти, мы проверяем его на $O^{\cap}$. Смысл тут в том, что если каждое активное убеждение исключает из пространства доксастически возможного все миры, которые с ним «не согласны», и мы перебираем все такие убеждения, присущие агенту, то в конце концов не исключенными остаются только те миры, которые ни одному ак-

тивному убеждению не противоречат. Эти миры и отражают содержание пассивных убеждений в их информационном аспекте. Тематический же аспект отражается в топике $\boldsymbol{\beta}$, который представляет собой слияние всех топиков, о которых агент когда-либо думал. Ситуация, когда топик какой-то пропозиции не попадает в $\boldsymbol{\beta}$, соответствует тому, что агент в принципе не знает, о чем говорит эта пропозиция, а не просто не думает об этом сейчас. Поэтому он, очевидно, не может иметь убеждения ни в истинности этой пропозиции, ни в ее ложности. Это очень хорошо соответствует интуитивному пониманию того, что происходит в таких случаях. Например, обосновывая, почему Цинь Ши-Хуанди ни в какой момент своей жизни не мог быть убежден ни в том, что Китай станет ядерной державой, ни в том, что Китай не станет ею, мы скажем, что Цинь Ши-Хуанди просто *не знал о* ядерном оружии. Интересно, что даже если мы образуем дизъюнкцию двух утверждений $p \vee \neg p$, мы не сможем истинно приписать убеждение в ней агенту, если для этого агента не выполняется условие $t\,(p) \sqsubseteq \boldsymbol{\beta}$.

Рассмотрим теперь то, как данная семантика работает непосредственно с эффектами фрейминга. Если говорить о кратковременной памяти, принятые семантические правила делают возможным построить такую модель M и взять в ней такую ячейку памяти O_a,[1] что для некоторых пропозиций p и q будет выполняться $M, (w, O_a) \Vdash B_A\, p$ и $M, (w, O_a) \Vdash \neg B_A\, q$, при том что пропозиции p и q являются необходимо эквивалентными на M. Однако мы не сможем взять такую O_a, чтобы при тех же условиях получить $M, (w, O_a) \Vdash B_A \neg q$.[2]

Пример 1. Возьмем модель $M_1 = <\{w_1, w_2\}, \{O, U\}, \{a, \boldsymbol{\beta}, c\}, \oplus, t, v>$, где $O = \{w_1\}$, $U = \{w_2\}$; $t\,(O) = \{a\}$, $t\,(U) = \{c\}$, $t\,(p) = a$, $t\,(q) = c$; $a \oplus c = \boldsymbol{\beta}$; $v\,(p) = v\,(q) = \{w_1\}$.

Для произвольного мира w в этой модели

$M_1, (w, O_a) \Vdash \Box(p \leftrightarrow q)$, т.к. $W \subseteq [[p \leftrightarrow q]]^{O_a}_M$,

$M_1, (w, O_a) \Vdash B_A\, p$, т.к. $O \subseteq \{w_1\}$ и $t\,(p) \sqsubseteq a$,

$M_1, (w, O_a) \Vdash \neg B_A\, q$, т.к. $M, (w, O_a) \nVdash B_A\, q$, т.к. неверно, что $t\,(q) \sqsubseteq a$.

Посмотрев на последнюю строчку в этом примере, легко понять, почему при таких условиях всегда будет $M, (w, O_a) \nVdash B_A \neg q$. Если p и q необходимо эквивалентны, единственно возможной причиной того, чтобы у агента при наличии активного убеждения, что p, отсутствовало активное убеждение, что q, является тематическая «несхваченность» пропозиции q в данном состоянии агента. Иными словами, агент сейчас не думает про то, о чем говорится в пропозиции q. Но если он сейчас про это не думает, то он не думает и о том, о чем говорится в пропозиции $\neg q$, поскольку у этих двух пропозиций одна тематика. А значит, агент не имеет сейчас активного убеждения, что $\neg q$. Это означает, что мы не можем истинно приписать никакому агенту ни в каком

---

[1] Мир w можно взять произвольно из множества миров модели.

[2] Приведу семантическое доказательство последнего факта. Пусть даны некоторые модель M, мир w и ячейка памяти O_a в M такие, что $M, (w, O_a) \Vdash B_A\, p$ и $M, (w, O_a) \Vdash \Box(p \leftrightarrow q)$. Предположим также, что $M, (w, O_a) \Vdash B_A \neg q$.

Поскольку $M, (w, O_a) \Vdash B_A\, p$, е.т.е. $O_a \subseteq [[p]]^{O_a}_M$ и $t\,(p) \sqsubseteq a$, $M, (w, O_a) \Vdash \Box(p \leftrightarrow q)$, е.т.е. $W \subseteq [[p \leftrightarrow q]]^{O_a}_M$, можно получить, что $O_a \subseteq [[q]]^{O_a}_M$. Далее, поскольку $M, (w, O_a) \Vdash B_A \neg q$, е.т.е. $O_a \subseteq [[\neg q]]^{O_a}_M$ и $t\,(\neg q) = t\,(q) \sqsubseteq a$, можно получить, что $O_a \subseteq [[\neg q]]^{O_a}_M$. Таким образом, $O_a \subseteq [[q]]^{O_a}_M$ и $O_a \subseteq [[\neg q]]^{O_a}_M$. Это было бы возможным, только если бы O_a было пусто. Но O_a непусто, согласно определению ячейки памяти. Значит, мы пришли к противоречию.

