

УДК 164–165

DOI: 10.17223/1998863X/46/3

С.П. Ковалёв, А.В. Родин

ПРОБЛЕМА ОБОСНОВАНИЯ В ФОРМАЛЬНОМ ПРЕДСТАВЛЕНИИ ЗНАНИЙ¹

Понятие знание, которое используется в компьютерной науке, не включает в себя обычное для философской эпистемологии требование обоснования знания. Такое положение вещей имеет глубокие корни в истории философской логики в XX в. и приводит к негативным практическим последствиям. Исследования последних лет в области теоретико-доказательной семантики, гомотопической теории типов и формальной эпистемологии позволяют наметить теоретическую основу для решения этой проблемы.

Ключевые слова: представление знаний, обоснование, доказательство, теоретико-доказательная семантика, гомотопическая теория типов.

1. Понятие знания в философской эпистемологии и в компьютерной науке

В центре современной эпистемологической дискуссии (по крайней мере в аналитической философской традиции) находится восходящая к Платону так называемая ЖТВ-концепция знания, согласно которой знание – это обоснованное истинное мнение (Justified True Belief). Это определение расшифровывается так: эпистемический агент A знает, что p (где p это некоторое высказывание), если и только если выполнены следующие три условия:

- пропозиция p истинна (является истинным высказыванием);
- агент A верит, что p ;
- A имеет достаточное основание g , для того чтобы верить, что p .

Если речь идет об общем знании, то агент A может специально не указываться: в этом случае говорят, что « p известно».

Заметим, что в рамках ЖТВ-теории предполагается, что предмет знания в общем случае – это пропозиция. Такое знание называют пропозициональным знанием или знанием-что (knowledge that).

В связи с ЖТВ-теорией широко обсуждаются так называемые «проблемы типа Гетъе» [1], которые показывают, что понятие обоснования нуждается в дальнейшем уточнении. В частности, необходимы критерии, по которым можно было бы оценивать предложенное обоснование данного утверждения и признавать или не признавать его достаточным. Предпосылка ЖТВ-теории, что всякое знание является пропозициональным, также широко обсуждается и оспаривается многими философами. Оставляя в стороне примеры другого рода, заметим, что согласно ЖТВ-теории знание теоремы Пифагора требует умения доказывать эту теорему, а не только веры в то, что утверждение этой

¹ Статья подготовлена в рамках исследовательского проекта РФФИ «Логические и эпистемологические аспекты конструктивного знания», грант 16-03-00364-ОГН.

теоремы истинно. Однако знание доказательства теоремы (в смысле умения ее доказывать) – это уже не пропозициональное, а процедурное знание (знание-как). Таким образом, понятие обоснования, которое используется в JTB-теории, по крайней мере в некоторых важных случаях, требует включения в рассмотрение понятия процедурного знания.

Компьютерная наука в отличие от философской эпистемологии не ставит своей целью построить общую теорию знаний и рассуждений. Тем не менее в некоторых монографиях по компьютерному представлению знаний делаются попытки неформально определить эти базовые понятия [2–4]. Ни одно из таких определений, из числа тех, которые нам удалось найти в компьютерной литературе, не включают в себя требование JTB-теории, чтобы знание включало в себя обоснование. Хотя представители компьютерных наук в последние годы много говорят и пишут о знаниях, как правило, они не проводят никакого различия между знанием высказывания p и самим этим высказыванием.

Вместе с тем проблема обоснования знаний не является чисто теоретической и тем более чисто философской. Ненадежность распространяемой и получаемой через электронные средства коммуникации информации в последние годы стала социальной и даже политической проблемой. Некоторые философы и публицисты говорят в этой связи о наступлении «эпохи пост-правды» [5]. На наш взгляд, в данном случае мы имеем дело с недостатком существующих компьютерных и коммуникационных технологий, который может и должен быть исправлен. В этой статье мы не предлагаем технических решений, но анализируем истоки проблемы и показываем, что она имеет фундаментальный логический аспект. Далее мы описываем некоторые новые подходы в логической семантике, которые, на наш взгляд, могут послужить теоретической основой для будущих технических решений.

2. Логическое следование и логический вывод

В своей последней статье с говорящим названием «Пренебрежение эпистемологическими соображениями в логике» Горан Зундольм [6] показывает, что в мейнстриме развития философской логики XX в. эпистемологические соображения систематически игнорировались. Опираясь на анализ Зундольма, мы показываем, что «пренебрежение» проблемой обоснования в теории и практике компьютерного представления знаний связано с тем, что создатели концепции искусственного интеллекта и их последователи до сих пор ориентировались на системы классической логики и логической семантики, в которых эпистемологические соображения являются второстепенными по сравнению с онтологическими соображениями.

