Проблемы логики и методологии науки

УДК 165.4
DOI:
10.15372/PS20190304

В.В. Целищев, А.В. Хлебалин

САМОРЕФЛЕКСИЯ НЕПРОТИВОРЕЧИВОСТИ В БАЗИСНОЙ ЛОГИКЕ ФОРМАЛИЗОВАННОГО МАТЕМАТИЧЕСКОГО ТЕОРЕТИЗИРОВАНИЯ[1]

Статья посвящена проблеме саморефлексии формальных систем. Традиционное убеждение в том, что в отличие от человека формальная система или компьютер не способны к саморефлексии, анализируется через экспликацию понятия «саморефлексия» применительно к формальным системам. Показано, что логика доказательства позволяет средствами формальной системы выразить свойство саморефлексии системы.

Ключевые слова: саморефлексия, формальные системы, геделево предложение, самореференция

V.V. Tselishchev, A.V. Khlebalin

SELF-REFLECTION OF CONSISTENCY IN BASIC LOGIC OF FORMALIZED MATHEMATICAL THEORIZATION

The paper deals with the problem of self-reflection of formal systems. The traditional belief that unlike a human, a formal system / computer is not capable of self-reflection is analyzed through explication of the concept of self-reflection with regard to formal systems. It is shown that GL enables to use means of a formal system to express self-reflection of the system.

Keywords: self-reflection; formal systems; Gödel-sentence; self-reference

[1] Работа выполнена в рамках программ фундаментальных научных исследований, определяемых Президиумом РАН.

Саморефлексия является свойством человеческого сознания, которое желательно воспроизвести в системах искусственного интеллекта. Одна из существенных проблем, встающих на этом пути, заключается в том, что в реальном мышлении саморефлексия балансирует на грани парадоксальности и зачастую попросту впадает в нее. Парадоксальность означает наличие противоречивости, в то время как формальная система должна быть непротиворечивой. Выход из этого затруднения может быть найден через определение «саморефлексивности» компьютера и сопоставление ее с человеческими возможностями в этом отношении. Само по себе это сопоставление является довольно затруднительным ввиду расплывчатости соответствующих категорий как для человеческого мышления, так и для компьютера. Однако математическая логика представляет собой удачный объект для исследования такого рода вопросов в виде проблемы истинности так называемого геделева предложения. Известная постановка вопроса Р. Пенроузом (чья позиция, как оказалось, очень сходна с позицией самого Геделя), заключающаяся в том, что человек знает истинность этого предложения, а компьютер – нет, говорит о принципиальной невозможности достижения машиной саморефлексии. Однако при более точной постановке вопроса этот вердикт не является заключительным. Цель данной статьи состоит в демонстрации возможностей формальной системы в отношении саморефлексии, что реализуется в понятии «знания» ею собственной непротиворечивости. При этом мы исходим из того, что конструкция геделева предложения является парадигмальным случаем самореференции и, далее, саморефлексии формальной системы.

Каков класс аргументов при сопоставлении возможностей «видения» человеком и машиной истинности геделева предложения? В данном случае двумя главными концепциями выступают истина и доказательство, семантическая и синтаксическая составляющие категории. Интуитивно оправданно предположение, что компьютер не может выйти за синтаксические рамки. Но эта интуиция фактически подменяет проблему уже готовым ответом, потому что при этом подразумевается, что компьютер не способен к саморефлексии. Последнее и является спорным вопросом.

Исходно саморефлексия связана с самореферентными выражениями. Они были причиной парадоксов, из которых наиболее известным являются парадокс Рассела и парадокс Лжеца. В философских дискуссиях понятие самореферентности в метаматематике играет важную роль. Здесь самореференция является отдельной темой сама по себе. Некоторые авторы, и среди них Р. Хек [9] и П. Милн [12], фокусируются на самореференции в метаматематике и спрашивают, например, на самом ли деле данное математическое предложение утверждает свою собственную недоказуемость (или любое другое свойство). Даже более важно то, что понятие самореференции используется в анализе парадоксов, и часто выражается надежда на то, что метаматематическое понятие самореференции прольет свет также на другие проблемы неформальных языков [8].