состоянии обладание двумя активными убеждениями в двух противоречащих друг другу пропозициях. Например, если формализовать конъюнкцию предложений «Вася убежден, что количество учеников в классе меньше 33» и «Вася убежден, что количество учеников в классе не меньше $\sqrt{1089}$», то эта конъюнкция будет ложной в любой модели (при формализации слова «убежден» с оператором B_A).

Если же говорить о долговременной памяти, то ситуация меняется: здесь уже можно построить такую модель M, в которой имеет место[1] $M, (w, O_a) \Vdash B_P p$ и $M, (w, O_a) \Vdash B_P \neg q$ при необходимо эквивалентных p и q. В качестве примера, иллюстрирующего такую ситуацию, можно взять модель M_1, определенную выше, и какую-то пару из мира и ячейки памяти в ней, скажем, w_1 и O_a. Действительно, при таких условиях оценки будет выполняться и

$M_1, (w_1, O_a) \Vdash B_P p$, т.к. $O^{\cap} = \emptyset \subseteq \{w_1\}$ и $t(p) \sqsubseteq \boldsymbol{\beta}$,

и

$M_1, (w_1, O_a) \Vdash B_P \neg q$, т.к. $O^{\cap} = \emptyset \subseteq \{w_2\}$ и $t(q) \sqsubseteq \boldsymbol{\beta}$.

Причина появления противоречащих друг другу пассивных убеждений в данном случае также проста: поскольку $O^{\cap}$ пусто, агент, согласно семантике, пассивно убежден вообще во всем, что тематически «схватывается» им в долговременной перспективе. Именно эта особенность делает данную семантику способной описывать ситуации, подобные той, о которой говорил Д. Льюис, – когда несколько разных фрагментов системы убеждений, каждый из которых является внутренне согласованным, рассмотренные вместе, влекут противоречие, которое остается незаметным для агента, поскольку все эти фрагменты никогда не активируются одновременно.

В целом, семантика логики эффектов фрейминга показывает, как именно может быть ограничена рациональность агента и как эту ограниченную рациональность можно достаточно легко моделировать. Причем здесь используется сразу два ограничения: ограничение, связанное с фрагментацией множества возможных миров, и ограничение, связанное с тематикой. Как отмечается в [20. P. 9], ни одного из этих ограничений как такового не достаточно для решения обсуждаемых проблем семантики пропозициональных установок. Если использовать фрагментацию без тематики, то мы получим систему, в которой агенту будут известны все, по крайней мере, простые тавтологии, даже если они выражены в понятиях, которыми агент не владеет. Если использовать тематику без фрагментации, то мы получим систему, в которой агент является всеведущим, по крайней мере, в рамках одной и той же темы. И то и другое нежелательно для семантики и логики, призванных описывать содержание пропозициональных установок реальных агентов.

Ф. Берто и А. Озгюн специально приводят три свойства своей логики, показывающих, что ни простая эквивалентность пропозиций φ и ψ при данных условиях оценки, ни пассивное убеждение агента в их эквивалентности, ни их эквивалентность плюс пассивное владение агентом списком тем ψ («знать, о чем это») не гарантируют, что агент, активно убежденный, что φ, будет активно убежден, что ψ:

1) *фрейминг а*

$\varphi \leftrightarrow \psi \nvDash B_A \varphi \leftrightarrow B_A \psi$,

[1] Здесь и ячейка памяти, и мир могут быть взяты произвольно.

2) *фрейминг b*

$B_A \varphi \wedge B_P (\varphi \leftrightarrow \psi) \nvdash B_A \psi$,

3) *фрейминг c*

$\varphi \leftrightarrow \psi \nvdash B_A \varphi \wedge B_P \bar{\psi} \rightarrow B_A \psi$.

Рассмотрим эти свойства на примере конкретной модели.

Пример 2. Возьмем модель $M_2 = <\{w_1, w_2\}, \{O, U\}, \{a, \boldsymbol{\beta}, c\}, \oplus, t, v>$, где $O = \{w_1\}$, $U = \{w_1\}$; $t(O) = \{a\}$, $t(U) = \{c\}$, $t(p) = a$, $t(q) = c$; $a \oplus c = \boldsymbol{\beta}$; $v(p) = \{w_1, w_2\}$, $v(q) = \{w_1\}$.

$M_2, (w_1, O_a) \Vdash B_A p$, т.к. $O \subseteq \{w_1, w_2\}$ и $t(p) \sqsubseteq a$,

$M_2, (w_1, O_a) \Vdash p \leftrightarrow q$, т.к. $M_2, (w_1, O_a) \Vdash p$ и $M_2, (w_1, O_a) \Vdash q$,

$M_2, (w_1, O_a) \Vdash B_P (p \leftrightarrow q)$, т.к. $O^{\cap} = \{w_1\} \subseteq \{w_1\}$[1] и $t(p) \oplus t(q) \sqsubseteq \boldsymbol{\beta}$,

$M_2, (w_1, O_a) \Vdash B_P \bar{q}$, т.к. $t(q) \sqsubseteq \boldsymbol{\beta}$,

$M_2, (w_1, O_a) \Vdash \neg B_A q$, т.к. неверно, что $t(q) \sqsubseteq a$.

