

Рассуждения о знаниях и проблема логического всеведения. Часть I. Модальный подход¹

Аннотация. В статье проведен обзор эпистемических модальных логик, формализующих рассуждения о знаниях. Особое внимание уделено связанной с ними так называемой проблеме логического всеведения и попыткам ее преодоления в рамках модального подхода. Анализ показал, что по способу решения данной проблемы эти попытки образуют два направления. Одно из них основано на искусственном ослаблении дедуктивных возможностей агента, другое направление имеет в своей основе введение в логический язык специальных модальных операторов, трактуемых как «ментальные усилия», необходимые для получения какого-либо знания, выраженного в виде логических формул.

Ключевые слова: эпистемическая логика, проблема логического всеведения, модальный подход, рассуждения с ограниченным ресурсом, многоагентные системы.

Введение

В последние годы в области распределенных систем искусственного интеллекта (DAI) был предложен ряд подходов к спецификации понятия «рациональный агент» в терминах ментальных категорий, таких как знание, вера, намерение, приверженность, обязательство, замысел.

В настоящее время не существует единого взгляда на то, какое сочетание качеств, выраженных в указанных выше терминах, наилучшим образом подходит для приемлемой во всех отношениях характеристики разумного агента. Однако в DAI-сообществе имеется согласие насчет того, что эпистемические понятия «знание», «убеждения» (= «вера») являются одними из важнейших в теории рациональных агентов. По существу, они играют в этой теории фундаментальную роль, так как моделирование ментальных свойств агентов, выраженных в других терминах, возможно только после моделирования их эпистемических свойств. Не вызывает сомнения, что моделирование эпистемических свойств рациональных агентов в качестве теоретического основания должно иметь формаль-

ную теорию знания – специализированную для этой цели эпистемическую логику. В настоящее время известно несколько различных эпистемических логических систем, предложенных для использования их в указанном выше качестве. Однако идеализации, использовавшиеся при построении этих систем, оказались слишком сильными для того, чтобы адекватно моделировать ментальность рациональных агентов, вынужденных решать свои задачи в динамической среде при ограниченном ресурсе времени, которым они располагают.

1. Модальные эпистемические логики

Эпистемическая логика впервые была введена в начале 1960-х годов для решения некоторых философских вопросов, связанных с формальным описанием знаний и убеждений. В дальнейшем к ней стали обращаться представители других научных дисциплин – лингвистики, экономики, теории игр, психологии, информатики. При этом в добавление к традиционным философским вопросам, относящимся к рассуждениям о знаниях, появились вопросы

¹ Работа выполнена при финансовой поддержке РФФИ.

прикладного характера – о вычислительной сложности такого рода рассуждений, о связи между знаниями и действиями etc. Среди различных подходов к моделированию знаний и убеждений наибольшее распространение получил модальный подход. Причина его популярности заключается в том, что многие системы модальной логики допускают вполне естественную эпистемическую интерпретацию, которая позволяет (после сопутствующих изменений) использовать основные технические результаты – метатеоремы исходной модальной логической системы. Ниже модальный подход будет изложен более подробно применительно к построению эпистемической логики для нескольких агентов.

1.1. Язык

Предположим, что имеется N агентов. Введем в пропозициональный язык N модальных операторов знаний $K_1, \dots, K_N$ (по одному для каждого агента). Далее, язык модальной эпистемической логики $\mathcal{L}_N^K$ формально определим следующим образом.

Определение 1 (язык модальной эпистемической логики). Пусть $Atom$ есть непустое счетное множество пропозициональных атомарных формул и $Agent = (1, \dots, N)$ множество агентов. $\mathcal{L}_N^K$ есть минимальное множество такое, что

- 1) $Atom \subset \mathcal{L}_N^K$;
- 2) если $\alpha \in \mathcal{L}_N^K$, то $\neg \alpha \in \mathcal{L}_N^K$;
- 3) если $\alpha \in \mathcal{L}_N^K$ и $\beta \in \mathcal{L}_N^K$, то $(\alpha \rightarrow \beta) \in \mathcal{L}_N^K$;
- 4) если $\alpha \in \mathcal{L}_N^K$ и $i \in Agent$, то $K_i \alpha \in \mathcal{L}_N^K$.

Поясняя данное определение, укажем, что, к примеру, формула вида $K_1 \alpha$ читается «агент 1 знает утверждение α », формула вида $K_1 K_2 \alpha$ читается «агент 1 знает, что агент 2 знает утверждение α ». Модальная глубина (*depth*) формул определяется обычным для систем модальной логики образом: $depth(\phi) = 0$ для всех $\phi \in Atom$; $depth(\neg \alpha) = depth(\alpha)$; $depth(\alpha \rightarrow \beta) = \max(depth(\alpha), depth(\beta))$; $depth(K_i \alpha) = depth(\alpha) + 1$. Условимся также, что пропозициональные связки $\wedge$ и $\vee$ в дальнейшем будут рассматриваться как аббревиатуры соответствующих суперпозиций связок $\neg$ и $\rightarrow$.

1.2. Аксиомы модальной эпистемической логики

Модальная эпистемическая логика для N агентов получается посредством объединения N модальных логик, по одной для каждого агента. Обычно для простоты агенты рассматриваются как гомогенные, т.е. поведение каждого из них подчиняется одной и той же логике. Тогда эпистемическая логика для N агентов рассматривается как объединение N копий некоторой модальной логики. Такая логика обозначается K_N .

Определение 2 (модальная эпистемическая логика K_N). K_N есть модальная эпистемическая логика, имеющая следующие аксиомы и правила вывода (для всех $i \in Agent$):

- (PC) все пропозициональные тавтологии;
- (K) $K_i(\alpha \rightarrow \beta) \rightarrow (K_i \alpha \rightarrow K_i \beta)$ (схема аксиомы дистрибутивности);
- (MP) modus ponens: из α и $(\alpha \rightarrow \beta)$ выводится β ;
- (NEC) из α выводится $K_i \alpha$.