Особое внимание на самореферентные утверждения было обращено в связи с техникой арифметизации синтаксиса Геделя, использованной им для доказательства своих теорем о неполноте арифметики. Кодирование синтаксических структур арифметическими средствами приводит к теоретическим выкладкам значительной сложности, что существенно затрудняло понимание природы самореференции, по крайней мере в случае геделевских теорем. Выходом из ситуации оказалось открытие так называемой логики доказательства, или модальной системы GL [4]. В определенном смысле это было крупным прорывом в проблематике самореферентности и других важных вопросов логики. Действительно, GL, по мнению А. Виссера, является настоящим «чудом»:

> «Свершилось чудо! С одной стороны, мы имеем класс фантастически сложных теорий в логике предикатов, теорий “с достаточно кодирующим потенциалом”, типа арифметики Пеано или теории множеств Цермело – Френкеля. С другой стороны, мы имеем определенные модальные пропозициональные теории поразительной простоты. Мы переводим модальные операторы модальных теорий в определенные специфические, фиксированные предикаты предикатной логики. Эти специальные предикаты содержат астрономическое число символов. Мы интерпретируем пропозициональные переменные как произвольные предикатные логические последовательности. И мы видим, что при такой интерпретации модальные теории обладают обоснованностью и полнотой. Они кодифицируют в точности схематические принципы в пределах их области действия. Так что чудеса все-таки случаются. Наше чудо включает переход в другую субстанцию. Мы осуществляем перевод между языками несравнимой сигнатуры. Модальный язык не содержит кванторов, а предикатные логические теории не содержат модальных операторов. Модальные операторы могут быть переведены в предикаты, потому что мы транссубстанционируем формулы, входящие в область действия модального оператора, в замкнутые термы (нумерические символы), представляющие собой коды геделевских номеров формул переводимой теории» [16, p. 1].

Геделево предложение, как известно, балансирует на грани парадокса. Самый простой способ демонстрации этого представлен в упрощенной версии GL Р. Смаллианом [3]. Сигнатура этой логики включает помимо пропозициональных связок константу $\perp$ (ложь) и предикат доказуемости B. В рамках этой логики предикат доказуемости B используется для формирования внешне парадоксального геделева предложения $p \equiv \neg Bp$. Это предложение эквивалентно утверждению о своей собственной недоказуемости. Такого рода саморефлексия представляет собой важную характеристику определенного класса формальных языков, которые лежат в основе того, что считается нами базовой логикой.

Рефлексия как свойственная сознанию процедура может имитироваться в понятийном аппарате логики доказательства. Для этой цели нужно ввести иерархию стадий «самосознания» с постепенным усилением этого самосознания до того момента, когда мы можем сказать, что формальная система «размышляет» над степенью обоснованности своих утверждений. Смаллиан приводит четыре типа систем, получающихся в ходе эволюции ко все большим выразительным свойствам систем. Система «доказывает», что принадлежит к типу 1, если в ней доказуемы все предложения формы BX, где X – любая тавтология, а также доказуемы все предложения вида $Bp\&B(p\supset q)\supset Bp$. Система «доказывает», что принадлежит к типу 2, если в ней доказуемы все предложения вида $B(Bp\&B(p\supset q)\supset Bp)$. Система «доказывает», что принадлежит к типу 3, если в ней доказуемы все предложения формы $Bp\supset BBp$. Система «доказывает», что принадлежит к типу 4, если в ней доказуемы все предложения формы $B(Bp \supset BBp)$.

Знание состоит в переходе от одного синтаксического уровня к другому, так что новые истинные предложения становятся доказуемыми в более широкой системе. Разумеется, эти новые предложения являются интуитивными предположениями о природе дедуктивных систем. Так, наиболее дискуссионная «новая» истина заключается в том, что если некоторое утверждение доказуемо в системе, то доказуемо в системе, что оно доказуемо. Основная мотивация перехода от одной системы к более «осознающей» состоит в том, что система типа n не «знает», что она типа n. И только следующая система n+1 «знает», что предшествующая ей система принадлежит к типу n. Тогда вполне естественно сказать, что система «знает», к какому типу она принадлежит, если она доказывает свою принадлежность к этому типу и на самом деле принадлежит к этому типу. Из перечисленных систем таковой является система 4. Как уже указывалось, это обстоятельство в значительной степени обязано доказуемости утверждения $B(Bp\supset BBp)$, эпистемическая суть которого состоит в значимости принципа $Bp\supset BBp$. Этот принцип действительно важен, являясь самой важной частью условий выводимости Гильберта – Бернайса для доказательства второй теоремы Геделя о неполноте арифметики.