Если разобрать то, что говорят эти формулы на нашем примере про Васю, получим следующее. Из эквивалентности пропозиции, что количество учеников в классе меньше 33, и пропозиции, что количество учеников в классе меньше √1089, нельзя вывести, что если Вася активно убежден в первой, то он активно убежден и во второй. Этот вывод нельзя будет получить и в том случае, если мы добавим условие, что Вася знает, что означает сам вопрос о соотношении количества учеников и √1089. И даже если известно, что Вася, в принципе, знает о том, что эти две пропозиции эквивалентны (например, мы заключаем о его знании из того, что однажды видели, как он верно решил задание «√1089 = ?» и объяснил свое решение), он может не прийти к убеждению во второй пропозиции просто потому, что сейчас он не актуализирует то свое знание. Как видно, в этом случае предсказания, которые дает нам логика эффектов фрейминга, вполне соответствуют интуиции, а также повседневному опыту и результатам психологических исследований.

Таким образом, свою основную задачу – корректно описывать эффекты фрейминга – данная семантика и основанная на ней логика успешно выполняют. Если так, возникает естественное стремление проверить, как этот формализм справляется с другими связанными проблемами, известными в семантике пропозициональных установок. К сожалению, особенности формализма в том его виде, который был представлен его авторами, не позволяют успешно применять его к классическим примерам проблемы нарушения подстановочности в косвенных контекстах (она же головоломка Фреге), поскольку в этих примерах используется подстановка имен, а не пропозициональных констант. Поэтому единственной целью здесь может быть проверка возможностей логики эффектов фрейминга в решении проблемы логического всеведения. И для того чтобы достичь этой цели, нам следует рассмотреть аксиоматику логики эффектов фрейминга, а затем показать, как теоремы данной логики соотносятся с интуитивными оценками возможностей рациональных агентов, а также с имеющимся опытом в этой области.

Аксиоматика и некоторые выводы о логическом всеведении

Относительно аксиоматики следует начать с замечания о том, что уже из семантических правил ясна валидность для данной логики всех классических

[1] В данной модели w_1 является единственным миром, в котором истинна формула $p \leftrightarrow q$.

пропозициональных тавтологий. Правило Modus Ponens также действует. Дополнительно действуют все аксиомы и правила модальной логики S5 для □. В сущности, логика эффектов фрейминга является расширением логики S5, и семантические правила ее включают все правила S5 без каких-либо изменений. Третьей, специфической для данной логики группой аксиом являются аксиомы для формул с B_A и B_P. Эта группа включает в себя:

1) *четыре аксиомы, одинаковые для B_A и B_P* (вместо звездочки в формулы подставляется один из двух субскриптов, т.е. A либо P[1]):

(C) $B_\star (\varphi \wedge \psi) \rightarrow (B_\star \varphi \wedge B_\star \psi)$,

(Ax1) $B_\star \varphi \rightarrow B_\star \overline{\varphi}$, где $\overline{\varphi} := \bigwedge_{x \in \mathrm{Var}(\varphi)} (x \vee \neg x)$,

(Ax2) $(\Box(\varphi \rightarrow \psi) \wedge B_\star \varphi \wedge B_\star \overline{\varphi}) \rightarrow B_\star \psi$,

(Ax3) $B_\star \varphi \rightarrow \Box B_\star \varphi$,

2) *одну аксиому только для B_A*:

(D) $B_A \varphi \rightarrow \neg B_A \neg\varphi$,

3) *одну аксиому, связывающую B_A и B_P*:

(Inc) $B_A \varphi \rightarrow B_P \varphi$.

Корректность и полнота данной аксиоматической системы относительно класса моделей, заданного семантикой логики эффектов фрейминга, доказывается авторами в приложении [4. P. 16–22].

Из тех примеров, которые приведены в предыдущем разделе, видно, что в логике эффектов фрейминга наиболее общие формулировки логического всеведения не выполняются ни для активных, ни для пассивных убеждений:

$\Box(\varphi \rightarrow \psi) \nvDash B_A \varphi \rightarrow B_A \psi$

и

$\Box(\varphi \rightarrow \psi) \nvDash B_P \varphi \rightarrow B_P \psi$.

Это означает отсутствие замкнутости системы убеждений относительно необходимой импликации. По тем же самым причинам, которые были изложены выше, не выполняются и более частные формулировки логического всеведения:

$\Box\varphi \nvDash B_A \varphi$

и

$\Box\varphi \nvDash B_P \varphi$.

Это означает, что агент может не знать об истинности каких-то теорем логики и математики и даже может не сразу правильно решать простые математические задания типа «√1089 = ?». Содержательно объяснить его затруднения в подобных случаях можно следующим образом: когда агент видит такое задание, у него нет причин сразу же думать о числе 33, а потому он может какое-то время (пока не произведет вычисления) не понимать, что именно это число является правильным ответом. Точнее, агент сначала думает об этом числе, используя для него только имя «√1089», но не имя «33», и если эти имена для него образуют различные топики (что вполне естественно предположить), он может не догадываться о том, что это имена одного и того же числа. Поэтому из убеждения, что правильным ответом на задание будет число √1089, он может не суметь перейти к убеждению, что правильным ответом на задание будет число 33. Иначе говоря, агент может

[1] Подстановку следует осуществлять так, чтобы в каждой инстанциации какой-либо аксиомы на всех местах вместо звездочки стоял один и тот же субскрипт.