Более «сильные» логики могут быть получены с помощью добавления к аксиомам K_N новых схем аксиом, выражающих другие желаемые принципы, составляющие концепцию знания. Часто используются следующие конструкции:

- (T) $K_i \alpha \rightarrow \alpha$;
- (D) $K_i \alpha \rightarrow \neg K_i \neg \alpha$;
- (4) $K_i \alpha \rightarrow K_i K_i \alpha$;
- (5) $\neg K_i \alpha \rightarrow K_i \neg K_i \alpha$.

Схема аксиом (T) устанавливает, что знания состоят только из истинных утверждений. Этот принцип обычно используется для того, чтобы отличать знание от веры – верить можно как в истинные, так и в ложные (на самом деле) утверждения. По этой причине схему (T) иногда называют «аксиомой знания», а системы, содержащие (T) (такие как $S4_N$ и $S5_N$) – логиками знания.

Схему аксиом (D) иногда называют «аксиомой непротиворечивости», так как она не допускает, чтобы знания агента включали какое-либо утверждение вместе с его отрицанием. Ясно, что условия, накладываемые на поведение агентов с помощью схемы (D), более слабые, чем условия, накладываемые с помощью схемы (T).

Схемы аксиом (4) и (5) называют, соответственно «аксиомами позитивной и негативной интроспекции» (от латинского *introspecto* – смотрю внутрь). Они говорят о том, что агент осведомлен обо всем, что он знает и чего не знает. Их конвенции, т.е. формулы $K_i K_i \alpha \rightarrow K_i \alpha$

и $K_i \neg K_i \alpha \rightarrow \neg K_i \alpha$ являются частными случаями схемы (Т). Вместе со своими конверсиями схемы (4) и (5) дают $K_i \alpha \leftrightarrow K_i K_i \alpha$ и $\neg K_i \alpha \leftrightarrow K_i \neg K_i \alpha$. Более подробно интроспекция описана в [1].

Ниже представлены наиболее известные эпистемические логики, являющиеся консервативными (= сохраняющими все теоремы) расширениями K_N :

T_N (получается с помощью добавления к K_N схемы аксиом (Т));

$S4_N$ (получается с помощью добавления к T_N схемы аксиом (4));

$S5_N$ (получается с помощью добавления к $S4_N$ схемы аксиом (5));

KD_N (получается с помощью добавления к K_N схемы аксиом (D));

$KD4_N$ (получается с помощью добавления к KD_N схемы аксиом (4));

$KD45_N$ (получается с помощью добавления к $KD4_N$ схемы аксиом (5)).

Легко убедиться, что следующие схемы правил вывода общезначимы для K_N и ее консервативных расширений:

(NEC) Из α выводится $K_i \alpha$ (необходимость);

(MON) Из $\alpha \rightarrow \beta$ выводится $K_i \alpha \rightarrow K_i \beta$ (монотонность);

(CGR) Из $\alpha \leftrightarrow \beta$ выводится $K_i \alpha \leftrightarrow K_i \beta$ (конгруэнтность);

(RK_N) Из $\alpha_1 \wedge \dots \wedge \alpha_n \rightarrow \beta$ выводится $K_i \alpha_1 \wedge \dots \wedge K_i \alpha_n \rightarrow K_i \beta$, для всех $n \in \omega$.

1.3. Семантика возможных миров для эпистемической логики

Семантика возможных миров является выражением интуитивной идеи о том, что агент может строить с использованием подходящего языка различные (логически) возможные модели мира. Обычно агент не знает, какая из этих моделей «правильная» (т.е. полностью согласуется с реальным миром), однако, получая информацию о реальном мире, он исключает из списка возможных те модели, которые с этой информацией несовместимы. Чем меньше список оставшихся возможных моделей мира, тем меньше неопределенности в знаниях агента, т.е. тем больше он об этом мире знает.

Зависимость множества моделей мира, которые агент i считает возможными, от его знаний о реальном мире может быть формально представлена с помощью бинарного отношения, скажем, R_i , заданного на множестве логически возможных моделей мира (далее на множестве возможных миров). Чтобы выразить идею о том, что для агента i возможный мир t совместим с его знаниями о реальном мире, «правильной» моделью которого в указанном выше смысле является возможный мир s , будем говорить, что между мирами s и t имеет место отношение R_i (t является эпистемической альтернативой s для агента i). Формально семантика возможных миров определяется следующим образом.

Определение 3. Модель M языка $\mathcal{L}_N^K$ состоит из непустого множества S возможных миров (или состояний), N бинарных отношений $R_1, \dots, R_N$, заданных на множестве S (по одному отношению на каждого агента), и оценочной функции $V: Atom \Rightarrow 2^S$. Отношение истинности формул $\models$ определяется на $\mathcal{L}_N^K$ рекурсивно следующим образом:

$M, s \models \phi$ если и только если $s \in V(\phi)$, для всех $\phi \in Atom$;

$M, s \models \neg \alpha$ если и только если не верно, что $M, s \models \alpha$;

$M, s \models \alpha \rightarrow \beta$ если и только если $M, s \models \beta$ или $M, s \not\models \alpha$;

$M, s \models K_i \alpha$ если и только если для всех $t \in S$ из $s R_i t$ следует $M, t \models \alpha$

Отношения $R_1, \dots, R_N$ называют отношениями эпистемической альтернативности (в других системах модальной логики более распространен термин «отношение достижимости»). Говорят, что формула α истинна по отношению к множеству (классу) моделей $\mathbb{N}$ если для каждой модели $M \in \mathbb{N}$ и каждого мира $s \in S$ имеет место $M, s \models \alpha$.

Легко убедиться, что в соответствии с Определением 3, если формула $\alpha_1 \wedge \dots \wedge \alpha_n \rightarrow \beta$ истинна по отношению к некоторому классу моделей для всех $i \in Agent$ и для каждого натурального числа $n = 0, 1, 2, \dots$, то истинна по отношению к этому же классу моделей и формула $K_i \alpha_1 \wedge \dots \wedge K_i \alpha_n \rightarrow K_i \beta$. Этот принцип может быть интерпретирован как замкнутость

знаний агента относительно логических законов: если i знает все посылки общезначимого правила вывода, то он знает и его заключение.

Если ограничить класс всех моделей, наложив подходящие условия на отношения эпистемической альтернативности, можно расширить множество истинных (по отношению к этому классу) формул. Хорошо известные результаты такого рода для модальной логики очевидным образом трансформируются для эпистемической логики. Они обобщены следующей теоремой [5, 7, 8, 12].