Вторая теорема Геделя о неполноте утверждает невозможность доказательства непротиворечивости формальной системы средствами этой формальной системы. «Размышления» о непротиворечивости формальной системы в рамках самой формальной системы являются наиболее ярким примером рефлексии, о которой говорилось выше. Любопытен механизм этой рефлексии, точнее, саморефлексии формальной системы [11]. Пусть имеется некоторое интуитивно истинное утверждение R, которое подлежит формализации. Далее, пусть $\vdash R$ означает, что система неопровержимо утверждает R. Тогда разумными представляются следующие утверждения.

Разумно предположить, что если система обладает выразительной силой, чтобы утверждать R, она имеет силу для утверждения $B(R)$, т.е. если в системе утверждается нечто, у нее есть возможности для доказательства этого. Символически (1) Если $\vdash R$, тогда $\vdash B(R)$.

Далее, предполагается, что система включает правило modus ponens. Символически (2) $\vdash B[(R_1)\&(B(R_1\to R_2)]\supset B(R_2)$.

Наконец, последнее правило говорит о том, что система «знает» (1), символически (3) $\vdash B(R)\supset B[B(R)]$.

Это вполне разумные предположения, и они и есть правила выводимости Гильберта – Бернайса. Возникает вопрос: какой смысл имели эти правила как дополнение классической логики? Дело в том, что они нужны тогда, когда мы имеем дело с саморефлексивными утверждениями, там, где обычная логика приводит к парадоксам (например, парадоксу Лжеца).

Саморефлексивность появляется в конструировании Геделем предложения G, которое содержательно утверждает собственную недоказуемость. Символически $G\equiv\neg B(G)$. Парадоксальность этого предложения не связана непременно с тем, что предложение доказывает то, что оно недоказуемо. Дело в том, что можно сконструировать аналогичное предложение с «утвердительным оттенком», а именно «предложение Генкина» формы $H\equiv B(H)$, которое утверждает свою собственную доказуемость. Доказуемость этого предложения была продемонстрирована в так называемой теореме Леба, приведшей к важнейшим результатам в отношении саморефлексии.

По аналогии с утверждениями мы можем говорить и о рефлексивных системах, потому что такими системами как раз и являются, например, арифметика Пеано и теория множеств Цермело – Френкеля. Важно то, что именно к такому классу систем принадлежит система, в которой доказуемы теоремы Геделя о неполноте. Каковы характерные черты рефлексивных систем? Система рефлексивна, если для каждого предложения q имеется по крайней мере одно предложение p, такое, что в системе доказуемо $p \equiv (Bp \rightarrow q)$. В этой терминологии устанавливается упомянутая выше теорема Леба:

> В каждой рефлексивной системе типа 4, если $Bp \supset p$ доказуемо в этой системе, то доказуемо и p.

Именно к таким системам относятся саморефлексивные предложения Геделя и Генкина. Интерес представляет некоторое усиление понятия рефлексивности формальной системы. В частности, система строго рефлексивна, если для каждого предложения q имеется предложение p, такое, что $p \equiv B(p \supset q)$ доказуемо в этой системе. Система называется лебовской в следующем случае: если в системе доказуемо предложение p, то в ней доказуемо предложение $Bp \supset p$. Из этого вытекает, что каждая рефлексивная система типа 4 является лебовской.

Легко заметить, что приведенные определения в значительной степени близки друг к другу. Но особенность логики, как и математики, состоит в том, что кажущиеся только нотационными модификации приводят к интересным результатам. Так, в частности, в несколько шагов логического доказательства можно показать фундаментальные результаты, например то, что теорема Геделя о неполноте является частным случаем теоремы Леба. Напомним, что результаты Леба в интуитивном плане даже более естественны, поскольку они касаются доказательства предложения Генкина $p \equiv B(p)$, которое не связано с парадоксальностью доказательного отрицания собственной недоказуемости. Рассмотрим следующий логический вывод из посылки о непротиворечивости системы в виде невозможности доказательства лжи, символически (1) $\neg B\bot$. Из (1) получаем эквивалентное ему выражение (2) $B\bot \supset \bot$. И по теореме Леба получаем (3) $\bot$. Противоречие между (1) и (3) говорит о том, что невозможно доказательство непротиворечивости системы в самой системе. Вторая теорема Геделя о неполноте оказывается следствием теоремы Леба, которая возникла из соображений о рефлексивности формальной системы, не связанной напрямую с геделевскими конструкциями.