не знать о том, что необходимо истинная пропозиция, что $\sqrt{1089} = 33$, истинна. Причем, как показывает формализм, он может не знать этого ни активно, ни пассивно.

Одно это, несомненно, представляет собой большой шаг в преодолении логического всеведения. Вместе в тем логическое всеведение устраняется в логике эффектов фрейминга не настолько полно, как может показаться сначала. В частности, в рамках принятой в данной логике аксиоматической системы доказывается следующая теорема.

Локальная дедуктивная замкнутость:

(1) $B_\star \varphi \wedge B_\star (\varphi \rightarrow \psi) \vdash B_\star \psi$.

То, что агент убежден в φ при некоторых условиях оценки и при тех же условиях также убежден в том, что из φ следует ψ, влечет, что он убежден в ψ. (Здесь слово «убежден» следует понимать как активную убежденность, если в формуле вместо $B_\star$ везде фигурирует B_A, и как пассивную убежденность, если там фигурирует B_P соответственно.)

Доказательство: (для B_A) 1. $B_A\varphi$ [Нур.]; 2. $B_A(\varphi \rightarrow \psi)$ [Нур.]; 3. $B_A(\varphi \wedge (\varphi \rightarrow \psi))$ [из 1 и 2 по C]; 4. $\varphi \wedge (\varphi \rightarrow \psi) \rightarrow \psi$ [проп.]; 5. $\Box(\varphi \wedge (\varphi \rightarrow \psi) \rightarrow \psi)$ [из 4 по S5]; 6. B_A [из 2 по Ax1]; 7. $B_A\, \psi^{-}$ [из 6 по Def. φ^{-}]; 8. $B_A\, \psi$ [из 3, 5, 7 по Ax2]. (Для B_P аналогично.)

В эпистемической логике свойство, аналогичное локальной дедуктивной замкнутости, называется замкнутостью относительно известной (known) импликации. Многие системы эпистемической логики имеют это свойство, и это не считается их недостатком. Однако можно привести аргументы в пользу того, чтобы попытаться избежать появления локальной дедуктивной замкнутости в тех формальных системах, которые предназначены для моделирования убеждений реальных агентов. Например, итерированное применение (1) гарантирует, что если агент убежден в посылках некоторого строгого доказательства и считает верным каждый шаг этого доказательства в отдельности, то он будет убежден и в заключении. Иначе говоря, наши агенты должны корректно использовать Modus Ponens «внутри» своей системы убеждений. Это требование выглядит вполне обоснованным и даже желательным, пока мы рассматриваем его на уровне теории. Вместе с тем оно показывает, что способности агента к дедуктивному выводу в данной логике трактуются как спонтанные, «автоматически» применяемые ко всему подряд и не требующие никаких дополнительных ресурсов, даже времени. Количество убеждений, которыми обладает каждый агент в каждый момент времени, согласно локальной дедуктивной замкнутости, является бесконечным, причем даже количество активных убеждений, не говоря уже о пассивных. Ясно, что это значительная идеализация.

Еще более сильную идеализацию означает следующая теорема.

Ограниченная дедуктивная замкнутость:

(2) $\Box(\varphi \rightarrow \psi) \vdash B_\star \varphi \wedge B_\star \psi^{-} \rightarrow B_\star \psi$.

Необходимая импликация между φ и ψ влечет, что если агент убежден в φ при некоторых условиях оценки и при тех же условиях он схватывает тематику ψ, то он убежден в ψ. (Здесь слово «схватывает» следует понимать как активное оперирование данной тематикой («думать про это»), если в формуле вместо $B_\star$ везде фигурирует B_A, и как пассивное владение ею («знать, о чем

это»), если там фигурирует B_P соответственно; и относительно понимания слова «убежден» релевантно то же самое, что и в примечании к (1).)

Доказательство: (для B_A) 1. $\Box(\varphi \rightarrow \psi)$ [Нур.]; 2. $B_A\,\varphi$ [Нур.]; 3. $B_A\,\overline{\psi}$ [Нур.]; 4. $B_A\,\psi$ [из 1, 2, 3 по Ax2]; 5. $B_A\,\varphi \wedge B_A\,\overline{\psi} \rightarrow B_A\,\psi$ [из 4, элим. 2, 3]. (Для B_P аналогично.)