Отождествление рассуждения с синтаксической процедурой будет бессодержательным, если не объяснить, как и почему данная синтаксическая процедура символическим образом выражает некоторое рассуждение. Семантическая теория, которая отвечает на этот вопрос, была предложена Альфредом Тарским в классической работе «О понятии логического следования» 1936 г. [7]. Тарский определяет это отношение так:

Формула B логически следует из совокупности формул A_1, \dots, A_n , если всякая интерпретация теории, при которой формулы A_1, \dots, A_n интер-

претируются как истинные высказывания, также интерпретирует формулу B как истинное высказывание.

Тарский использует понятие логического следования в качестве эрзаца понятия логического *вывода*, не определяя это последнее понятие формально. Стандартные логические исчисления обладают свойством *корректности* (soundness): если формула B выводится (т.е. порождается в согласии с синтаксическими правилами дедукции) из формул A_1, \dots, A_n , то B логически следует из и формул A_1, \dots, A_n в смысле приведенного выше определения. Такое свойство логического исчисления позволяет думать о синтаксическом выводе (т.е. «механической» дедукции) как о логическом выводе, т.е. процедуре, которая позволяет устанавливать логические следствия из данных предпосылок. Такую семантику логического вывода называют *теоретико-модельной*, чтобы отличить ее от *теоретико-доказательной* семантики, о которой мы скажем ниже.

Теоретико-модельная семантика логического вывода не несет никакой эпистемологической нагрузки. Отношение логического следования по Тарскому – это отношение между формулами, которое формулируется в терминах истинных предложений и их классов. Классическое понятие истины, которое здесь используется, предполагает, что истинность предложения никак не зависит от того, знает какой-либо эпистемический агент это предложение или нет. Таким образом, данная логико-семантическая схема рассуждений не имеет никакого эпистемологического содержания.

Могут возразить, что в нашем кратком описании классической логической семантики мы просто забыли упомянуть о таком важном эпистемологически нагруженном понятии, как *доказательство*. Этим словом Гильберт и Бернайс [8] стали называть синтаксические выводы, т.е. цепочки формул, включающих ряд предпосылок A_1, \dots, A_n (они могут иметь статус *аксиом* теории), из которых в соответствии с правилами дедукции данной теории выводится искомая формула B : такую цепочку вслед за Гильбертом и Бернайсом часто называют доказательством формулы B на основании предпосылок (аксиом) A_1, \dots, A_n . Однако, как убедительно аргументирует Правиц [9], такое понятие доказательства является чисто техническим и также не имеет эпистемологического содержания. Чтобы формальные доказательства выполняли эпистемические функции и работали как доказательства в обычном смысле слова, на них необходимо накладывать дополнительные требования. Сами по себе эти синтаксические конструкции позволяют лишь определенным образом структурировать классы высказываний.

Такая структуризация совершенно необходима при компьютерном представлении знаний в форме теорий, поскольку любой компьютер обладает конечными ресурсами памяти и в принципе не способен хранить в явной форме все предложения теории [10]. За уменьшение расхода памяти приходится платить увеличением времени, требуемого для проверки истинности того или иного утверждения. Достижение баланса между расходом памяти и скоростью обработки относится к числу классических проблем теории алгоритмов и практики программирования, однако постановка и решение этой технической проблемы по-прежнему никак не учитывает эпистемическую функцию логического вывода.

3. Формальные онтологии

Теоретико-модельную логическую семантику в духе Тарского можно простым образом связать с онтологией, т.е. теорией о том, какого рода сущности существуют («бывают») в мире, по крайней мере в интересующей нас части мира. Для этого достаточно принять следующий тезис:

Для всякого истинного высказывания существует некоторая вещь (или вещи), которая делает это высказывание истинным.

Такую вещь по отношению к данному высказыванию в современной философской логике и аналитической метафизике принято называть *фактором истины* (truth-maker) этого высказывания, а приведенный выше тезис – тезисом *реализма факторов истины* (truth-maker realism, TMR).