Механизм «размышления» системы о собственной непротиворечивости можно продемонстрировать на следующем примере. Принимая во внимание довольно естественное свойство так называемой регулярности, а именно то, что из $p \supset q$ доказуемо $Bp \supset Bq$, получаем:

1. $\neg Bp$	посылка
2. $\bot \supset p$	тавтология
3. $B\bot \supset Bp$	2, регулярность
4. $\neg Bp \supset B\bot$	3, обращение
5. $\neg Bp \supset (B\bot \supset \bot)$	4
6. $B\bot \supset \bot$	1 и 5
7. $\bot$	6, лебовская система

Заключение противоречит предположению о непротиворечивости системы, и, стало быть, посылка должна быть отброшена. Факт получения противоречия записывается в виде $B\neg Bp \supset B\bot$.

Такого рода заключения о формальной системе, полученные «внутри» нее самой, позволяют употреблять термины «рефлексия», «самореференция» в отношении символических структур. Но важно понять, что сами эти термины принадлежат к естественному языку и относятся к неформальному, интуитивному пониманию концепций. Таким образом, можно считать, что геделево предложение использует концепцию самореференции, идея которой выразима на естественном языке (геделево предложение «говорит о самом себе»). С точки зрения математики геделево предложение в арифметике Пеано представляет собой универсальную квантификацию, справедливую для всех натуральных чисел. Возникает вопрос: в какой степени на формальное представление геделева предложения можно перенести идею самореференции? Здесь действительно есть проблема, поскольку соотносимыми при самореференции в этом предложении оказываются разные категории, а именно синтаксические (само предложение) и числа в системе геделевской нумерации (геделев номер предложения). Строго говоря, тут нет самореференции.

Связь синтаксических структур с арифметическими осуществляется Геделем путем кодирования первых вторыми. Но понимание природы кодирования выходит за пределы собственно арифметики. Тогда геделево предложение, претендующее на самореферентность, является в определенном смысле артефактом процедуры кодирования. Артефакт по своей природе может не соответствовать исходным целям символических конструкций. И интуитивное восприятие геделевской конструкции затруднено именно в силу этого фактора. Действительно, «трудность восприятия теоремы (Геделя. – *Авт.*) связана с тем, что реальное арифметическое построение геделева предложения, создаваемое по прямому рецепту Геделя, может оказаться исключительно трудным для понимания и не представлять математического интереса. Поэтому даже сами математики зачастую относятся к таким конструкциям с полным пренебрежением» [2, с. 180].

В литературе существует масса попыток дать более или менее интуитивное представление о механизме получения самореферентных предложений типа геделева. В этих попытках самое трудное состоит в «увязывании» семантических и синтаксических понятий. Весьма удачным можно считать пример конструирования геделева предложения, демонстрирующий механизм самореференции [3]. Пусть k есть предложение, говорящее, что некоторое другое предложение p, могущее быть истинным или ложным, истинно. Очевидно, что k истинно, когда истинно p, и ложно, когда ложно p. На языке пропозициональной логики это обстоятельство выражается через $k \equiv p$. Если в формальной системе утверждается некоторое предложение q, тогда в этой системе предложение $k \equiv q$ истинно. Ясно, что k есть утверждение метаязыка, но это обстоятельство не препятствует установлению эквивалентности $k \equiv q$, поскольку это выражение является переводом утверждения о том, что утверждается предложением q.

В этом последнем «невинном» замечании присутствует вся машинерия геделевской конструкции, включая эффект кодирования. Эффект кодирования в данном примере трансформируется в то обстоятельство, что в формальной системе $k \equiv q$ доказуемо. Такая формальная система подразумевает два свойства: во-первых, это обоснованность, т.е. все доказуемые утверждения являются истинными; во-вторых, это полнота, т.е. все истинные утверждения доказуемы. Неверно было говорить напрямую, что в данном случае мы имеем обоснованность и полноту, поскольку ситуация осложнена «переводом». Переменная k есть часть метаязыка, а q принадлежит к объект-языку. Установление их взаимосвязи и есть результат «перевода». В какой степени два указанных свойства формальной системы совместимы, зависит от силы формальной системы. В частности, Гедель показал, что полнота в его формальной системе не выполняется.

Предлагаемый Смаллианом механизм позволяет сконструировать парадоксальное предложение $k \equiv \neg Bk$, где B есть предикат доказуемости.