Хотя это и не логическое всеведение, кажется, что мы подходим уже достаточно близко к нему. Чтобы стало ясно, в чем здесь проблема, рассмотрим это свойство в контексте приведенного выше примера про шахматы. Допустим, φ будет большой конъюнкцией пропозиций, описывающих положение фигур на доске и правила игры в шахматы, а ψ будет пропозицией, что Qe7 при данной позиции является началом выигрышной стратегии. Между этими пропозициями существует необходимая импликация, что означает выполнение посылки теоремы об ограниченной дедуктивной замкнутости. Далее, наш агент видит позицию на доске и знает правила игры, а значит, он убежден, что φ. Кроме того, он рассматривает последовательно все ходы, возможные при данной позиции, в том числе и Qe7, с целью найти тот ход, который привел бы его к выигрышу. Можно сказать поэтому, что он в какой-то момент задумывается и о том, истинно ли ψ. Тогда, согласно ограниченной дедуктивной замкнутости, агент должен быть убежден, что ψ истинно. Однако он все же делает другой ход, что невозможно объяснить с помощью обсуждаемого формализма. Как видим, логика эффектов фрейминга имеет применительно к этому примеру ровно те же затруднения, что и мультимодельный подход.

Что на это могут ответить сторонники данных подходов? Есть как минимум два варианта. Во-первых, они могут сказать, что в нашем примере недостаточно тонко различены состояния агента. То, что агент в один и тот же относительно короткий период времени обдумывает все пропозиции, являющиеся конъюнктами в φ, а также пропозицию ψ, не означает, что он *знает* всю конъюнкцию φ в пределах одного и того же состояния, а тем более не означает, что он обдумывает ψ в пределах этого же состояния. Даже если считать, что множество возможностей, которые он рассматривает в это время, не изменяется, можно сказать, что изменяются рассматриваемые им топики, и ни в одном из случаев не активируется такой ячейки памяти, топик которой включал бы все топики, необходимые для перехода к убеждению, что ψ. Подобная стратегия, кажется, подразумевается в [20. P. 10], в примечании 9. Однако это не очень убедительная стратегия, поскольку она может быть опровергнута более простыми примерами такого рода. Та же самая теорема Ферма представляет собой настолько простое предложение, что кажется весьма контринтуитивным предполагать, что в процессе его восприятия и обдумывания агент пребывает в нескольких разных состояниях.

Во-вторых, сторонник логики эффектов фрейминга мог бы сказать, что вывод, который мы получили, правильный, но он не является причиной для того, чтобы отвергать обсуждаемый формализм. Нужно просто подобрать более точную его интерпретацию. Такой ход часто используется в семантике пропозициональных установок. Пример подал уже Я. Хинтикка, который в более поздних своих работах писал, что логика, которую он построил в [3], является не логикой собственно знания и убеждения, а логикой чего-то вроде

информированности [21. P. 26]. Например, мы могли бы говорить, что операторы B_A и B_P выражают не то, какие активные и пассивные убеждения агент фактически имеет, а то, какие убеждения он мог бы иметь при данных условиях оценки «при наилучшем раскладе», так сказать. В конце концов, на способность рационального агента к производству определенного вывода влияет не только сама логическая возможность такого вывода для него, но и то, насколько агент опытен в этом деле, насколько он доверяет своим логическим способностям, насколько он в состоянии сосредоточиться в данный момент и т.д. Ничего из этих факторов не отражено в том объяснении, которое дает нам логика эффектов фрейминга, поэтому неудивительно, что объяснение является частичным. Вместе с тем даже частичное объяснение лучше, чем никакого, и если сравнить то частичное объяснение, которое давала нам доксастическая логика Хинтикки, с объяснением логики эффектов фрейминга, очевидно, что последняя объясняет намного больше.

Во многом с тем, что сказано в предыдущем абзаце, можно с готовностью согласиться. Логика (по крайней мере, логика эффектов фрейминга) действительно не предназначена для того, чтобы быть формализацией полной психологической теории в области убеждений или в какой бы то ни было области. Она просто решает другие задачи. Однако, коль скоро в качестве интерпретации логических формул используются психологические понятия, критика выводов, полученных в рамках логики, за то, что они не вполне соответствуют психологической реальности, должна считаться правомерной. В конце концов, именно подобного рода критика привела к появлению подходов, альтернативных подходу Я. Хинтикки, в частности к появлению самой логики эффектов фрейминга. Надо полагать, что и в дальнейшем такая критика способна играть стимулирующую роль. Именно для этого я ее здесь привожу.

Дополнительно для активных убеждений в логике эффектов фрейминга можно вывести следующую теорему:

(3) $\vdash B_A \varphi \rightarrow \Diamond\varphi$.

Если агент активно убежден в чем-то, то это возможно.

Доказательство: 1. $B_A \varphi$ [Нур.]; 2. $\neg\Diamond\varphi$ [Нур.]; 3. $\Box\neg\varphi$ [из 2 по Def. $\Diamond$]; 4. $\Box(\varphi \rightarrow \neg\varphi)$ [из 3 по S5]; 5. $B_A \overline{\varphi}$ [из 1 по Ax1]; 6. B_A [из 5 по Def. $\overline{\varphi}$]; 7. $B_A \neg\varphi$ [из 1, 4, 6 по Ax2]; 8. $\neg B_A \neg\varphi$ [из 1 по D]; 9. $B_A \neg\varphi \wedge \neg B_A \neg\varphi$ [из 7, 8]; 10. $\Diamond\varphi$ [из 9, элим. 2]; 11. $B_A \varphi \rightarrow \Diamond\varphi$ [из 10, элим. 1].