Теорема 1.

1. K_N – множество формул, истинных по отношению к классу всех моделей с N отношениями эпистемической альтернативности.

2. T_N – множество формул, истинных по отношению к классу всех моделей с N рефлексивными отношениями эпистемической альтернативности.

3. $S4_N$ – множество формул, истинных по отношению к классу всех моделей с N рефлексивными и транзитивными отношениями эпистемической альтернативности.

4. $S5_N$ – полна и семантически непротиворечива по отношению к классу всех моделей с N отношениями эпистемической альтернативности, каждое из которых является отношением эквивалентности.

5. KD_N – множество формул, истинных по отношению к классу всех моделей с N отношениями эпистемической альтернативности, удовлетворяющих свойству серийности².

6. $KD4_N$ – множество формул, истинных по отношению к классу всех моделей с N отношениями эпистемической альтернативности, удовлетворяющих свойствам серийности и транзитивности.

7. $KD45_N$ – полна и семантически непротиворечива по отношению к классу всех моделей с N отношениями эпистемической альтернативности, удовлетворяющих свойствам серийности и евклидовости³.

8. K_N , T_N , $S4_N$, $S5_N$, KD_N , $KD4_N$, $KD45_N$ разрешимы.

Логика $S5_N$ рассматривается многими как стандартная логика рационального знания, а $KD45_N$ – как стандартная логика убеждения.

² Т.е. для всех $s, s', s'' \in S$ из $s R_i s'$ и $s R_i s''$ следует $s' R_i s''$.

³ Т.е. для всякого $s \in S$ существует $s' \in S$, такой, что имеет место $s R_i s'$.

2. Проблема логического всеведения

Трактовка эпистемической логики как одной из ветвей модальной логики дает ряд очевидных преимуществ, но требует за них довольно высокую цену. Наиболее серьезная претензия к традиционному модальному подходу состоит в том, что связанная с ним идеализация способностей агентов к рассуждениям подчас оказывается чрезмерной. Эта проблема известна как «проблема логического всеведения» и может возникать в различных формах. Наиболее сильная ее форма проявляется следующим образом.

Пусть Λ – любая модальная логика, содержащая K_N . Для любых $\Gamma \subseteq \mathcal{L}_N^K$, $\alpha \in \mathcal{L}_N^K$, $i \in Agent$, если $\Gamma \vdash_{\Lambda} \alpha$, то $K_i(\Gamma) \vdash_{\Lambda} K_i \alpha$, где $K_i(\Gamma) =_{\text{def}} \{K_i \gamma : \gamma \in \Gamma\}$.

Т.е. если агент знает все формулы из множества Γ , а α логически следует из Γ , то агент знает α . Например, если агент знает все аксиомы теории чисел, то он знает и все ее теоремы (включая и те, которые на самом деле пока математическому сообществу не известны). Существуют также более слабые формы проблемы логического всеведения, перечисленные в [5]:

- агент знает все общезначимые формулы (правило (NEC));

- знания агента замкнуты относительно логической импликации: если агент i знает α и $\alpha \rightarrow \beta$ есть общезначимая формула, то агент знает β (правило (MON));

- знания агента замкнуты относительно эквивалентности: если агент i знает α и $\alpha \leftrightarrow \beta$ есть общезначимая формула, то агент знает β (правило (CGR));

- знания агента замкнуты относительно материальной импликации: если агент i знает α и $\alpha \rightarrow \beta$, то агент знает β (аксиома K);

- знания агента замкнуты относительно конъюнкции: если агент знает α и β , то агент знает $\alpha \wedge \beta$ (аксиома C).

В отношении схем аксиом (D), (4), (5) также установлено [3, 15], что они при определенных обстоятельствах слишком сильны для реалистичного агента, подчас наделяя его даже способностью мгновенно решать алгоритмически неразрешимые проблемы.

Таким образом, если рассматривать модальную эпистемическую логику как средство для

описания действительных знаний агента (т.е. таких знаний, которые всегда у него «под рукой» и готовы для применения в любой момент), то перечисленные выше свойства замкнутости этих знаний вступают в противоречие с ограниченными ресурсами (в первую очередь, времени) которыми располагает агент в реальной действительности. Из-за этих ограничений агент на практике оказывается просто неспособен обрабатывать весь объем информации, которой он якобы обладает исходя из аксиом и правил вывода модальной эпистемической логики, описывающей его знания и дедуктивные возможности. Некоторые логические истины или сравнительно простые следствия из известных фактов он действительно способен устанавливать почти мгновенно, тогда как получение другой информации требует значительного времени и серьезных ментальных усилий. В модальных эпистемических логиках, рассматривавшихся выше, эти две ситуации неразличимы, что делает их мало пригодными для моделирования рассуждений агента, чьи ресурсы ограничены.

Проблема логического всеведения показывает, что точка зрения на модальную эпистемическую логику как на средство для описания действительных знаний реальных (а не идеализированных) агентов не вполне оправдана. Определенный уровень идеализации полезен и даже необходим в философских и научных исследованиях. Однако идеализация, которая делается при использовании модальной эпистемической логики для указанной выше цели, столь сильна, что ментальные возможности идеализированного агента имеют мало общего с таковыми возможностями реальных агентов. Но если модальная эпистемическая логика не описывает действительные знания реальных агентов, то что же она на самом деле все-таки описывает? Некоторые исследователи считают, что законы модальной эпистемической логики приемлемы и для практических целей, если формулу $K_i \alpha$ читать «агент i знает α имплицитно» [4,14], « α логически следует из знаний агента i » [5], «агент i является носителем информации α » [2], « α является потенциальным знанием агента i », вместо «агент i знает α ». Хотя указанные трактовки формулы $K_i \alpha$ могут отличаться друг от друга в деталях, все они имеют практически один и тот же смысл: подразумевается знание, представленное в инфор-

мационном состоянии агента имплицитно (далее, имплицитное знание). Таким образом, модальная эпистемическая логика вполне подходит для формализации концепции «имплицитного знания».