Парадоксальность такого рода утверждений обсуждалась ранее. Но является ли это предложение на самом деле парадоксальным? Рассмотрим следующий логический вывод:

1.	$k \equiv \neg Bk$	посылка
2.	$\neg k$	посылка
3.	Bk	1 и 2
4.	k	обоснованность
5.	$k \,\&\neg\, k$	2 и 4
6.	$k \,\&\neg\, k$	закон исключенного третьего
7.	k	2, 5 и 6 – отказ от 2
8.	Bk	1-7 – доказательство k
9.	$\neg k$	1, 8
10.	$k \,\&\, \neg k$	7 и 9 – противоречие

В данном выводе, приближенном в своей буквальности к «машине», мы получаем парадокс. На самом деле, парадокс возникает за счет незаконного перехода от 3 к 4. Этот переход предполагает обоснованность формальной системы, т.е. существования доказательства, что p, достаточно для заключения об истинности p. Но в процессе доказательства формальная система «не знает», что она обоснована, и, значит, неверен переход от 3 к 4. Это свидетельствует о важности элементов «саморефлексии» формальной системы. Следующей ступенью в саморефлексии подобного рода является наличие в ней доказательства, что система обоснована, т.е. доказательства того, что в ней недоказуемы ложные утверждения. Другими словами, для любого предложения p в системе доказуемо, что если p доказуемо, то p должно быть истинным. Повышение «саморефлексии» формальной системы должно привести к ослаблению парадоксальности фактически геделева предложения $k \equiv \neg Bk$.

В свете такой методологии мы подходим к одному из самых важных вопросов о рефлексии формальной системы: может ли непротиворечивая формальная система «знать» о собственной непротиворечивости? Можно показать, что ни в одной лебовской формальной системе типа 4 не является доказуемым утверждение о непротиворечивости этой системы. Действительно, представленное Смаллианом описание переводного устройства как аналога конструирования геделева предложения показало, что многие проблемы могут быть сведены к пропозициональной модальной логике.

Геделево предложение и парадокс Лжеца – очевидные примеры самореферентных предложений, позволившие установить ограничения на саморефлексии формальной системы. Вместе с тем подлинным источником самореференции является диагональный метод: саморефрентное предложение имплицитно определяется диагонализацией, через которую оно указывает на самое себя. Имплицитные определения трансформируются в эксплицитные через вычисление. Представление самореферентных предложений как фиксированных точек позволяет превращать их в предмет вычисления, причем сами вычисления в определенных случаях могут быть формализованы в самой теории. К. Сморинский выразил суть такого подхода к самореферентным предложениям следующим лозунгом: «Вычисляйте фиксированные точки!» [14]. Исторически первым примером использования вычисления фиксированной точки была вторая теорема Геделя, вторым – лебовское решение проблемы Хенкина. Как указывает К. Сморинский, «в каждом из этих случаев результаты затемняли вычисления и, как следствие, проблема вычисления самореферентных предложений игнорировалась» [14, p. 18]. Развитие вычислительного подхода к самореферентным предложениям стало возможным в результате экспансии модальной логики в арифметику, что позволило поставить вопрос об эксплицитной вычислимости самореференции прежде всего в рамках самой формальной системы. Представление саморефрентного предложения как фиксированной точки достигает наибольшего развития в бестиповых теориях истины.

Возможность приписывать истинность внутри формальной системы ее собственным предложениям является частью концепции саморефлексии системы. Существенные ограничения на эту возможность выявляет семантическая теория истины А. Тарского. Он формулирует известное требование к адекватной теории истины, выраженное *Конвенцией T*: p – истинное предложение L, если и только если φ (где p – имя φ, а переменная φ пробегает по всем предложениям L). В своей теории Тарский дает начало типовому подходу в теории истины, согласно которому понятие «быть истинным в языке L» может быть определено только в более богатом метаязыке L^*, а его теорема о неопределимости понятия истинного предложения языка L в самом L демонстрирует ограничения саморефлексии любой достаточно богатой формальной системы в отношении истинности собственных предложений.

Результат Тарского в отношении определения понятия истины во многом аналогичен результату Геделя в отношении ограничений доказуемости для формальной системы. Оба они, получив обозначение ограничительных, демонстрируют существенные ограничения на саморефлексию формальных систем. И оба результата получат широкое распространение в других областях философской логики. Например, теорема о неопределимости Тарского будет распространена на случай эпистемической логики в связи с парадоксом Каплан – Монтегю [15]. Но наравне с этим нарастающим влиянием результата Тарского можно заметить развитие альтернативной ему тенденции избежать обращения к типовому подходу. Одно из самых первых проявлений этой тенденции можно увидеть уже в 1930-х годах, и оно связано с разработкой бестипового лямбда-исчисления [5], в котором функции применимы к самой себе как к объектам. Это оказалось одним из самых ранних проявлений становления бестипового подхода, который ныне нужно рассматривать как доминирующий в формальных теориях истины.