Сама по себе выводимость формулы, представленной в (3), кажется достаточно безобидным свойством данной логики. Более того, она имеет простое семантическое объяснение. Поскольку условием истинности формул вида $B_A \varphi$ относительно ячейки памяти O_a является конъюнкция, один из конъюнктов которой гласит, что φ должно быть истинно на множестве O, и поскольку в определении ячейки памяти сказано, что множество O непусто, отсюда следует, что если $B_A \varphi$ истинно, то существует, по крайней мере, один мир, в котором истинно φ. А поскольку условием истинности формул вида $\Diamond\varphi$ является как раз существование такого мира, получаем, что $\Diamond\varphi$ истинно. Все вполне закономерно и понятно. Однако оказывается, что теорема (3) имеет крайне проблематичные в философском смысле следствия. Я рассмотрю три из них.

Первым следствием (3) является выводимость формулы, которую можно считать формализацией так называемого *принципа следования возможности из «мыслимости»*[1]:

(3a) $\vdash B_A \Diamond\varphi \rightarrow \Diamond\varphi$.

Если агент активно убежден в возможности чего-то, то это действительно возможно.

Доказательство: 1. $B_A \Diamond\varphi$ [Нур.]; 2. $\Diamond\Diamond\varphi$ [из 1 по (3)]; 3. $\Diamond\varphi$ [из 2 по S5].

Принцип следования возможности из «мыслимости» или «представимости» (conceivability) известен еще со времен средневековой философии, и он еще с тех пор вызывал у философов обоснованные сомнения. Кажется, это означает что-то слишком сильное – что реальность каким-то образом зависит от мышления. Не вдаваясь здесь в эти дискуссии, можно отметить, по крайней мере, что данный принцип предъявляет весьма сильные требования к тому, как агент понимает модальности. Более отчетливо это проявляется, если рассмотреть другое следствие (3), близкое по смыслу к обсуждаемому принципу:

(3b) $\vdash \neg\Box\varphi \rightarrow \neg B_A \Box\varphi$.

Агент не убежден в необходимости того, что не является необходимым.

Доказательство: 1. $B_A \Box\varphi$ [Нур.]; 2. $\Diamond\Box\varphi$ [из 1 по (3)]; 3. $\Box\varphi$ [из 3 по S5]; 4. $B_A \Box\varphi \rightarrow \Box\varphi$ [из 3, элим. 1]; 5. $\neg\Box\varphi \rightarrow \neg B_A \Box\varphi$ [из 4 по контрапозиции].

Это означает не что иное, как *инфаллибиллизм в области дедуктивных наук*. Согласно данной логике, агент не может, например, ошибаться в доказательствах: ведь если ошибочное доказательство он примет за верное, то сочтет, что полученный им результат имеет необходимый характер, что противоречит сказанному в (3b). Причем, обратим внимание, агент не только не может *знать* о необходимости того, что необходимым не является (что было бы вполне оправданным выводом, если бы мы придерживались классической инфаллибилистской концепции знания). Он не может быть *убежденным* в необходимости этого, что почти полностью стирает различие между знанием и убеждением применительно к области дедуктивных наук. Между тем, думается, никто не станет спорить, что в логике и математике (как дисциплинах) убеждения присутствуют и играют роль, существенную и отчетливо отличимую от той роли, которую в этих дисциплинах играет знание. Доказательства, даже ошибочные, могут быть убедительными и могут пользоваться *доверием*. Так, мы не всегда проверяем доказательства теорем, встречающихся в учебниках, поскольку полагаем, что до нас их проверило уже много людей, в том числе специалистов в этой области, и вероятность ошибки крайне мала. И тем не менее такая вероятность есть, потому что все мы люди, а люди могут ошибаться.

Еще более подозрительной выглядит теорема, которую я буду называть *теоремой о необходимости «мыслимо» необходимого*:

(3c) $\vdash B_A \Diamond\Box\varphi \rightarrow \Box\varphi$.

Если агент активно убежден в возможности чего-то как необходимого, то это действительно является необходимым.

[1] Под «мыслимостью» здесь и далее я понимаю, в соответствии с традицией употребления этого понятия, то, что некоторый агент схватывает содержание φ как такое, которое не содержит в себе противоречия и, следовательно, является возможным в логическом смысле.

Будучи непосредственным следствием подстановочного случая теоремы (3) по S5, теорема о необходимости «мыслимо» необходимого существенно отличается от (3) по тому, какой смысл из нее вычитывается. В частности, философ при взгляде на формулу в (3c) сразу заметит сходство ее с логической формой хорошо известного и крайне дискуссионного доказательства, а именно онтологического доказательства бытия Бога. Конечно, для того чтобы получить из этой формы само доказательство, требуется провести сначала большую концептуальную работу, и тем не менее появление формул такого рода в качестве доказуемых в логике эффектов фрейминга показывает, что данная логика может иметь серьезные и не для всех приемлемые философские следствия. Думается, этот вопрос стоило бы исследовать более подробно.

Есть еще несколько интересных свойств логики эффектов фрейминга, таких как прямая и обратная интроспекция для активных и пассивных убеждений, обратная негативная интроспекция для активных убеждений и т.д. Но эти свойства не относятся к теме настоящей статьи, поэтому они не будут здесь обсуждаться.