TMR не решает вопроса о том, каков фактор или факторы истины каждого конкретного истинного предложения. Однако если речь идет о первопорядковой теории, то простейшим ответом на вопрос о факторах истины предложений этой теории будет онтологическое допущение о существовании индивидов (с их свойствами и отношениями), с помощью которых данная теория интерпретируется. Именно такой подход используется в *формальной* онтологии, которая в настоящее время является не только философской, но и компьютерной дисциплиной, имеющей приложения в искусственном интеллекте, и в первую очередь в области компьютерного представления знаний. Основным вариантом использования онтологии в инженерии компьютерных систем является проектирование структур информационного обеспечения [11]: баз данных, протоколов межпрограммного взаимодействия, форм пользовательского интерфейса и отчетов. Языки описания формальных онтологий предусматривают возможность задавать нетривиальные аксиомы и выводить из них утверждения, не записанные в структуре в явном виде.

Использование философских подходов наработок в теории и практике компьютерного представления знаний можно только приветствовать, но в данном случае вызывает недоумение тот факт, что для решения задач представления знаний компьютерная наука заимствует из философии именно онтологические, а не эпистемологические подходы.

4. Формальная эпистемология и конструктивная логика

В рамках описанной выше стандартной логической архитектуры компьютерного представления знаний истинностные значения приписываются высказываниям «с точки зрения Всезнающего Высшего Существа» без учета того, каким образом то или иное истинностное значение стало известным. Как нужно изменить эту стандартную архитектуру, чтобы учесть эпистемологическую сторону дела? Для решения этой задачи лучше подходит иная концепция логики, которую сегодня по историческим причинам принято называть *конструктивной*. Истина (предложения) понимается в конструктивной логике как существование доказательства (этого предложения), но при этом понятие доказательства понимается более широко, чем в классическом случае; в частности, доказательствами (или, как принято говорить в конструктивной логике, *свидетельствами*) при таком подходе могут считаться сами факторы истины. В последние годы были также построены новые семантики логического вывода, которые сегодня называют *теоретико-*

доказательными [12]. Идея такой семантики состоит в том, чтобы синтаксическим процедурам и правилам вывода одних формул из других поставить в соответствие подробно эксплицированные эпистемологические процедуры. Сохранение истинности при всех возможных интерпретациях по-прежнему мыслится как необходимое, но уже не как достаточное требование для того, чтобы данная синтаксическая процедура представляла логический вывод. С этим связано введение дополнительных (по сравнению со случаем теоретико-модельной семантики) синтаксических ограничений.

Представим информационную систему, которая не просто выдает пользователю в удобном виде некоторую полезную информацию по его запросу, осуществляя при этом поиск и обработку данных, но и предоставляет пользователю дополнительно соответствующее формальное обоснование в виде доступного для человеческого понимания и анализа описания выполненной процедуры. С помощью такого описания пользователь сможет (1) отождествить первоначальный источник нужной ему информации и (2) увидеть, по крайней мере в общих чертах, как именно информационная система обрабатывала исходные данные, чтобы ответить на его запрос [13]. Теоретико-доказательная семантика выводов может быть использована в такой ситуации для спецификации требований, при соблюдении которых предоставляемые компьютерной системой доказательства надежности информации будут валидными.

5. Гомотопическая теория типов как новая теоретическая модель для компьютерного представления знаний

Говоря о формальных выводах, мы подразумевали, что формулы, представляющие посылки и заключения таких выводов, интерпретируются как высказывания. Сейчас кратко опишем формальную систему ГТТ (гомотопическая теория типов), которая имеет более широкие семантические возможности. Синтаксис ГТТ в первоначальном варианте этой теории совпадает с синтаксисом теории типов Мартина – Лефа (ТТМЛ) [14]. Базовые формулы этого исчисления имеют вид $a : A$, где A называется *типом*, а a – *термом* данного типа. Сами такие формулы называются *суждениями* и, согласно первоначальной идее Мартина – Лефа, всегда допускают любую из следующих неформальных интерпретаций:

- a является доказательством пропозиции A ;
- a является элементом множества A ;
- a является решением задачи A ;
- a является реализацией намерения A .

Существенный прогресс в исследовании ТТМЛ, который заставил пересмотреть некоторые ключевые аспекты семантики этой теории, был достигнут в течение последнего десятилетия после того, как Владимир Воеводский установил связь между ТТМЛ и геометрической теорией гомотопий [15]. ТТМЛ и аналогичные теории с новой гомотопической семантикой сегодня принято называть *гомотопической теорией типов* (ГТТ). В рамках гомотопической семантики типы интерпретируются как абстрактные пространства (для которых определены основные понятия теории гомотопий), а термы – как точки таких пространств. ГТТ позволила установить, что «не все типы одинаковы»: только типы специального вида (гомотопического уровня), а

именно типы, содержащие самое большее единственный терм, можно отождествить с высказываниями; типы другого специального вида отождествляются с множествами; типы более высоких уровней мы оставим в стороне. Таким образом, первоначальная идея о том, что всякий тип допускает интерпретации как высказывание, множество и т.д., в контексте ГТТ не выглядит убедительной. Однако по всякому данному типу высшего порядка можно каноническим образом построить соответствующее высказывание, искусственно отождествляя все термы данного высшего типа. Такую процедуру в ГТТ называют *обрезанием* (truncation).