Однако с практической точки зрения для агента более важны его действительные знания, которые естественно назвать эксплицитными. Этими знаниями агент может пользоваться в тот момент, когда они ему потребуются, а не когда-либо в будущем, так как в отличие от имплицитных знаний они уже осознаны им в полной мере. Ввиду важности для агентов эксплицитных знаний, их формализация является предметом исследований, результатом которых стали несколько логических систем.

3. Логика для агентов с искусственно ослабленными дедуктивными способностями

Лишить агента логического всеведения можно различными способами. Например, он может быть ограничен в своих логических (дедуктивных) способностях и не знать некоторых аксиом и правил вывода или отказаться от их использования. Возможно также, что агент не озабочен наличием возможных следствий из известных ему утверждений и даже не будет пытаться их вывести. Ниже будут рассмотрены решения, полученные таким путем в рамках модального подхода.

Естественным образом решение (обычно частичное) проблемы логического всеведения достигается с помощью ослабления стандартной модальной эпистемической логики. Например, не все теоремы стандартной модальной эпистемической логики можно сделать доказуемыми в ее ослабленной версии. При этом, множество формул, известных агенту, оказывается не обязательно замкнутым относительно законов классической пропозициональной логики.

При конструировании таких ослабленных систем может быть постулировано, например, что агент знает только «тривиальные» логические истины, но не знает более «изошренных». Механизм дедукции при этом оказывается неполным, т.е. недостаточно мощным, чтобы вывести все следствия (отвечающие семантике классической пропозициональной логики) из знаний агента [6, 9].

Примерно тот же эффект имеет использование при построении модальной эпистемической логики в качестве основы для расширения некоторых неклассических логик, в частности, некоторых вариантов логики релевантности [5, 15], вместо классической пропозициональной логики. При этом проблему логического всеведения удается (частично) преодолеть благодаря тому, что отношение логического следования, определяемое для указанных неклассических логик, можно рассматривать как подмножество отношения логического следования, определенного для классической пропозициональной логики. Заметим, что в данном случае агент перестает быть логически всеведущим с точки зрения классической логики, но остается таковым с точки зрения неклассической.

Хотя проблема логического всеведения в ее наиболее сильной форме преодолевается с помощью ослабления дедуктивных способностей агента, некоторые ее более слабые формы обычно остаются в силе [11]. Ниже будут рассмотрены две системы, связанные с ослаблением дедуктивных способностей и позволяющие преодолеть проблему логического всеведения во всех ее формах, которые обсуждались в разделе 2.

3.1. Невозможные возможные миры

Один из путей исчерпывающего решения проблемы логического всеведения получил название подхода на основе невозможных миров ("impossible worlds approach" [9, 16]. Интуитивно данный подход можно представить следующим образом. Агент может считать некоторые модели реального мира возможными, хотя логически они невозможны. Например, логически противоречивая формула не может быть истинной. Однако агент, не имея достаточно ресурсов для того, чтобы это противоречие увидеть, просто предполагает, что данная формула истинна. Таким образом, агент рассматривает некоторые миры как возможные, хотя логически они невозможны.

Определение 4 (структуры невозможных возможных миров). Модель невозможных миров M для языка $\mathcal{L}_N^K$ состоит из непустого множества миров S , множества возможных миров $W \subseteq S$ (элементы из $S \setminus W$ будем называть невозможными мирами), N бинарных отношений $R_1, \dots, R_N$ (эпистемической альтернативности), заданных на множестве S (по одному от-

ношению на каждого агента), и оценочной функции $V: \mathcal{L}_N^K \times S \Rightarrow \{1, 0\}$, присваивающей вполне произвольно истинностные значения формулам языка $\mathcal{L}_N^K$ в невозможных мирах и ведущей себя стандартно в возможных мирах, т.е. если $s \in W$, то:

- $V(\neg\alpha, s) = 1$ если и только если $V(\alpha, s) = 0$;
- $V(\alpha \rightarrow \beta, s) = 1$ если и только если $V(\alpha, s) = 0$ или $V(\beta, s) = 1$;
- $V(K_i \alpha, s) = 1$ если и только если $V(\alpha, t) = 1$ для всех $t \in S$, таких, что $s R_i t$.

Отношение истинности формул $\models$ определяется следующим образом: $M, s \models \alpha$ если и только если верно, что $V(\alpha, s) = 1$. Общезначимость определяется только по отношению к возможным мирам: α общезначима по отношению к моделям невозможных возможных миров если и только если для каждой такой модели M и возможного мира $s \in W$ имеем $M, s \models \alpha$.

Так как знания оцениваются по отношению ко всем состояниям, а логические законы для некоторых состояний не действуют, в данном случае удастся обойти любую форму логического всеведения или сразу все. Например, тавтология $\alpha \vee \neg \alpha$ может быть ложной только в невозможном мире, но агент может считать этот мир возможным, поэтому $K_i(\alpha \vee \neg \alpha)$ не будет универсально истинной. Другими словами, правило вывода (NEC) перестает быть общезначимым.

3.2. Осознание

Другое решение проблемы логического всеведения опирается на интуитивное основание, связанное с тем, что агент должен осознавать (aware) свои знания: нельзя знать что-либо, в то же время совершенно не осознавая этого. Заметим, что одним из возможных уточнений фразы «агент i осознает α » является «агент i способен оценить истинностное значение α за время t ».

Определение 5 (структуры осознания). Модель осознания M для языка $\mathcal{L}_N^K$ состоит из непустого множества S состояний, N бинарных отношений $R_1, \dots, R_N$, заданных на множестве S (по одному отношению на каждого агента), N функций $A_1, \dots, A_N$, связывающих каждое состояние с множеством формул, т.е. $A_i: S \Rightarrow 2^{\mathcal{L}_N^K}$ и оценочной функции $V: Atom \Rightarrow 2^S$. Семантика для атомарных формул, отрицаний

и импликаций определяется как обычно (Определение 3). Для формул вида $K_i \alpha$, имеем:

• $M, s \models K_i \alpha$ если и только если $\alpha \in A_i(s)$ и для всех $t \in S$ из $s R_i t$ следует $M, t \models \alpha$.