5. Заключение

Резюмируя сказанное выше, можно заключить, что логика эффектов фрейминга в том виде, в котором она представлена в статье Ф. Берто и А. Озгюн, действительно успешно справляется с проблемой эффектов фрейминга на пропозициональном уровне, а вот с проблемой логического всеведения справляется несколько хуже. Таких очевидно чрезмерных идеализаций, как замкнутость системы убеждений относительно необходимой импликации и убеждение во всех необходимо истинных пропозициях, данная логика не поддерживает. Однако те свойства убеждений, которые в ней доказываются (а именно локальная и ограниченная дедуктивная замкнутость, а также свойства (3a)–(3c) для активных убеждений), все же очень сильны и не всеми могут считаться приемлемыми.

В качестве предположения о том, как можно дополнительно ограничить рациональность агента, моделируемую в логике эффектов фрейминга, и тем самым увеличить ее объяснительные возможности, можно рассмотреть расширение этой логики до логики первого порядка с индексикалами. Попытки использования индексикальной стратегии для объяснения эффектов фрейминга и информативности необходимо истинных предложений уже предпринимались, например, в [7] и выглядят достаточно убедительными. Им недоставало лишь хорошей формализации. Думается, тот формализм, который используют Ф. Берто и А. Озгюн, вполне совместим с индексикальностью.

В действительности элементы индексикальной стратегии уже имплицитно применялись выше при интерпретации результатов, которые в формальном виде дает логика эффектов фрейминга – конкретно для объяснения затруднений агента с задачей «$\sqrt{1089} = ?$». Как отмечает при разборе этого примера Дж. Перри в своей книге, правильно решить задачу агенту мешает не отсутствие у него знания об истинности пропозиции, что $\sqrt{1089} = 33$ как таковое. Наиболее непосредственно агенту не хватает знания, что *это число* (т.е. число, которое является ответом на задание) имеет имя «33» [7. P. 144]. Содержание такого знания не является необходимо истинной пропозицией, и потому неосведомленность агента в данном случае вообще не представляла бы собой объяснительной проблемы с точки зрения семантики возможных

миров. Таким образом, проблема лишь в том, чтобы понять, как нужно модифицировать имеющуюся семантику, чтобы она была способна работать с «внутренними» индексикалами[1], подобными *этому числу*. Очевидно, что началом движения к этой цели должно быть преобразование ее в семантику для логики первого порядка.

Существуют, конечно, и другие стратегии борьбы с логическим всеведением в рамках общего подхода семантики возможных миров (например, введение невозможных миров и/или каких-то ограничений на длину логического вывода). Однако эти стратегии дают семантику, которая достаточно сложно формализуется и имеет свои проблемы. На мой взгляд, важное преимущество логики эффектов фрейминга в этом смысле состоит в том, что она имеет не только интуитивно понятную и обоснованную концептуальными соображениями семантику, но и аксиоматику, для которой даже доказаны теоремы о корректности и полноте. Можно надеяться, что в том случае, если эта логика будет расширена, по крайней мере, до логики первого порядка, ее также будет достаточно легко аксиоматизировать, что обеспечивает возможность вывода в ней интересных и в формальном, и в содержательном смысле теорем.

Список источников

1. *Kahneman D., Tversky A.* Choices, Values, and Frames // American Psychologist. 1984. Vol. 39. P. 341–50.

2. *Kahneman D.* Thinking: fast and slow. London : Penguin, 2011.

3. *Hintikka J.* Knowledge and Belief: An Introduction to the Logic of the Two Notions. Ithaca, New York : Cornell University Press, 1962.

4. *Berto F., Özgün A.* The Logic of Framing Effects // Journal of Philosophical Logic. 2021. https://doi.org/10.1007/s10992-022-09694-0

5. *Stalnaker R.C.* Inquiry. Cambridge, Mass. : MIT / Bradford Books, 1984.

6. *Stalnaker R.* Propositions // Issues in the Philosophy of Language / ed. by Alfred Mackay and Daniel Merrill. New Haven : Yale Press, 1976.

7. *Perry J.* Knowledge, Possibility and Consciousness. Cambridge, MA : MIT Press, 2001.

8. *Soames S.* Substitutivity // On Being and Saying: Essays in Honor of Richard Cartwright / ed. by J.J. Thomson. Cambridge, Mass. : MIT Press, 1988.

9. *Salmon N.U.* Frege's Puzzle. Cambridge, MA : MIT Press, 1986.

10. *Schiffer S.* Naming and Knowing // Midwest Studies in Philosophy. 1977. Vol. II. P. 28–41.

11. *Schiffer S.* Belief Ascription // The Journal of Philosophy. 1992. Vol. 89, № 10. P. 499–521.

12. *Моисеева А.Ю.* De re приписывания убеждений и спецификация понятий // Философия науки. 2016. № 4 (71). С. 40–56.

13. *Schiffer S.* The Basis of Reference // Erkenntnis. 1978. № 13. P. 171–206.

14. *Schiffer S.* The 'Fido'-Fido Theory of Belief // Philosophical Perspectives. 1987. Vol. 1. Metaphysics. P. 455–480.