Покажем теперь, каким образом ГТТ может быть использована в качестве теоретической модели для систем представления знаний. Синтаксис ГТТ представляет собой систему правил, которые по отношению к пропозициональным типам (высказываниям) применяются как логические правила (правила вывода), а по отношению к типам более высоких гомотопических уровней – как правила построений непропозициональных конструкций, которые используются для верификации соответствующих высказываний. В настоящее время такой подход уже успешно используется для формализации и компьютерной проверки математических доказательств. Мы предполагаем, что он может быть использован и за пределами чистой математики. В задачах представления технологических знаний предлагаемый подход открывает возможность не только описывать технологические процедуры с помощью программного кода, но и моделировать на компьютере процесс изготовления изделия и проверять на этой модели, будет ли спроектированное изделие обладать требуемыми характеристиками.

6. Заключение

В последние годы в компьютерном представлении знаний широко используются подходы анализа больших данных, машинного обучения и искусственных нейронных сетей. Существует даже мнение, что классические средства представления знания, основанные на явном символьном выражении фактов и законов, безнадежно устарели и скоро будут заменены нейронными сетями, хранящими знания в неявном распределенном виде. Однако распределенное знание ненадежно и трудноверифицируемо, легко допускает ввод в заблуждение [16]. Поэтому, на наш взгляд, нейронные сети и другие средства анализа больших данных могут значительно обогатить, но никоим образом не заменить собой логические подходы и инструменты в компьютерном представлении знания.

Для успешного использования логических подходов в компьютерном представлении знаний недостаточно пользоваться классической логической архитектурой и связанными с этой архитектурой онтологиями, которые, как показывает наш анализ, полностью лишены всякого эпистемологического содержания. Хотя методы формальной эпистемологии пока еще недостаточно развиты, чтобы уже сегодня найти применение в компьютерной науке и технологии, совместная работа в этой области логически ориентированных философов и представителей компьютерных наук открывает многообещающие перспективы как в теоретическом, так и в практическом плане.

Литература

1. Hetherington S. Gettier Problems // Internet Encyclopedia of Philosophy. URL: <https://www.iep.utm.edu/gettier/> (дата обращения: 20.11.18).
2. Jakus G., Milutinović V., Omerović S., Tomažič S. Concepts, Ontologies and Knowledge Representation. Springer Science & Business Media, 2013. 67 p.
3. Abraham A., Grosan C. Intelligent Systems : A Modern Approach. Springer, 2011. 450 p.
4. Lakemeyer G., Nebel B. (Eds.) Foundations of Knowledge Representation and Reasoning. Springer, 1994. 355 p.
5. Fuller S. Post-Truth: Knowledge as a Power Game. CUP, 2018. 207 p.
6. Sundholm G. The Neglect of Epistemic Considerations in Logic: the Case of Epistemic Assumptions // Forthcoming in Topoi. Springer, 2018. URL: https://www.researchgate.net/publication/325547849_The_Neglect_of_Epistemic_Considerations_in_Logic_The_Case_of_Epistemic_Assumptions (дата обращения: 20.11.18).
7. Tarski A. On the Concept of Logical Consequence // Logic, Semantics, Metamathematics. Hackett Publ., 1983. P. 409–420.
8. Гильберт Д., Бернайс П. Основания математики. Т. 2 : Теория доказательств. М. : Наука, 1979. 557 с.
9. Prawitz D. On the Idea of the General Proof Theory // Synthese. 1974. 27, № 1–2. P. 63–77.
10. Muggleton S., De Raedt L. Inductive Logic Programming: Theory and Methods // The Journal of Logic Programming. 1994. Vol. 19–20. P. 629–679.
11. Ковалёв С.П. Применение онтологий при разработке распределенных автоматизированных информационно-измерительных систем // Автометрия. 2008. Т. 44, № 2. С. 41–49.
12. Piesha Th., and Schroeder-Heister, P. (Eds.) Advances in Proof-Theoretic Semantics. Springer, 2015.
13. Васильев С.Н. и др. Интеллектуальное управление динамическими системами. М.: Физматлит, 2000. 352 с.
14. Martin-Lof P. Intuitionistic Type Theory. BIBLIOPOLIS, 1984. 91 p.
15. Univalent Foundations Program, Homotopy Type Theory. IAS Princeton, 2013. 473 p.
16. Szegedy C. et al. Intriguing Properties of Neural Networks. CoRR abs/1312.6199. 2013.