Интуитивно $A_i(s)$ мыслится как множество формул, которые агент i осознает в состоянии s , а отношения $R_1, \dots, R_N$ используются для моделирования имплицитных знаний. Множество формул, которые агент осознает, может быть произвольным и, значит, не обязательно замкнутым относительно логических законов.

Так как данный подход допускает возможность, когда агент осознает некоторую формулу, но не осознает формулы, являющиеся ее логическими следствиями, в общем случае теоремы эпистемических модальных систем здесь не выполняются, а некоторые правила вывода не являются общезначимыми. Это обстоятельство позволяет обойти все формы логического всеведения, рассмотренные выше. Другой способ убедиться в том, что структура осознания позволяет моделировать агентов, не являющихся логически всеведущими, указывает следующая теорема [5, 16].

Теорема 2.

1. Пусть $M = (S, W, R_1, \dots, R_N, V)$ есть модель невозможных возможных миров. Тогда существует модель осознания $M' = (W', R'_1, \dots, R'_N, A_1, \dots, A_N, V')$ такая, что для любой формулы α языка $\mathcal{L}_N^K$ и каждого $s \in W$ имеем, что $M, s \models \alpha$ если и только если $M', s \models \alpha$.

2. Пусть $M = (S, R_1, \dots, R_N, A_1, \dots, A_N, V)$ есть модель осознания. Тогда существует модель невозможных возможных миров $M' = (W', R'_1, \dots, R'_N, A_1, \dots, A_N, V')$ такая, что для любой формулы α языка $\mathcal{L}_N^K$ и каждого $s \in W$ имеем, что $M, s \models \alpha$ если и только если $M', s \models \alpha$.

Очевидным следствием данной теоремы является вывод о том, что использование структуры осознания потенциально позволяет решить проблему логического всеведения в том же объеме, что и использование структуры невозможных возможных миров.

4. Метарассуждения: рассуждения о рассуждениях

Дедуктивные способности у реальных людей (не обязательно агентов) могут быть самыми разными, от вполне совершенных, до весьма

посредственных. Но вне зависимости от этих способностей люди не являются логически всеведущими. Одна из (достаточных) причин очевидна - для осознания следствий из известных фактов человеку необходимы ментальные усилия, приложение которых приводит к расходу имеющихся ресурсов, в первую очередь, располагаемого времени. Если этих ресурсов не хватит, то человек не сможет (не успеет к определенному сроку) должным образом осознать некоторые, быть может, важные для него факты, хотя они и составляют часть его имплицитных знаний. Этого бы не случилось, если бы человек действительно был логически всеведущим. Эти же соображения остаются в силе и для рациональных агентов (не обязательно людей), о моделировании ментальных свойств которых шла речь до сих пор.

Во всех представленных выше эпистемических модальных логиках состояние знаний агента рассматривалось как неизменное, пребывающее в статике. При этом если знания замкнуты относительно всех логических законов (аксиом и правил вывода) данной логики, то за исключением тривиальных случаев мы имели дело с логически всеведущим агентом, иначе - с агентом, чьи дедуктивные способности искусственно ослаблены. По всей видимости, ключевым моментом здесь является статичность знаний, о которых способен рассуждать агент. Эта статичность не позволяет установить связь между знаниями агента и располагаемым ресурсом времени. Чтобы выйти за эти рамки, необходимо расширить дедуктивные возможности агента таким образом, чтобы он был способен рассуждать не только о состоянии знаний (своем и других агентов), но и о том, как это состояние изменяется во времени в результате рассуждений. Говоря более общо, рациональный и реалистичный агент должен обладать способностью к метарассуждению определенного рода. Ниже будет рассмотрена логическая система, дающая агентам такую способность.

5. Динамическая эпистемическая логика

Идея, лежащая в основе логической системы, введенной в [10], состоит в том, что для осознания каких-либо имеющихся у агента имплицитных знаний ему необходимы определенные ментальные действия («reasoning actions»), соответствующие применениям правил вывода, например, отвечающих схеме правил вывода

«modus ponens» (аксиомы тоже трактуются как правила вывода, у которых отсутствуют посылки). Пока эти действия не будут выполнены, агент не осознает даже очевидные следствия из тех утверждений, которые он уже успел осознать. Те знания, которые агент успел осознать, трактуются как его (текущие) эксплицитные знания (в противовес имплицитным). Для формализации ментальных действий использованы принципы одной из форм динамической логики [13]. С этой целью в язык эпистемической логики добавлено множество базовых действий. Полученный таким образом язык включает формулы вида $[R_i] K_i \alpha$ или $\langle R_i \rangle K_i \alpha$, которые трактуются, соответственно, как «всегда после использования правила вывода R (или иногда после использования правила вывода R) агент i знает α ». Осознание (далее эксплицитное знание) агентом формулы в результате применения ментального действия поясним на примере аксиомы, выражающей принцип применения правила modus ponens: $K_i \alpha \wedge K_i (\alpha \rightarrow \beta) \rightarrow \langle MP_i \rangle K_i \beta$. Данная аксиома «говорит», что если агент знает α и знает, что α имплицирует β , то при условии выполнения шага вывода, состоящего в корректном применении правила modus ponens (MP), он будет знать β . С помощью программных связей динамической логики (таких как композиция или итерация) формализуется концепция, согласно которой агент осознает некоторые следствия из известных ему эксплицитно формул посредством корректного выполнения последовательности шагов вывода. Например, предположим, что агент эксплицитно знает конъюнкцию формул α и $(\alpha \rightarrow \beta)$, т.е. $K_i (\alpha \wedge (\alpha \rightarrow \beta))$. Во всех нормальных модальных системах формула $K_i (\alpha \wedge \beta)$ в этом случае будет теоремой. Однако в свете вышеизложенного мы теперь не можем сказать, что агент знает эту формулу эксплицитно. Можно только сказать, что если агент будет рассуждать корректно, тогда только он будет знать $(\alpha \wedge \beta)$. Пусть CE, CI, MP будут означать, соответственно, правило вывода элиминации конъюнкций, правило вывода введения конъюнкций и modus ponens, а «;» означает операцию композиции действий. Тогда теоремой будет $K_i (\alpha \wedge (\alpha \rightarrow \beta)) \rightarrow \langle CE; CI; MP \rangle K_i (\alpha \wedge \beta)$, а не $K_i (\alpha \wedge (\alpha \rightarrow \beta)) \rightarrow K_i (\alpha \wedge \beta)$ как при стандартном модальном подходе.