Основу бестиповых теорий составили теории С. Крипке, Х. Херцберга и А. Гупты. Наиболее частым аргументом в пользу бестиповых теорий является указание на неестественность типов. Но такой аргумент представляется скорее идеологическим или даже риторическим, ведь аксиоматические теории, как типовые, так и бестиповые, чаще всего имеют дело с арифметикой Пеано или иными формализованными языками. Существуют сугубо технические преимущества бестиповых теорий перед типовыми. Так, бестиповые теории в гораздо большей степени обеспечивают общность утверждений. Например, если мы хотим утверждать что-либо обо всех предложениях теории, мы хотим это утверждать и о предложениях, включающих предикат истины, а не усложнять теорию дополнительными средствами, которые требуются для языка, содержащего типы. Бестиповая теория истины позволяет легче и естественнее делать обобщения: С. Феферман [6] показал, что бестиповые теории способны доказать много больше обобщений, чем типовые.

Бестиповые формальные теории истины являются подлинной альтернативой типовому подходу. Теорема о неопределимости Тарского имеет своим следствием то, что не может быть частичной модели M, такой, что $M \models \varphi \leftrightarrow T(\varphi)$ для всех $\varphi \in L_T$, но может быть такая модель M, что для всех $\varphi \in L_T$ справедливо $M \models \varphi \Leftrightarrow M \models T(\varphi)$. С. Крипке доказывает существование такой модели, названной моделью неподвижной точки и кладет начало семантике неподвижной точки в формальных исследованиях истины. Модель неподвижной точки Крипке начинает строиться с определения двух множеств предложений S_{1e} и S_{2a} – истинных и ложных соответственно.

Рассмотрим модель $M_1 = \langle \mathbb{N}, S_{1e}, S_{1a} \rangle$. Эта модель генерирует новую модель M_2. Рассмотрим множества $S_{2e} = \{\varphi \in LT \mid M1 \mid= \varphi\}$ и $S_{2a} = \{\varphi \in LT \mid M1 \mid= \neg\, \varphi\}$, тогда второй частичной моделью будет $M_2 = \langle \mathbb{N}, S_{2e}, S_{2a} \rangle$. Модель M_2 аналогично генерирует третью модель, и так далее. В итоге мы достигаем в итоге фиксированной точки, в которой не получаем новых истины и ложности при генерировании следующей модели. Крипке утверждает, что стадии генерирования моделей фиксированной точки отражают способ обучения значению предиката истины носителем языка.

Мы не будем здесь подробно освещать построение моделей и достижение фиксированной точки (типичное изложение построения теории и оценивания ее логико-математического аппарата см. в [5]). Сам Крипке так характеризует свой результат: «На начальной стадии (в L_0) $T(x)$ является полностью неопределенным. Это соответствует начальной стадии, в которой субъект не владеет пониманием понятия истины. Задавая характеристику истины с помощью правил оценивания Клини, субъект может легко подняться на уровень L_1, т.е. он может оценивать различные предложения как истинные или ложные без какого-либо знания о $T(x)$ – в частности, он может оценивать все те предложения, которые не содержат $T(x)$. Как только оценивание произведено, он сразу же расширяет $T(x)$, как в L_1. Затем он может использовать новую интерпретацию $T(x)$, с тем чтобы большее число предложений оценить как истинные или ложные и расшириться до L_2 и т.д. В конце концов, процесс "насыщается", субъект достигает фиксированной точки L_σ. (*Будучи фиксированной точкой, L_σ является языком, который содержит свой собственный предикат истины.*)» [1, с. 169].

Отличие стратегии Крипке от стратегии Тарского в том, что в случае иерархии типовых теорий мы имеем дело с разными предикатами истины на разных уровнях, тогда как в теории Крипке используется единственный предикат истины. Теория Крипке трансформировала синтаксическую иерархию предиката истины Тарского в семантику бестипового предиката истины так же, как итеративная концепция в своем понятии перечисления синтаксическую теоретико-типовую иерархию Рассела. «Уместным будет сказать, что теория Крипке соотносится с теорией Тарского так же, как *ZFC* соотносится с теорией типов» [5].