Sergei P. Kovalyov, Institute of Control Sciences, Russian Academy of Sciences (Moscow, Russian Federation).

E-mail: kovalyov@sibnet.ru

Andrei V. Rodin, Institute of Philosophy, Russian Academy of Sciences (Moscow, Russian Federation).

E-mail: andrei@philomatica.org

Vestnik Tomskogo gosudarstvennogo universiteta. Filosofiya. Sotsiologiya. Politologiya – Tomsk State University Journal of Philosophy, Sociology and Political Science. 2018. 46. pp. 14–29.

DOI: 10.17223/1998863X/46/3

THE PROBLEM OF JUSTIFICATION IN FORMAL KNOWLEDGE REPRESENTATION

Keywords: knowledge representation; justification; proof; proof-theoretic semantics; homotopy type theory.

The concept of knowledge used in the computer science unlike the one widely used in today's philosophical epistemology (knowledge as justified true belief) does not involve the requirement according to which knowledge has to be justified. The neglect of this issue in the computer science originates from the philosophical logic of the 20th century where for various historical and conceptual reasons the concept of justification has been left out of the mainstream and as a consequence remained without a proper formal treatment. This situation has perceivable negative outcomes in the existing knowledge representation technologies, which makes a certain thinker become skeptical about the capacities of such technologies to convey truths or even about the truth concept itself. Recent research in the fields of proof-theoretic semantics, homotopy type theory and formal epistemology help us to outline a theoretical background for solving this problem.

References

1. Hetherington, S. (n.d.) *Gettier Problems*. [Online] Available from: <https://www.iep.utm.edu/gettier/>. (Accessed: 20th November 2018)

2. Jakus, G., Milutinović, V., Omerović, S., Tomažič, S. (2013) *Concepts, Ontologies and Knowledge Representation*. Springer Science & Business Media.
3. Abraham, A. & Grosan, C. (2011) *Intelligent Systems: A Modern Approach*. Springer.
4. Lakemeyer, G. & Nebel, B. (eds) (1994) *Foundations of Knowledge Representation and Reasoning*. Springer.
5. Fuller, S. (2018) *Post-Truth: Knowledge as a Power Game*. Cambridge University Press.
6. Sundholm, G. (n.d.) The Neglect of Epistemic Considerations in Logic: the Case of Epistemic Assumptions. *Topoi*. [in print]. [Online] Available from: https://www.researchgate.net/publication/325547849_The_Neglect_of_Epistemic_Considerations_in_Logic_The_Case_of_Epistemic_Assumptions. (Accessed: 20th November 18)
7. Tarski, A. (1983) On the Concept of Logical Consequence. In: Corcoran, J. (ed.) *Logic, Semantics, Metamathematics*. Hackett Publ. pp. 409–420
8. Hilbert, D. & Bernays, P. (1979) *Osnovaniya matematiki* [Foundations of Mathematics]. Vol. 2. Translated from German by N. Nagornyy. Moscow: Nauka.
9. Prawitz, D. (1974) On the Idea of the General Proof Theory. *Synthese*. 27(1-2). pp. 63–77. DOI: 10.1007/BF00660889
10. Muggleton, S. & De Raedt, L. (1994) Inductive Logic Programming: Theory and Methods. *The Journal of Logic Programming*. 19–20. pp. 629–679. DOI: 10.1016/0743-1066(94)90035-3
11. Kovalev, S.P. (2008) Domain Engineering of Distributed Measurement Systems. *Avtometriya*. 44(2). p. 41–49. (In Russian).
12. Piesha, Th. & Schroeder-Heister, P. (eds). *Advances in Proof-Theoretic Semantics*. Springer.
13. Vasiliev, S.N. et al. (2000) *Intel'lektnoe upravlenie dinamicheskimi sistemami* [Intellectual Control of Dynamic Systems]. Moscow: FISMATLIT.
14. Martin-Lof, P. (1984) *Intuitionistic Type Theory*. BIBLIOPOLIS.
15. Univalent Foundations Program. (2013) *Homotopy Type Theory*. IAS Princeton.
16. Szegedy, C. et al. (2013) *Intriguing Properties of Neural Networks*. CoRR abs/1312.6199.