15. *Salmon N.U.* Illogical Belief // Philosophical Perspectives. 1989. Vol. 3. Philosophy of Mind and Action Theory. P. 243–285.

16. *Lewis D.* Logic for Equivocators // Noûs. 1982. Vol. 16. P. 431–441.

17. *Berto F., Jago M.* Impossible Worlds. Oxford : Oxford University Press, 2019. http://dx.doi.org/10.1093/oso/9780198812791.001.0001

18. *Yablo S.* Aboutness. Princeton, NJ: Princeton University Press, 2014.

19. *Berto F.* Aboutness in imagination // Philosophical Studies. 2018. Vol. 175. P. 1871–1886.

20. *Hawke P., Özgün A., Berto F.* The Fundamental Problem of Logical Omniscience // Journal of Philosophical Logic. 2019. https://doi.org/10.1007/s10992-019-09536-6

21. *Hintikka J.* Socratic Epistemology: Explorations of Knowledge-Seeking by Questions. Cambridge : Cambridge University Press, 2007.

[1] Под внутренними индексикалами Перри понимает знаки, которые заводятся агентом в рамках его внутреннего идиолекта и указывают на некоторый объект посредством его роли в определенном контексте, в рамках которого агент этот объект воспринимает.

References

1. Kahneman, D. & Tversky, A. (1984) Choices, Values, and Frames. *American Psychologist*. 39. pp. 341–50.
2. Kahneman, D. (2011) *Thinking: Fast and Slow*. London: Penguin.
3. Hintikka, J. (1962) *Knowledge and Belief: An Introduction to the Logic of the Two Notions*. Ithaca, N.Y.: Cornell University Press.
4. Berto, F. & Özgün, A. (2021) The Logic of Framing Effects. *Journal of Philosophical Logic*. 52. pp. 939–962. DOI: 10.1007/s10992-022-09694-0
5. Stalnaker, R.C. (1984) *Inquiry*. Cambridge, Mass.: MIT / Bradford Books.
6. Stalnaker, R. (1976) Propositions. In: Mackay, A. & Merrill, D. (eds) *Issues in the Philosophy of Language*. New Haven: Yale Press.
7. Perry, J. (2001) *Knowledge, Possibility and Consciousness*. Cambridge, MA: MIT Press.
8. Soames, S. (1988) Substitutivity. In: Thomson, J.J. (ed.) *On Being and Saying: Essays in Honor of Richard Cartwright*. Cambridge, Mass.: MIT Press.
9. Salmon, N.U. (1986) *Frege's Puzzle*. Cambridge, MA: MIT Press.
10. Schiffer, S. (1977) Naming and Knowing. *Midwest Studies in Philosophy*. 2. pp. 28–41.
11. Schiffer, S. (1992) Belief Ascription. *The Journal of Philosophy*. 89(10). pp. 499–521.
12. Moiseeva, A.Yu. (2016) De re pripisyvaniya ubezhdeniy i spetsifikatsiya ponyatiy [De re ascriptions of beliefs and notion specification]. *Filosofiya nauki*. 4(71). pp. 40–56.
13. Schiffer, S. (1978) The Basis of Reference. *Erkenntnis*. 13. pp. 171–206.
14. Schiffer, S. (1987) The 'Fido'-Fido Theory of Belief. *Philosophical Perspectives*. 1. pp. 455–480.
15. Salmon, N.U. (1989) Illogical Belief. *Philosophical Perspectives*. 3. pp. 243–285.
16. Lewis, D. (1982) Logic for Equivocators. *Noûs*. 16. pp. 431–441.
17. Berto, F. & Jago, M. (2019) *Impossible Worlds*. Oxford: Oxford University Press. DOI: 10.1093/oso/9780198812791.001.0001
18. Yablo, S. (2014) *Aboutness*. Princeton, NJ: Princeton University Press.
19. Berto, F. (2018) Aboutness in imagination. *Philosophical Studies*. 175. pp. 1871–1886.
20. Hawke, P., Özgün, A. & Berto, F. (2019) The Fundamental Problem of Logical Omniscience. *Journal of Philosophical Logic*. 49. pp. 727–766. DOI: 10.1007/s10992-019-09536-6
21. Hintikka, J. (2007) *Socratic Epistemology: Explorations of Knowledge-Seeking by Questions*. Cambridge: Cambridge University Press.

Сведения об авторе:
Моисеева А.Ю. – кандидат философских наук, сотрудник Русского общества истории и философии науки (Москва, Россия); научный сотрудник Международной лаборатории логики, лингвистики и формальной философии Национального исследовательского университета Высшая школа экономики (Москва, Россия). E-mail: abyssian03@gmail.com

Автор заявляет об отсутствии конфликта интересов.

Information about the author:
Moiseeva A.Yu. – Cand. Sci. (Philosophy), researcher, Russian Society for the History and Philosophy of Science (Moscow, Russia); research officer, HSE University (Moscow, Russia). E-mail: abyssian03@gmail.com

The author declares no conflicts of interests.

Статья поступила в редакцию 15.12.2023;
одобрена после рецензирования 17.01.2024; принята к публикации 04.03.2024
The article was submitted 15.12.2023;
approved after reviewing 17.01.2024; accepted for publication 04.03.2024