Стремясь к большей общности, автор динамической эпистемической логики Хо Нгок Дук ввел вспомогательное действие (модальный оператор) F_i , трактуемое следующим образом: выполнить некоторое базовое действие (какое - заранее неизвестно); повторить (с последующим выполнением) конечное число раз (но заранее неизвестно какое) недетерминированный выбор базового действия из всего множества таких действий. Таким образом подчеркивается, что нам необходимо знать только то, что для получения в будущем некоторого нового эксплицитного знания агенту необходимо выполнить некоторые шаги рассуждения. Заметим, что в динамической эпистемической логике данное вспомогательное действие ведет себя аналогично модальному оператору «иногда в будущем» $\langle F \rangle$ минимальной темпоральной логики **K4**. Формально язык динамической эпистемической логики, который называется $\mathcal{L}_N^{DE}$, определяется следующим образом.

Определение 6 (язык динамической эпистемической логики). Пусть $Agent = (1, \dots, N)$ множество агентов, $\mathcal{L}_N^{DE}$ есть минимальное множество такое, что

- 1) $\mathcal{L}_N^K \subseteq \mathcal{L}_N^{DE}$;
- 2) если $\alpha \in \mathcal{L}_N^{DE}$ то $\neg \alpha \in \mathcal{L}_N^{DE}$;
- 3) если $\alpha \in \mathcal{L}_N^{DE}$ и $\beta \in \mathcal{L}_N^{DE}$ то $(\alpha \rightarrow \beta) \in \mathcal{L}_N^{DE}$;
- 4) если $\alpha \in \mathcal{L}_N^{DE}$ и $i \in Agent$, то $\langle F_i \rangle \alpha \in \mathcal{L}_N^{DE}$.

Конъюнкция и дизъюнкция определяются обычным образом. $[F_i] \alpha$ является аббревиатурой для $\neg \langle F_i \rangle \neg \alpha$. Формула $\langle F_i \rangle \alpha$ читается «то, что α истинна, будет эксплицитно известно агенту i , если у него будет правильный ход мысли», формула $[F_i] \alpha$ читается «то, что α истинна, обязательно будет эксплицитно известно агенту i , при любом ходе его мысли».

Определение 7 (динамическая эпистемическая логика DEK_N). Пусть $\mathcal{L}_N^{K+}$ есть минимальное подмножество множества $\mathcal{L}_N^K$, включающее все объективные формулы и замкнутое относительно следующего условия: если $\alpha, \beta \in \mathcal{L}_N^{K+}$ и $i \in Agent$, то $\{(\alpha \wedge \beta), (\alpha \vee \beta), K_i \alpha\} \subseteq \mathcal{L}_N^{K+}$. **DEK_N** есть эпистемиче-

ская логика, имеющая следующие аксиомы и правила вывода (для всех $i \in Agent$):

- (PC1) $(\alpha \rightarrow (\beta \rightarrow \alpha))$;
 (PC2) $(\alpha \rightarrow (\beta \rightarrow \gamma)) \rightarrow ((\alpha \rightarrow \beta) \rightarrow (\alpha \rightarrow \gamma))$;
 (PC3) $(\neg \beta \rightarrow \neg \alpha) \rightarrow (\alpha \rightarrow \beta)$
 (TL1) $[F_i] (\alpha \rightarrow \beta) \rightarrow ([F_i] \alpha \rightarrow [F_i] \beta)$;
 (TL2) $[F_i] \alpha \rightarrow [F_i] [F_i] \alpha$;
 (DE1) $K_i \alpha \wedge K_i (\alpha \rightarrow \beta) \rightarrow < F_i > K_i \beta$ (замыкание относительно импликации);
 (DE2) $K_i \alpha \rightarrow [F_i] K_i \alpha$, для всех $\alpha \subseteq \mathcal{L}_N^{K+}$ (устойчивость);
 (DE3) $< F_i > (\alpha \rightarrow (\beta \rightarrow \alpha))$;
 (DE4) $< F_i > (\alpha \rightarrow (\beta \rightarrow \gamma)) \rightarrow ((\alpha \rightarrow \beta) \rightarrow (\alpha \rightarrow \gamma))$;
 (DE5) $< F_i > (\neg \beta \rightarrow \neg \alpha) \rightarrow (\alpha \rightarrow \beta)$.

Правила вывода:

- (MP) modus ponens: из α и $(\alpha \rightarrow \beta)$ выводится β ;
 (NEC_i) из α выводится $[F_i] \alpha$.

Аксиомы (PC1) - (PC3) вместе с (MP) аксиоматизируют классическое пропозициональное исчисление. Вместе с (TL1), (TL2) и (NEC_i) они образуют полную аксиоматизацию минимальной темпоральной логики транзитивного времени. Остальные аксиомы вместе с правилами вывода описывают динамику знаний. Аксиома (DE1) «говорит», что агенты умеют пользоваться правилом вывода modus ponens. Аксиома (DE2) «говорит», что агенты не забывают в будущем того, что они знают в настоящем. Аксиомы (DE3) - (DE5) устанавливают способность агентов пользоваться в своих рассуждениях аксиомами (PC1) - (PC3) классической логики.

Понятия доказательства, теоремы и непротиворечивого множества формул (по отношению к логике DEK_N) определяются обычным образом. Отношение доказуемости для DEK_N обозначается $\vdash^{DEK_N}$. Будем также говорить, что формула α PC-доказуема (обозначается $\vdash_{PC} \alpha$) если существует доказательство данной формулы с использованием только аксиом (PC1) - (PC3) и правила (MP).

Исследование свойств логики DEK_N , проведенное в [8] позволило установить, что она непротиворечива и свободна от проблемы логического всеведения во всех ее формах, рассмотренных в разделе 2. Последнее утверждение составляет сущность Теоремы 3.

Теорема 3.

1. Следующие правила вывода не являются общезначимыми в системе DEK_N :

- (NEC) Из α выводится $K_i \alpha$;
 (MON) Из $\alpha \rightarrow \beta$ выводится $K_i \alpha \rightarrow K_i \beta$;
 (CGR) Из $\alpha \leftrightarrow \beta$ выводится $K_i \alpha \leftrightarrow K_i \beta$.