Ограничения на саморефлексию формальных систем (как в случае с понятием истины, так и в случае с понятием доказуемости) основаны на концепции самореферентных предложений. В случае концепции истины это предложение Лжеца, в случае теорем о неполноте – геделево предложение. Теория истины С. Крипке послужила источником подхода к трактовке этих предложений, выражающих ограничения на саморефлексию формальной системы, который уточняет традиционную интерпретацию ограничительных результатов. Х. Гайфман предложил следующий подход к интерпретации истинности предложения Лжеца [7]. Он принимает в расчет обозначения Лжеца (Liar's tokens). Пусть c обозначает $\neg T(c)$, тогда предложение будет лишено значения истинности, и пусть b будет другим именем того же самого предложения $\neg T(c)$, тогда $\neg T(b)$ будет обладать значением истины – истинно. Аналогичным может быть подход в концепции доказуемости: если в исходной теории T доказуемо $c =$ '$\neg Pr(c)$', тогда $\neg Pr(c)$ будет недоказуемо; но если b будет другим именем $\vdash_T b = c$, тогда $\neg Pr(b)$ будет доказуемо.

Теория истины С. Крипке стимулировала бурное развитие бестиповых теорий и их аксиоматических построений. Нам нет необходимости здесь характеризовать его (см. [13]). Экспансия богатых средств математической логики в область логики философской приводит в последние десятилетия к очень интересным результатам, свидетельствующим о существовании тесной концептуальной связи между центральными концепциями философской логики. Одним из вызывающих интерес направлений исследования такой связи является именно исследование связи концепций истины и доказуемости, в котором целью является создание интегрированной теории бестиповой истины и доказуемости [10].

Сказанное позволяет представить концепцию саморефлексии формальной системы предметом логико-философского исследования. Типичное отрицание самой возможности такой саморефлексии оказывается несостоятельным. Безусловно, предложенные выше идеи нуждаются в кропотливой технической разработке. Тем не менее философское значение их уже вполне понятно: они демонстрируют, что определенная степень саморефлексии присуща формальным системам.

Литература

1. *Крипке С.* Очерк теории истины // Язык. Истина. Существование / Под ред. Суровцева В.А. – Томск: Изд-во ТГУ, 2002. – С. 151–183.
2. *Пенроуз Р., Хокинг С., Шимони А.Ю., Картрайт Н.* Большое, малое и человеческий разум. – М.: Мир, 2004. – 191 с.
3. *Смаллиан Р.* Вовеки неразрешимое. – М.: Канон+, 2013. – 303 с.
4. *Boolos G.* The Logic of Provability. – Cambridge: Cambridge University Press, 1996. – 316 p.

5. *Cantini A*. Paradoxes, self-reference and truth in the twentieth century // Handbook of the History of Logic. – 2009. – Vol. – 5. –P. 875–1013.

6. *Feferman S.* Reflecting on incompleteness // Journal of Symbolic Logic. – 1991. – Vol. 56. – P. 1–49.

7. *Gaifman H.* Pointers to propositions // Circularity, Definition, and Truth / Ed. by A. Chapuis & A. Gupta. – New Delhi: Indian Council of Philosophical Research, 2000. – P. 79–121.

8. *Halbach V. Visser A.* Self-reference in arithmetic // The Review of Symbolic Logic. – 2014. – Vol. 7, No. 4. –P. 671–697.

9. *Heck R.* Self-reference and the language of arithmetic // Philosophia Mathematica. – 2007. – Vol. 15, No. 1. – P. 1–29.

10. *Leitgeb H.* On the probabilistic convention *T* // Review of Symbolic Logic. – 2008. – No. 1. – P. 218–224.

11. *Lindstrem P.* Penrose's new argument // Journal of Philosophical Logic. – 2001. – Vol. 30. – P. 241–250.

12. *Milne P.* On Gödel sentences and what they say // Philosophia Mathematica. – 2007. – Vol. 15, No. 2. –P. 193–226.

13. *Sheard M.* A guide to truth predicate in modern era // The Journal of Symbolic Logic. – 1994. – Vol. 59, No. 3. –P. 1032–1054.

14. *Smorynski C.* Calculating self-referential statement, I: Explicit calculation // Studia Logica. – 1979. – Vol. 38, No. 1. – P.17–36.