2. Следующие формулы не доказуемы в системе DEK_N :

- $K_i (\alpha \rightarrow \beta) \rightarrow (K_i \alpha \rightarrow K_i \beta)$;
 $K_i \beta \rightarrow K_i (\alpha \rightarrow \beta)$;
 $K_i (\alpha \wedge \beta) \rightarrow K_i \alpha$;
 $K_i (\alpha \wedge \beta) \rightarrow K_i \alpha \wedge K_i \beta$;
 $K_i \alpha \rightarrow K_i (\alpha \vee \beta)$;
 $K_i \alpha \wedge K_i \beta \rightarrow K_i (\alpha \wedge \beta)$;
 $K_i \alpha \rightarrow K_i (\alpha \vee \beta)$;
 $K_i \alpha \vee K_i \beta \rightarrow K_i (\alpha \vee \beta)$;
 $K_i \neg \neg \alpha \rightarrow K_i \alpha$;
 $K_i \alpha \rightarrow K_i \neg \neg \alpha$;
 $K_i \alpha \rightarrow K_i K_i \alpha$;
 $\neg K_i \alpha \rightarrow K_i \neg K_i \alpha$.

Таким образом, агенты, описываемые данной логикой, не являются логически всеведущими. В тоже время есть основания считать их рациональными, так как в *принципе* они способны осознать все классические теоремы и вывести все следствия из имеющихся у них знаний при условии проведения корректных рассуждений. Этот факт устанавливается следующей (мета) теоремой [10] (под объективными понимаются формулы, не содержащие модальных операторов).

Теорема 4.

Пусть α, β есть объективные формулы:

- (NEC_{de}) Если $\vdash_{PC} \alpha$, то $\vdash^{DEK_N} < F_i > K_i \alpha$;
 (MON_{de}) Если $\vdash_{PC} \alpha \rightarrow \beta$, то $\vdash^{DEK_N} K_i \alpha \rightarrow < F_i > K_i \beta$;
 (MON*_{de}) Если $\vdash_{PC} \alpha \rightarrow \beta$, то $\vdash^{DEK_N} < F_i > K_i \alpha \rightarrow < F_i > K_i \beta$.

В [10] показано, что если α и β являются объективными формулами, то теоремами DEK_N , например, являются следующие формулы:

- 1) $K_i \beta \rightarrow < F_i > K_i (\alpha \rightarrow \beta)$;
- 2) $K_i (\alpha \wedge \beta) \rightarrow < F_i > K_i \alpha$;
- 3) $K_i (\alpha \wedge \beta) \rightarrow < F_i > K_i \alpha \wedge < F_i > K_i \beta$;
- 4) $K_i \alpha \rightarrow < F_i > K_i (\alpha \vee \beta)$;
- 5) $K_i \alpha \vee K_i \beta \rightarrow < F_i > K_i (\alpha \vee \beta)$;
- 6) $K_i \neg \neg \alpha \rightarrow < F_i > K_i \alpha$;
- 7) $< F_i > K_i \beta \rightarrow < F_i > K_i (\alpha \rightarrow \beta)$;
- 8) $< F_i > K_i (\alpha \wedge \beta) \rightarrow < F_i > K_i \alpha$;
- 9) $K_i \alpha \rightarrow < F_i > K_i (\alpha \vee \beta)$;
- 10) $< F_i > K_i \alpha \rightarrow < F_i > K_i (\alpha \vee \beta)$.

Однако формула $\langle F_i \rangle K_i \alpha \wedge \langle F_i \rangle K_i (\alpha \rightarrow \beta) \rightarrow \langle F_i \rangle K_i \beta$ не доказуема в $\mathbf{DEK}_N$ и, вопреки интуитивным ожиданиям, истинность формул $\langle F_i \rangle K_i \alpha$ и $\langle F_i \rangle K_i (\alpha \rightarrow \beta)$ не означает истинности $\langle F_i \rangle K_i \beta$. Говоря более общо, в $\mathbf{DEK}_N$ не является общезначимым никакое правило вывода, имеющее две или более посылки, каждая из которых становится известной после некоторых рассуждений. Эта проблема преодолевается с помощью добавления новой аксиомы:

(TL3) $\langle F_i \rangle [F_i] \alpha \rightarrow [F_i] \langle F_i \rangle \alpha$.

Данная аксиома называется аксиомой направленности. Система, получаемая из $\mathbf{DEK}_N$ с помощью добавления аксиомы **(TL3)** называется $\mathbf{DEK}_N^*$. В этой системе доказуемы формулы $\langle F_i \rangle K_i \alpha \wedge \langle F_i \rangle K_i (\alpha \rightarrow \beta) \rightarrow \langle F_i \rangle K_i \beta$ и $\langle F_i \rangle K_i \alpha \wedge \langle F_i \rangle K_i \beta \rightarrow \langle F_i \rangle K_i (\alpha \wedge \beta)$. Данное обстоятельство используется в [10] для доказательства (мета) теоремы, устанавливающей связь между системами $\mathbf{K}_N$ и $\mathbf{DEK}_N^*$.

Теорема 5.

Пусть $\alpha \in \mathcal{L}_N^K$ есть формула с модальной глубиной, не превосходящей 1. Пусть $\alpha' \in \mathcal{L}_N^{DE}$ – формула, полученная из α одновременной заменой K_i во всех вхождениях на $\langle F_i \rangle K_i$. Тогда α является теоремой $\mathbf{K}_N$ если и только если α' является теоремой $\mathbf{DEK}_N^*$.

Заключение

Рассмотренные выше подходы к решению проблемы логического всеведения образуют два направления. Одно из них основано на искусственном ослаблении дедуктивных возможностей агента, другое направление имеет в своей основе введение в логический язык специальных модальных операторов, трактуемых как необходимые для получения какого либо знания (выраженного в виде логических формул) умственные усилия.