15. *Thomason R.* A note on syntactical treatments of modality // Synthese. – 1980. – Vol. 44. – P. 391–395.

16. *Visser A.* An overview of Interpretability Logic. Advances in Modal Logic // Ed. by M. Kracht, M. de Rijke, H. Wansing & M. Zakharyaschev. – CSLI Publications, 1998. – Vol. 1. – P. 307–359.

References

1. *Kripke, S.* (2002). Ocherk teorii istiny [Outline of a theory of truth]. In: Surovtsev, V.A. (Ed.). Yazyk. Istina. Sushchestvovanie [Language. Truth. Life]. Tomsk, Tomsk State University Publ., 151–183. (In Russ.).

2. *Penrose, R., S. Hawking, A. Shimony & N. Cartwright.* (2004). Bolshoe, maloe i chelovecheskiy razum [The Large, the Small and the Human Mind]. Moscow, Mir Publ., 191. (In Russ.).

3. *Smallyan, R.* (2013). Voveki nerazreshimoe [Forever Undecided]. Moscow, Kanon+ Publ., 303. (In Russ.).

4. *Boolos, G.* (1996). The Logic of Provability. Cambridge, Cambridge University Press, 316.

5. *Cantini, A*. (2009). Paradoxes, self-reference and truth in the twentieth century. Handbook of the History of Logic, 5, 875–1013.

6. *Feferman, S.* (1991). Reflecting on incompleteness. Journal of Symbolic Logic, 56, 1–49.

7. *Gaifman, H.* (2000). Pointers to propositions. In: Chapuis, A. & A. Gupta (Eds). Circularity, Definition, and Truth. New Delhi, Indian Council of Philosophical Research, 79–121.

8. *Halbach, V. & A. Visser*. (2014). Self-reference in arithmetic. The Review of Symbolic Logic, Vol. 7, No. 4, 671–697.

9. *Heck, R.* (2007). Self-reference and the language of arithmetic. Philosophia Mathematica, Vol. 15, No. 1, 1–29.

10. *Leitgeb, H.* (2008). On the probabilistic convention *T*. Review of Symbolic Logic, 1, 218–224.
11. *Lindstrem, P.* (2001). Penrose's new argument. Journal of Philosophical Logic, 30, 241–250.
12. *Milne, P.* (2007). On Gödel sentences and what they say. Philosophia Mathematica, Vol. 15, No. 2, 193–226.
13. *Sheard, M.* (1994). A guide to truth predicate in modern era. The Journal of Symbolic Logic, Vol. 59, No. 3, 1032–1054.
14. *Smorynski, C.* (1979). Calculating self-referential statement, I: Explicit calculation. Studia Logica, Vol. 38, No. 1, 17–36.
15. *Thomason, R.* (1980). A note on syntactical treatments of modality. Synthese, 44, 391–395.
16. *Visser, A.* (1998). An overview of Interpretability Logic. In: Kracht, M., M. de Rijke, H. Wansing & M. Zakharyaschev (Eds.). Advances in Modal Logic. CSLI Publications. Vol. 1, 307–359.

Информация об авторах

Целищев Виталий Валентинович – доктор философских наук, профессор, кафедра гносеологии и истории философии Новосибирского национального исследовательского государственного университета (630090, Новосибирск, ул. Пирогова, 2); научный руководитель Института философии и права СО РАН (630090, Новосибирск, ул. Николаева, 8, e-mail:leitval@gmail.com).

Хлебалин Александр Валерьевич – кандидат философских наук, старший научный сотрудник Институт философии и права СО РАН (630090, Новосибирск, ул. Николаева, 8, e-mail: sasha_khl@mail.ru).

Information about the authors

Tselishchev Vitaliy Valentinovich – Doctor of Sciences (Philosophy), Professor, the Department of Gnoseology and History of Philosophy at Novosibirsk National Research State University (2, Pirogov st., Novosibirsk, 630090, Russia); Scientific Director at the Institute of Philosophy and Law, Siberian Branch of the Russian Academy of Sciences (8, Nikolaev st., Novosibirsk, 630090, Russia, e-mail: leitval@gmail.com).

Khlebalin Aleksandr Valerievich – Candidate of Sciences (Philosophy), Researcher at the Institute of Philosophy and Law, Siberian Branch of the Russian Akademy of Sciences (8, Nikolaev st., Novosibirsk, 630090 Russia, e-mail: sasha_khl@mail.ru)

Дата поступления 29.07.2019