Ослабление дедуктивных способностей агента хотя и решает полностью или частично (в зависимости от того, каким конкретно способом оно достигается) проблему логического всеведения, но в то же время порождает другие проблемы. В частности, границы дедуктивных способностей агента во всех случаях определяются вполне субъективно. При этом в ход идут такие категории, как «очевидные тавтологии», «простые правила вывода», «тривиальные следствия»

и т.д. Важно иметь в виду, что при наложении такого рода ограничений агент «навсегда» лишается возможности знать некоторые формулы, которые без этих ограничений составляли бы часть его имплицитных знаний и хотя бы потенциально могли быть агентом использованы. Таким образом, лишение агента способности проводить рассуждения, не являющиеся «очевидными» или «более или менее простыми» (при любом уточнении этих понятий) по сути равносильно его искусственному «оглуплению» – он становится «менее рациональным», а то и вовсе перестает им быть, если проблема логического всеведения оказывается решенной в полном объеме. Еще одна серьезная проблема заключается в том, что при ограничении дедуктивных способностей агента любым из известных способов, весьма сложно определить границы класса прикладных задач, решение которых будет после этого все еще по силам агенту с ограниченными способностями. Исключения составляют только вырожденные случаи, когда, например, агент вообще ничего не знает («total ignoramus» – «тотальный невежда»), не может делать даже элементарных выводов («complete idiot» – «законченный идиот»), не верит ни во что, кроме противоречий («ultimate fool» – «абсолютный дурак»). В каждом из этих случаев известно, что агенту вообще никакая прикладная задача не по силам, однако именно поэтому такие случаи не представляют практического интереса. Таким образом, ослабление дедуктивных способностей агентов если и позволяет избавить их от логического всеведения (подчас попутно избавляя их и от рационального поведения), но все же не делает по настоящему реалистичными.

Второе направление, имеющее в своей основе введение в логический язык специальных модальных операторов, не требует для решения проблемы логического всеведения каких-либо компромиссов в отношении дедуктивных особенностей агента. Сопоставляя языки $\mathcal{L}_N^{DE}$ и $\mathcal{L}_N^K$, можно сказать, что последний, в некотором смысле, играет роль внутреннего языка агентов, тогда как первый дает возможность рассуждать о меняющемся во времени состоянии их знаний некоему внешнему наблюдателю. Однако в части метарассуждений выразительные возможности языка $\mathcal{L}_N^{DE}$ ограничены. Они позволяют «наблюдателю» узнать, рассу-

жая о конкретном агенте, что после некоторого раздумья этот агент способен решить поставленную задачу. Но узнать, решит ли ее агент к требуемому сроку, «наблюдателю» не помогут уже никакие метарассуждения, отвечающие семантике **ДЕК***_N. О том, как данная проблема решается вне рамок модального подхода, будет изложено во второй части данной статьи.

Литература

1. В.Н. Вагин, Е.Ю. Головина, А.А. Загорянская, М.В. Фомина. Достоверный и правдоподобный вывод в интеллектуальных системах. Монография. Под ред. В.Н. Вагина и Д.А. Пospelova. М.: Физматлит, 2004, 704 с.
2. J. Barwise, On the model theory of common knowledge, The situation in logic, CSLI lectures notes, Center of the Study of Language and Information, 1989, pp. 201-220.
3. K. Binmore and H. S. Shin, Algorithmic knowledge and game theory, Knowledge, belief, and strategic interaction (C. Bicchieri and M.-L. Dalla-Chiara, eds.) MIT Press, 1990, pp. 15-32.
4. R. Fagin and J. Y. Halpern, Belief, awareness and limited reasoning, Artificial Intelligence **34** (1988), 39-76.
5. R. Fagin, J.Y. Halpern, Y. Moses, and M.Y. Vardi, Reasoning about knowledge, MIT Press, Cambridge, Mass., 1995.
6. E. Gillet and P. Gochet, Le logique de la connaissance: Le probleme de l'omniscience logique, Dialectica **47** (1993), no. 2-3, 143-171.
7. R. Goldblatt. Logics of time and computations. CSLI, Stanford, 1987.
8. G.E. Hughes and M.J. Cresswell, A new introduction to modal logic, Routledge, London, 1968.
9. J. Hintikka, Impossible possible worlds vindicated, Journal of Philosophical logic **4** (1975), 475-484.
10. D.N. Ho. Resource bounded reasoning about knowledge. Dissertation zur erlangung des Akademischen grades doctor rerum naturalium. Universitat Leipzig. 2001.
11. D. N. Ho, Logical omniscience vs. logical ignorance. On a dilemma of epistemic logic, Progress in Artificial Intelligence. Proceedings of EPIA'95, LNAI, vol. 990, Springer Verlag, 1995, pp. 237-248.
12. J.Y. Halpern and Y. Moses, A guide to completeness and complexity for modal logics of knowledge and belief, Artificial Intelligence **54** (1992), 319-379.
13. D. Kozen, J. Tiurzyn. Logics of program. Handbook of Theoretical Computer Science (J. van Leeuwen, ed.), vol. B, North-Holland, 1990, p.p. 789-840.
14. H. Levesque, A logic of implicit and explicit belief, Proceedings AAAI-84 (Austin, TX), 1984, pp. 198-202.
15. H.-S. Shin and Timothy Williamson, Representing the knowledge of turing machines, Theory and Decision **37** (1994), no. 1, 125-146.
16. H. Wansing, A general possible worlds framework for reasoning about knowledge and belief, Studia Logica **49** (1990), no. 4, 523-539.

Виньков Михаил Михайлович. Заведующий отделом Российского научно-исследовательского института информационных технологий и систем автоматизированного проектирования. Окончил в 1977 г. Московский авиационный институт, к.т.н., доцент. Автор 43 научных работ. Область научных интересов: динамические экспертные системы, индуктивное обучение, неклассические логики, формализация рассуждений.

Фоминых Игорь Борисович. Заместитель директора по научной работе Российского научно-исследовательского института информационных технологий и систем автоматизированного проектирования. Окончил МГТУ им. Н.Э.Баумана в 1963 г. и МГУ им. М.В.Ломоносова в 1969 г., д.т.н., профессор. Имеет 92 печатные работы, в том числе 3 монографии. Область научных интересов: динамические экспертные системы, нейроинформатика сети, неклассические логики, информационные проблемы мозга.